

# **Non-Canonical Word Order and Subject-Object Asymmetry in Korean Case Ellipsis**

**Hanjung Lee**  
Sungkyunkwan University

**Nayoun Kim**  
Sungkyunkwan University

Proceedings of the 19th International Conference on  
Head-Driven Phrase Structure Grammar

Chungnam National University Daejeon

Stefan Müller (Editor)

2012

CSLI Publications

pages 427–442

<http://csli-publications.stanford.edu/HPSG/2012>

Lee, Hanjung, & Kim, Nayoun. (2012). Non-Canonical Word Order and Subject-Object Asymmetry in Korean Case Ellipsis. In Stefan Müller (Ed.): *Proceedings of the 19th International Conference on Head-Driven Phrase Structure Grammar, Chungnam National University Daejeon* (pp. 427–442). Stanford, CA: CSLI Publications.

## Abstract

The dispreference for subject case ellipsis in OSV sentences has been analyzed as resulting from a violation of a structural requirement on the position of bare subject NPs (Ahn and Cho 2006a, 2006b, 2007). In this study, we present evidence from an acceptability rating experiment demonstrating that OSV sentences containing a case-ellipsed subject exhibit acceptability patterns different from ungrammatical sentences violating a core syntactic principle on case assignment and that these sentences are judged acceptable when the subject refers to expected, predictable information in context. This evidence supports the conclusion that the dispreference for subject case ellipsis in OSV sentences is due to violations of probabilistic constraints that favor case marking for rare types of subjects and such violations can be remedied by non-syntactic information.

## 1 Introduction

Ellipsis is the phenomenon whereby speakers omit from an utterance normally obligatory elements of syntactic structure. One common type of ellipsis in Korean is case ellipsis, whereby case markers like *-i/-ka* and *-(l)ul* are omitted.

Although case ellipsis is possible for both subjects and objects, a comparison between subject and object case ellipsis as found in corpus and acceptability data shows that in general, subject case ellipsis occurs less frequently and is also less acceptable than object case ellipsis (Kim 2008; S. Lee 2009; H. Lee 2010, 2011a). One particular case of this general subject-object asymmetry in case ellipsis is the dispreference for subject case ellipsis in sentences that have the non-canonical OSV word order. Ahn and Cho (2006a, 2006b, 2007) observe that whereas a case-ellipsed direct object can appear in the non-canonical, sentence-initial position without resulting in ill-formedness, a subject cannot appear without following case marker in sentences that have the non-canonical OSV order:

- (1) a. Chelswu-lul     Mary-ka     manna-ss-e.  
Chelsoo-Acc     Mary-Nom     meet-Pst-Ind  
'Mary met Chelsoo.'
- b. Chelswu     Mary-ka     manna-ss-e.  
Chelsoo(-Acc)     Mary-Nom     meet-Pst-Ind  
'Mary met Chelsoo.'
- c. \*Chelswu-lul     Mary     manna-ss-e.  
Chelsoo-Acc     Mary(-Nom)     meet-Pst-Ind  
'Mary met Chelsoo.' (Ahn and Cho 2007: 54)

The main purpose of this paper is to provide a new probability-based analysis of this particular asymmetry between subject and object in case ellipsis that can also explain the general subject-object asymmetry in case ellipsis. In section 2, we will

first show that previous syntactic accounts are problematic because OSV sentences with a case-ellipsed subject that have been predicted to be syntactically ill-formed are acceptable when the subject refers to expected, predictable information in context. Section 3 introduces the rationale behind our probability-based account of case ellipsis. In probabilistic models of grammar (Boersma and Hayes 2001; Bod, Hay and Jannedy 2003; Bresnan 2007; Bresnan and Ford 2010), grammatical constraints are defined in terms of graded preferences, weights or rankings, rather than categorical or discrete levels of grammaticality. These models are well-suited to account for case ellipsis because they can describe syntactic phenomena in terms of grammaticality that emerges from preferences that develop over phrases and constructions. In turn, such preferences can be linked to factors that affect processing difficulty, e.g., frequency/probability of use, prototypicality, etc. Section 4 presents evidence from an acceptability rating experiment demonstrating that OSV sentences containing a case-ellipsed subject exhibit acceptability patterns different from ungrammatical sentences violating a core syntactic principle on case assignment and that these sentences are judged acceptable when the subject refers to expected, predictable information in context. This evidence supports the conclusion that the dispreference for subject case ellipsis in OSV sentences is due to violations of probabilistic constraints that penalize form reduction for rare types of subjects and such violations can be remedied by non-syntactic information.

