

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

Hanjung Lee and Nayoun Kim

Department of English Language and Literature, Sungkyunkwan University,
25-2 Sungkyunkwan-ro, Jongro-ku, Seoul 110-745, Korea
{hanjung, na881017}@skku.edu

Summary. Case ellipsis for subjects and objects in Korean exhibits several clear asymmetries in acceptability. The main purpose of this study is to provide a new probability-based analysis of a particular type of subject-object asymmetry in case ellipsis found in OSV sentences that can also explain the general subject-object asymmetry in case ellipsis. In previous syntactic accounts of case ellipsis, the unacceptability of OSV sentences with the case-ellipsed subject has been attributed to the violation of a structural condition which requires the bare subject NP to occupy the sentence-initial, Spec-Force position (Ahn and Cho, 2006a, 2006b, 2007). However, experimental evidence suggests that OSV sentences with the case-ellipsed subject are judged acceptable when the subject represents expected, predictable information in context. We argue that speakers' judgments of non-canonical OSV sentences with the case-ellipsed subject or object reflect their knowledge of probabilistic properties of argument NPs rather than purely structural properties of syntactic derivations.

Keywords: case ellipsis, case marking, subject-object asymmetry, predictability, usage probability

1 Introduction

Particle ellipsis is a phenomenon in which speakers omit NP-final particles. One common type of particle ellipsis in Korean is case ellipsis, whereby case markers such as *-i/-ka* (nominative) and *-(l)ul* (accusative) are omitted. Although case ellipsis is possible for subjects and objects, a comparison between patterns of subject and object case ellipsis 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, 2011).

One particular case of this general subject-object asymmetry in case ellipsis we will focus on in this study is the restriction of the appearance of case-ellipsed subjects to sentences that have a canonical SOV 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, subjects cannot appear without following case markers in sentences that have the noncanonical OSV order:

- (1) a. Chelswu-lul Mary-ka manna-ss-e.
 Chelsoo-Acc Mary-Nom meet-Pst-Inc
 'Mary met Chelsoo.'
 b. Chelswu Mary-ka manna-ss-e.
 Chelsoo(-Acc) Mary-Nom meet-Pst-Inc
 'Mary met Chelsoo.'

* HPSG 2012 Conference/Ellipsis Workshop, July 18-21, 2012, Chungnam National University, Daejeon, Korea
Copyright © 2012 by Hanjung Lee and Nayoun Kim

- (2) a. Chelswu-lul Mary-ka manna-ss-e.
 Chelsoo-Acc Mary-Nom meet-Pst-Inc
 ‘Mary met Chelsoo.’
 b. *Chelswu-lul Mary manna-ss-e.
 Chelsoo-Acc Mary(-Nom) meet-Pst-Inc
 ‘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 the following we will first show that previous syntactic accounts are problematic because OSV sentences with an unmarked subject that have been predicted to be syntactically ill-formed by previous syntactic accounts are acceptable when the subject represents expected, predictable information in context. We then present an alternative account of case ellipsis that explains the particular subject-object asymmetry in case ellipsis found in OSV sentences and the general subject-object asymmetry in case ellipsis in a unified way in terms of asymmetries in the probabilistic properties associated with subjects and objects.

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, rests on the following key assumptions defended in their earlier research:

- (3) a. An unmarked subject NP is a left-dislocated bare NP (that undergoes a movement out of DP/ Φ P, stranding a resumptive pronoun in Φ). The landing site of this NP is the Spec-Force position where it is assigned a generalized theta-role “aboutness.”
 b. All nominals including subject NPs in their canonical A-position must be projected to DP or Φ 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 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 (2b) can be accounted for by the assumption (3a). Under Ahn and Cho's account, the bare subject NP *Mary* in (2b) occupies the Spec-T position, not the sentence-initial