## 2 Ahn and Cho's Syntactic Account

Ahn and Cho (2006a, 2006b, 2007) offer an account for the subject-object asymmetry in case ellipsis found in OSV sentences as well as other asymmetries which we will not discuss here. Their analysis, couched within the structural framework of the Minimalist Program (Chomsky 1995), rests on the following key assumptions defended in their earlier research:

- (2) a. An unmarked subject NP is a left-dislocated bare NP (that undergoes movement out of DP/ $\Phi$ P, stranding a resumptive pronoun in  $\Phi$ ).<sup>1</sup> The landing site of this NP is the Spec-Force position where it is assigned a generalized theta-role "aboutness."<sup>2</sup>
- b. All nominals including subject NPs in their canonical A-position must be projected to DP or  $\Phi$ P and hence cannot be a bare NP.
- c. Case markers on moved nominals must be pronounced unless they are left-dislocated.
- d. Unmarked object NPs in their canonical complement position can be part of a syntactic complex predicate. When generated inside a VP, they are bare

---

<sup>1</sup> In Ahn and Cho's analysis, an unmarked NP is treated as a bare NP, a noun phrase that is not required to be projected to DP or  $\Phi$ P. They assume three independent layers of nominal projections: NP,  $\Phi$ P, and DP. They further assume that D is correlated with Case in Korean and that  $\Phi$  is the projection of pronominal features such as number, person, gender, etc.

<sup>2</sup> Force may express the illocutionary force, modality, or the clausal type (Ahn and Cho 2006a).

NPs and are thus allowed to be caseless.

- e. An object that is overtly case-marked by the accusative case markers is a purely optional counterpart of the unmarked object.

The ill-formedness of case ellipsis for the subject NP in (1c) can be accounted for by the assumption (2a). Under Ahn and Cho's account, the bare subject NP *Mary* in (1c) occupies the Spec-T position, not the sentence-initial, Spec-Force position, as shown in (3):

- (3) [FP[DP John-ul]; [F[TP[NP Mary]; [T'[vP t<sub>j</sub> ...]]T] F]]

The derivation (3) is predicted to be ill-formed because the bare NP cannot occupy Spec-T, where the formal feature checking of  $\Phi/D$ -features is required.

However, OSV sentences with a bare subject NP, which Ahn and Cho predict to be ill-formed, are acceptable in an appropriate discourse context. Consider the following example:

- (4) A: ecey Minswu-ka i cip-ul sa-le o-ass-ta.  
 yesterday Minsoo-Nom this house-Acc buy-to come-Pst-Ind  
 haciman na-nun ku salam-hanthey nay cip an phal-a.  
 but I-Top that person-to my house(-Acc) not sell  
 '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 hay.  
 this house(-Acc) that person(-Nom) easily give up not do.  
 'He won't give up this house easily.'

In Korean, the OSV order typically marks the object as prominent information such as topic or contrastive focus, and the subject as new information (Choi 1999). This is illustrated in B's utterance in (4), where *i cip* 'this house' is the topic, and the subject *ku salam* 'that person', referring to *Minsoo*, is included within the comment (the information that is asserted about the topic), and represents the backgrounded part of the sentence, namely the part of the sentence which neither topic nor focus is assigned. In this context, unlike in Ahn and Cho's example (1c), case ellipsis for both the subject and the object is just as felicitous as case marking.

The above example is in sharp contrast to the case in which the subject in the OSV sentence is the focus:

- (5) A: ney-ka Minhi-lul cohahay?  
 you-Nom Minhi-Acc like  
 'Is it you who likes Minhi?'  
 B: ani, Minhi-lul Minswu-ka/\*Minswu cohahan-ta-ko!  
 no, Minhi-Acc Minsoo-Nom/Minswu(-Nom) like-Ind-QT  
 'No, it's Minsoo who likes Minhi!'

In B's utterance in (5), there is a clear preference for case marking for the subject.

The contrast between (4) and (5) shows that information predictability plays a crucial role in determining the acceptability of OSV sentences with a case-ellipsed subject. Such sentences, although unacceptable out of context, are judged natural by speakers when the subject represents expected, predictable information in context. In this case, both case ellipsis and case marking for the subject is felicitous.

### 3 Usage Probability and Subject-Object Asymmetry in Case Ellipsis

In this paper we propose a new usage-based account of the particular subject-object asymmetry in case ellipsis found in OSV sentences that can also explain other types of subject-object asymmetries in case ellipsis noted in the literature in a unified way. Our account of case ellipsis is based on the notion of 'usage probability.' We use the term 'usage probability' to refer to the probability of use of syntactic elements. One important factor that contributes to the high probability of subjecthood or objecthood is frequent associations between certain properties and grammatical functions. For example, subjects are more frequently animate than inanimate and definite than indefinite across languages; objects have the opposite default associations. In response to this, certain patterns of case marking have evolved, whereby more frequent types of objects (e.g., inanimate and indefinite objects) can be unmarked while rare types of objects (e.g., animate and definite objects) are overtly case-marked (Aissen 2003). Conversely, subjects can be unmarked when they are animate and definite, and overtly case-marked when they are inanimate and indefinite. The generalization that suggests itself here is that inanimacy and indefiniteness makes objecthood more likely because of their frequent association, and this increased probability permits zero object marking, by economy principles (6). When a direct object is animate or definite and objecthood is less likely, it is explicitly case-marked. Conversely, subjects can be unmarked when they are animate and definite and subjecthood is highly likely.

- (6) a. The more predictable a sign is, the shorter it is.  
b. The more frequent a sign is, the shorter it is.

Haspelmath (2008: 5) argues that any efficient sign system in which costs correlate with signal length follows the Zipfian principles in (6) (see also Bybee and Hopper (2001) and Hawkins (2004)). Evidence from syntactic reduction provides support for the pervasive effect of these principles on language use. In a study using a database of spontaneous English, Wasow, Jaeger and Orr (2011) found that speakers are less likely to mention the relativizer *that* in non-subject-extracted relative clauses (e.g., *I like the way (that) it vibrates*) when the relative clause is predictable (see also Jaeger (2006)). Further evidence comes from the optional *that*-mentioning in English complement clauses. Jaeger (2006) analyzed the same spontaneous database that was employed for the study by Wasow et al. (2011) and found that speakers are less likely to mention the complementizer *that* when the

complement clause is predictable.

In this paper, we will propose that the acceptability difference between subject and object case ellipsis in non-canonical OSV sentences reflects asymmetries in probabilistic properties of argument NPs.

In Korean, SOV is regarded as the canonical order and is also documented to be the most frequent word order. Kim (2008) has examined argument realization patterns by analyzing the use of what he calls ‘zero-marking’, nominative and accusative case markers and other particles such as *-(n)un* in 9,249 clausal units produced by 20 native speakers of Korean. His study found that among the various patterns of word order observed in conversational Korean, OV (51.95%), SOV (19.37%), V (19.23%) and SV (6.42%) are the most common patterns. The dominance of the OV pattern, which involves subject ellipsis and overtly realized object NPs, is due to asymmetries in the probabilistic patterns of the use of subjects and objects. Bloom (1990, 1993) argues that the tendency to omit subject NPs over object NPs may be due to discourse factors surrounding subjects, which tend to be more given than objects. Subjects are typically given information (that is, previously mentioned and already activated), while objects typically convey newly introduced information, it is not surprising that subjects tend to be omitted more than objects. Similarly, DuBois (1987) discusses why discourse factors explain why object ellipsis occurs less frequently and why transitive predicates tend to have more subject ellipsis than intransitive predicates. DuBois argues that there tends to be only one lexical argument (which contributes new information) in a clause in Sacapultec Maya. The lexical argument appears preferentially in the S (intransitive subject) or O (direct object) roles, but rarely in the A (transitive subject) role. This is because human agents (which occupy the A role) tend to be topic and given information in the sentence, while objects tend to be new information, and intransitive clauses tend to be used when new human referents are introduced. As a result, the A role tends to be reduced to an overt or zero pronoun.

Kim’s study also found that OSV occurs very rarely (1.92% of 3,692 tokens of transitive clauses). This order typically marks the object as prominent information such as topic or contrastive focus, and the subject as new information (Choi 1999). The correlation between the OSV order and the new subject can lead to a particular bias toward the form of the subject NP. Because newness is a rare and unexpected property for the subject, the current probability-based account predicts that the case-marked form to be preferred over the unmarked form as the suitable form for this less probable subject type, i.e., new information subject. This explains why case ellipsis for focus subject in (5B) is not acceptable.

The acceptability of (1b), repeated below as (7), can be explained similarly. Case marking on the sentence-initial object is motivated by the fact that objecthood is less likely in the sentence-initial position than in the immediately preverbal position. However, the unmarked object form is not ruled out because it is compatible with the probabilistic property of the OSV order (i.e., marking the object as given) and with the general information status of elements occurring the sentence-initial position (i.e., high predictability and low information content).

- (7) Chelswu(-lul)            Mary-ka            manna-ss-e.  
       Chelsoo(-Acc)            Mary-Nom            meet-Pst-Ind  
       ‘Mary met Chelsoo.’

Our probability-based account further predicts that the sentence (1c), although unacceptable out of context, is judged natural by speakers when the predictability of the subject referent increases. One such case is when the subject has a higher degree of givenness as in (4), repeated here as (8).

- (8) A: ecey            Minswu-ka            i cip-ul            sa-le            o-ass-ta.  
       yesterday Minsoo-Nom    this house-Acc    buy-to            come-Pst-Ind  
       haciman na-nun ku salam-hanthey nay cip            an phal-a.  
       but            I-Top    that person-to            my house(-Acc)    not sell  
       ‘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    hay.  
       this house(-Acc) that person(-Nom) easily    give up    not do.  
       ‘He won’t give up this house easily.’

In B’s utterance in (8), the referent of *i cip* ‘this house’ is the topic of the sentence, and the referent of *ku saram* ‘that person’, *Minsoo*, is part of the relationally new information predicated about the topic, as indicated by its occurrence in the immediately preverbal position. However, it is referentially given by virtue of having been previously mentioned in context, i.e., in A’s utterance. The referential givenness of the subject referent contributes to higher referential predictability (Jaeger 2006), and the increased predictability may in turn increase preference for the unmarked subject form. In this case, both case ellipsis and case marking with respect to the subject is felicitous: while the relational newness of the subject referent favors the use of case marking, use of case ellipsis for the subject in (8B) is also felicitous because of the increased predictability of the subject referent.

Using case markers to mark less probable phrases has been argued to have a processing advantage (Jaeger 2010): 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 on 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 subject (e.g., new subjects, focused subjects, subjects occurring in the non-initial

position following the object, etc.).<sup>3</sup> Violations of such constraints, unlike violations of core syntactic principles, give rise to mild unacceptability and can be remedied by non-syntactic information. This explains why referential predictability improves the acceptability of case ellipsis for rare types of subjects. The predictability condition on acceptable case ellipsis may be an important component of the recoverability condition on ellipsis phenomena in general, and thus it is not surprising that satisfying it improves the acceptability of unacceptable case ellipsis induced by violations of the probabilistic constraints that penalize zero marking for rare subject types.

This view of case marking can also account for the fact that in general, subject case ellipsis occurs less frequently and is also less acceptable than object case ellipsis (Kim 2008; S. Lee 2009; H. Lee 2010, 2011a). Given the high frequency of overt realization of object NPs and the rarity of overt realization of subject NPs, it is not surprising that case ellipsis is more acceptable for the more frequent type of explicit NPs, i.e., overt objects, whereas case marking is more acceptable for the rare type of explicit NPs, i.e., overt subjects.

A similar explanation is possible for the acceptability contrast between (1b) and (1c). As noted above, the subject in OSV sentences can be considered more marked than the object. It is *doubly marked* due to its association with two rare properties for subject, i.e., overt realization and non-canonical syntactic position, whereas the object in OSV sentences is associated with only one property unexpected for object, i.e., non-canonical syntactic position. Thus, case marking is more strongly enforced for the subject than for the object, and this explains why case ellipsis for the subject in OSV sentences is not acceptable out of context.

#### 4 Experimental Data

This section reports a rating experiment that elicits speakers' judgments on the acceptability of OSV sentences containing the case-marked or unmarked form of the transitive subject. Although acceptability judgments would probably not reflect actual performance in the same way as naturally occurring data would, the acceptability judgment task was chosen for the following reasons. First, given that the type of sentences that we are of interest in this study, i.e., OSV sentences, are not highly frequent in either written or spoken Korean, it was considered necessary to use elicitation tasks, for it was unpredictable to what extent data collection methods that use naturally occurring data would provide data rich enough for the present purposes. Second, the acceptability judgment task was used because it was considered very important to tightly control contexts and factors that are known to affect the frequency of case ellipsis in an experimental setting.

Our central hypothesis is that the degree of the acceptability of case ellipsis for the subject in such sentences is correlated with the degree of the subject referent's predictability in context. This hypothesis predicts: i) Case ellipsis is more acceptable

---

<sup>3</sup> See H. Lee (2003) for formalization of these constraints as Optimality-theoretic constraints.

in the high predictability condition than in the low predictability condition, whereas the pattern of the acceptability of case marking is reversed; and ii) In the high predictability condition, case ellipsis for the subject of OSV sentences is judged acceptable, whereas in the low predictability condition, case ellipsis for the subject of OSV sentences is judged mildly unacceptable.

Sixty undergraduate students of a university in Seoul participated in the experiment. Each participant was asked to read short conversations between speakers and indicate to what degree the two subject forms were suitable in the given context. To do this, they had to rate the acceptability of sentences containing a case-marked or unmarked subject by assigning them a grade from 1-5 on a five-point rating scale.

The predictability of the subject in OSV sentences was manipulated by means of variation in context sentences (sentences uttered by the first speaker). In the high predictability condition, the referent of the subject in the target OSV sentences is referentially given, i.e., introduced in the previous speaker's utterance. On the other hand, in the low predictability condition, the subject in the target OSV sentences is not only new to the discourse but also functions as a contrastive focus or an informational focus. Consistently with the typical information structure of OSV sentences, in both conditions, the referent of the object in the target OSV sentences is the topic of the sentence, and the referent of the subject is relationally new by virtue of being part of the comment predicated about the topic (the high predictability condition) or by virtue of being in focus (the low predictability condition). A sample stimulus translated into English is shown in (9):

(9) Example of judgment task in questionnaire

Instruction: Please read through the following conversations, then make a judgment on underlined sentences in each conversation by assigning them grades from 1-5. Use the following scale to make your judgments:

- 1 = Completely Unacceptable      2 = Unacceptable  
 3 = Just Barely Acceptable      4 = Acceptable  
 5 = Completely Acceptable

1) [High predictability condition]

A: ecey      Minswu-ka      i      cip-ul      sa-le      o-ass-e.  
 yesterday Minsoo-Nom      this house-Acc      buy-to      come-Pst-Ind  
 haciman      na-nun      ku      salam-hanthey      nay      cip-ul      an      pha-l-keya.  
 but      I-Top      that      person-to      my house-Acc      not sell-will  
 'Minsoo came (here) yesterday to buy this house. But I won't sell my house to him.'

B-1: i	cip-ul	<u>ku salam-i</u>	swipkey	phoki	an	ha-l-keya.
	this house-Acc	that person-Nom	easily	give up	not	do.
	‘He won’t give up this house easily.’ Your score: _____					
B-2: i	cip-ul	<u>ku salam</u>	swipkey	phoki	an	ha-l-keya.
	this house-Acc	that person(-Nom)	easily	give up	not	do.
	‘He won’t give up this house easily.’ Your score: _____					

To keep the influence of factors other than the predictability of the subject referent in the target sentences to a minimum, we have further controlled the items in the questionnaire in the following way:

- (10) a. Only definite subjects and objects were included as the head of subject and object NPs in the target OSV sentences. The following four kinds of definite subjects were included in our stimuli: pronouns, names, kinship terms and definite descriptions.
- b. Only human subjects and inanimate objects were included since they represent the most prototypical types of transitive subjects and direct objects in terms of animacy.
- c. In view of the finding that the previous occurrence of a parallel structure affects speakers’ use of syntactic structures (Bock 1986; Gries 2005; Szendrői 2005), the form of the subject and the object in the context sentences has been controlled by keeping them consistently case-marked. Furthermore, the word order of the context sentences was consistently SOV.
- d. In view of the finding that longer NPs are more likely to be marked with an overt particle than shorter NPs (Ono, Thompson and Suzuki 2000; Kim 2008), the length of the subject NP has been controlled so that the subject NP in the target sentences did not contain any phrase modifying the head noun and did not exceed 4 syllables.

There were 40 items per each predictability condition, and there were two versions of the target sentences of each item: one version contained a case-marked subject and the other version an unmarked subject. These 80 items were combined with 60 fillers belonging to another experiment. The stimuli and fillers were combined in three different orders for each list, to avoid ordering effect.

Thus, the experiment followed  $2 \times 2$  design, where the factors were 1) the subject’s predictability (high vs. low) and 2) subject form (case-marked vs. unmarked). The two versions of the target sentences were presented in a factorial design so that half the participants saw 30 stimuli with a case-marked subject, and half saw 30 stimuli with an unmarked subject.

The results of the ANOVA indicate a significant main effect of subject predictability ( $F(1, 118) = 24.88, p = .000$ ;  $F(1, 38) = 5.59, p = .029$ ). As shown in Table 1, the mean judgments for case-marked subjects in the OSV sentences were higher in the low predictability condition than in the high predictability condition,

whereas the pattern of the mean judgments for unmarked subjects was reversed. This confirms our first major prediction that whereas case marking is more acceptable in the low predictability condition than in the high predictability condition, case ellipsis is more acceptable in the high predictability condition than in the low predictability condition. However, the acceptability of the OSV sentences with the unmarked form of the high-predictability subject counters to the predictions of purely syntactic accounts (e.g., Ahn and Cho (2006a, 2006b, 2007)) because such accounts predict OSV sentences with an unmarked subject to be syntactically ill-formed.

Table 1. Average ratings

	High predictability	Low predictability	Means
Subj-Nom	3.46	3.98	3.72
Subj- $\emptyset$	3.72	2.47	3.10

The results of the ANOVA also indicate a significant main effect of subject form ( $F(1, 118) = 75.47, p = .000$ ;  $F(1, 38) = 47.46, p = .000$ ). As predicted, in the high predictability condition, the mean judgments for unmarked subjects were significantly higher than those for case-marked subject. By contrast, in the low predictability condition, case-marked subjects showed higher acceptability values than unmarked subjects.

Also noteworthy is that whereas only the case-marked subject form is judged acceptable and the case-ellipsed form is judged mildly unacceptable in the low predictability condition, both forms are judged acceptable showing acceptability values higher than 3 in the high predictability condition. This supports the second major prediction tested in this experiment. As discussed earlier, case marking is felicitous because the (relational) newness of the subject referent favors the use of case marking. Use of case ellipsis for the subject in OSV sentences is also expected to be felicitous because it matches the increased expectancy of the reduced form in the high predictability condition. However, the acceptability of both forms in the high predictability condition contrasts directly with the predictions of Ahn and Cho's syntactic account (2006a, 2006b, 2007) because on their account, OSV sentences with an unmarked subject are predicted to be syntactically ill-formed.

We also found a significant interaction between subject form and subject predictability ( $F(1, 118) = 216.77, p = .000$ ;  $F(1, 38) = 78.19, p = .000$ ). As shown in Table 1, the subjects' scores of the acceptability of OSV sentences containing the unmarked subject increase from the low predictability condition to the high predictability condition, whereas the acceptability of sentences containing the case-marked subject shows the opposite pattern. Thus, the results of this analysis indicate that the degree of the acceptability of case ellipsis for the subject in OSV sentences is correlated positively with the degree of subject predictability, whereas the degree of the acceptability of case marking for the subject in OSV sentences is

correlated negatively with the degree of subject predictability.

Overall, our results clearly indicate that speakers' judgments of OSV sentences with a case-marked or unmarked subject are sensitive to the degree of the predictability of the subject referent in context. The fact that the OSV sentences containing an unmarked subject were judged not only acceptable but also more acceptable than those with a case-marked subject in the high predictability condition provides strong support for the view advocated here that speakers' judgments of acceptability are affected by satisfaction or violation of probabilistic expectations about form reduction and predictability.

However, there is one crucial methodological limitation in this experiment. In this study, acceptability data were obtained from speakers' judgments on the written sentences. This constitutes a limitation because it does not take into consideration the possible effect of the prosody of non-canonical word order on judgments of acceptability. It has been observed that scrambling has a prosodic effect of shifting the intermediate phrase boundary to the left along with the scrambled phrase. As a consequence, the post-scrambled position gets de-accented or prosodically reduced (Jun 1993; Kenstowicz and Sohn 1997). The unavailability of natural prosody is likely to have contributed to low judgments of OSV sentences, particularly those containing a case-ellipsed subject.

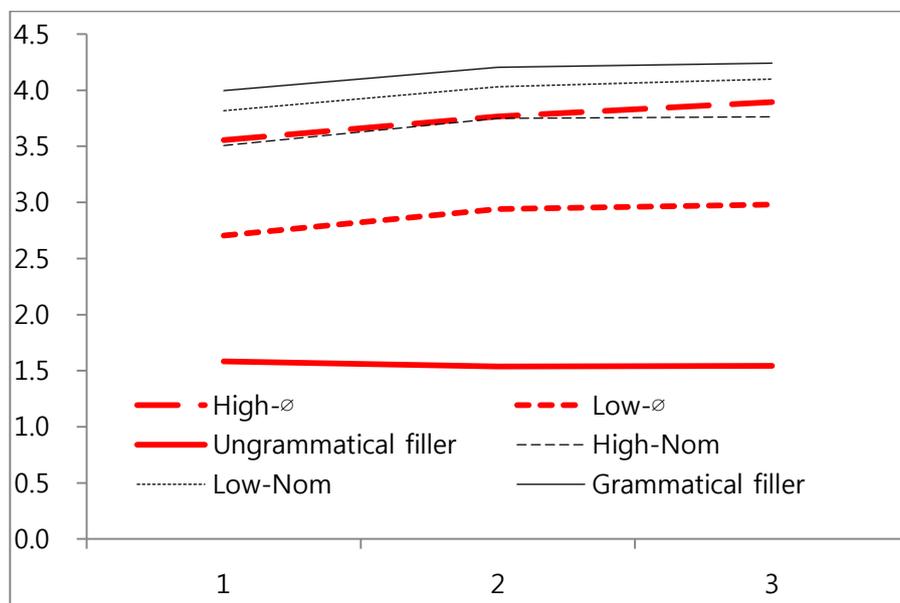
In a follow-up experiment designed to investigate this possibility, we elicited acceptability data by asking participants to listen to conversations between two speakers spoken with natural prosody. The overall results converge with the basic findings of the experiment reported in this section, showing that whereas only the case-marked subject form is judged acceptable and the case-ellipsed form is judged mildly unacceptable in the low predictability condition, both forms are judged acceptable showing acceptability values higher than 3 in the high predictability condition. Furthermore, the results also indicate that OSV sentences were judged higher in this follow-up experiment than in the experiment reported here. As Table 2 shows, this effect of increased ratings was most noticeable in the OSV sentences with a low-predictability subject that is case-ellipsed: these sentences were judged significantly higher (means: 2.87) than ungrammatical filler items (means: 1.55) where the direct object of agentive transitive verbs were marked with nominative case markers instead of accusative case markers. However, it is unexpected under purely syntactic accounts why the OSV sentences that are predicted to be ill-formed show acceptability values close to moderately grammatical level and why acceptability judgments for these sentences change with manipulation of non-syntactic information such as context and prosody.

Table 2. Average ratings

	High predictability	Low predictability	Means
Subj-Nom	3.67	3.98	3.83
Subj-Ø	3.74	2.87	3.31

Additional support for the hypothesis that the dispreference for subject case ellipsis in OSV sentences is due to violations of probabilistic constraints that favor case marking for rare types of subjects comes from an analysis of patterns of acceptability exhibited by the experimental items and filler items. As Figure 1 shows, OSV sentences in all of the four experimental conditions exhibited acceptability patterns identical to those of grammatical filler items in that acceptability judgments increase over the course of the experiment. But acceptability judgments for ungrammatical filler items violating a core syntactic principle on case assignment were not significantly affected by list position and stayed constant over the course of the experiment.

Figure 1. Patterns of acceptability judgments



This effect of increased ratings after repeated exposure to structurally similar sentences follows naturally from the perspective that the OSV sentences with an unmarked subject are hard-to-process constructions which get better with experience.<sup>4</sup> However, the common upward trend in acceptability judgments for

<sup>4</sup> Several studies suggest that only grammatical strings including moderately grammatical sentences and grammatical sentences that are difficult to process get better with repeated exposure. Hofmeister, Jaeger, Amon, Sag and Snider (In press), for example, present evidence that Superiority violations in English, which has traditionally been considered violations of syntactic constraints, show gradient acceptability (rather than being categorically out or ungrammatical) and that judgments on certain types of Superiority violations increase with exposure while processing times decrease. Their analysis further shows that the observed gradience in acceptability is correlated with processing difficulty at the verb. Based on this evidence, Hofmeister et al. argue that Superiority effects reflect online processing costs that can be attenuated with repeated exposure.

the experimental OSV sentences and the grammatical filler items is left unexplained in purely syntactic accounts of case ellipsis.

## 5 Conclusion

In this paper, we have presented a new account of variable case marking that can explain the subject-object differences in case ellipsis by the interaction between grammatical constraints on the use of case ellipsis and the predictability condition on ellipsis recoverability: the unacceptability of case ellipsis for subjects in noncanonical OSV sentences and non-specific subjects, and *wh*-word subjects is caused by violations of probabilistic linguistic constraints that penalize the use of case ellipsis for rare types of subjects, which can be viewed as a grammaticalization of the speakers' preference to avoid form reduction for less frequent types of phrases for increased processing efficiency.

Violations of these constraints, unlike violations of core syntactic principles, give rise to mild unacceptability and can be remedied by manipulation of non-syntactic information. This explains why case ellipsis for rare types of subjects is judged acceptable when the subject represents expected, predictable information in context. These results provide strong support for the view that grammatical asymmetries manifested by the subject-object asymmetries in case ellipsis should be explained by asymmetries in the usage probability of the properties of argument NPs in context, not by categorical syntactic constraints. Our results also add to the growing body of evidence that native speakers' knowledge of grammar includes access to fine-grained predictability and probability (Jaeger 2006, 2010; Bresnan 2007, Bresnan and Ford 2010; H. Lee 2010, 2011a, 2011b).

## References

- Ahn, H.-D. and Cho, S. 2006a. *Wh*-topics and Unpronounced Cases on *Wh*-phrases. *Studies in Generative Grammar* 16, 61-90.
- Ahn, H.-D. and Cho, S. 2006b. Layered Nominal Structures: Implications for Caseless Nominals. *Korean Journal of Linguistics* 31, 165-185.
- Ahn, H.-D. and Cho, S. 2007. Subject-Object Asymmetries of Morphological Case Realization. *Language and Information* 11, 53-76.
- Aissen, J. 2003. Differential Object Marking: Iconicity vs. Economy. *Natural Language and Linguistic Theory* 21, 435-483.
- Arnold, J. 1998. *Reference Form and Discourse Patterns*. Ph.D. dissertation, Stanford University.
- Bloom, P. 1990. Subjectless Sentences in Child Language. *Linguistic Inquiry* 21, 491-504.
- Bloom, P. 1993. Grammatical Continuity in Language Development: The Case of Subjectless Sentences. *Linguistic Inquiry* 24, 721-734.
- Bock, J. K. 1986. Syntactic Persistence in Language Production. *Cognitive Psychology* 18, 355-387.

- Bod, R., Hay, J. and Jannedy, S. (eds.). 2003. *Probabilistic Linguistics*. Cambridge: MIT Press.
- Boersma, P. and Hayes, B. 2001. Empirical Tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32, 45-86.
- Bresnan, J. 2007. Is Knowledge of Syntax Probabilistic? Experiments with the English Dative Alternation. In S. Featherston and W. Sternefeld (eds.), *Roots: Linguistics in Search of Its Evidential Base (Studies in Generative Grammar)*, pp 75-96, Berlin: Mouton de Gruyter.
- Bresnan, J. and Ford, M. 2010. Predicting Syntax: Processing Dative Constructions in American and Australian Varieties of English. *Language* 86, 168-213.
- Bybee, J. and Hopper, P. (eds.). 2001. *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins.
- Choi, H.-W. 1999. *Optimizing Structure in Context: Scrambling and Information Structure*. Stanford: CSLI Publications.
- Chomsky, N. 1995. *The Minimalist Program*. Cambridge: MIT Press.
- Gries, S. 2005. Syntactic Priming: A Corpus-based Approach. *Journal of Psycholinguistic Research* 34, 365-399.
- Haspelmath, M. 2008. Frequency vs. Iconicity in Explaining Grammatical Asymmetries. *Cognitive Linguistics* 19, 1-33.
- Hawkins, J. 2004. *Efficiency and Complexity in Grammar*. Oxford: Oxford University Press.
- Hofmeister, P., Jaeger, F. T., Arnon, I., Sag, I. and Snider, N. In press. The Source Ambiguity Problem: Distinguishing the Effects of Grammar and Processing on Acceptability Judgments. *Language and Cognitive Processes*.
- Jaeger, F. T. 2006. *Redundancy and Syntactic Reduction in Spontaneous Speech*. Ph.D. dissertation, Stanford University.
- Jaeger, F. T. 2010. Redundancy and Reduction: Speakers Manage Syntactic Information Density. *Cognitive Psychology* 61, 23-62.
- Jun, S.-A. 1993. *The Phonetics and Phonology of Korean Prosody*. Ph.D. dissertation, Ohio State University.
- Kenstowicz, M. and Sohn, H.-S. 1997. Phrasing and Focus in Northern Kyungsang Korean. *MIT Working Papers in Linguistics* 30, 25-47.
- Kim, T. 2008. *Subject and Object Markings in Conversational Korean*. Ph.D. dissertation, State University of New York at Buffalo.
- Lee, H. 2003. Prominence Mismatch and Markedness Reduction in Word Order. *Natural Language and Linguistic Theory* 21, 617-680.
- Lee, H. 2010. *Explaining Variation in Korean Case Ellipsis: Economy versus Iconicity*. *Journal of East Asian Linguistics* 19, 291-318.
- Lee, H. 2011a. *Contrastive Focus, Usage Probability and Gradients in Korean Case Ellipsis*. *Discourse and Cognition* 18, 215-240.
- Lee, H. 2011b. *Gradients in Korean Case Ellipsis: An Experimental Investigation*. *Lingua* 121, 20-34.

- Lee, S. 2009. The Characteristics of Case Omission in Spoken Korean. *Journal of the Humanities* 43, 29-59. [written in Korean]
- Ono, T., Thompson, S. A., and Suzuki, R. 2000. The Pragmatic Nature of the So-called Subject Marker *ga* in Japanese: Evidence from Conversation. *Discourse Studies* 2, 55-84.
- Szmrecsányi, B. 2005. Language Users as Creatures of Habit: A Corpus-based Analysis of Persistence in Spoken English. *Corpus Linguistics and Linguistic Theory* 1, 113-149.
- Wasow, T., Jaeger, F. T., and Orr, D. 2011. Lexical Variation in Relativizer Frequency. In H. Wiese and H. Simon (eds.), *Expecting the Unexpected: Exceptions in Grammar*, pp 175-196. Berlin and New York: Mouton de Gruyter.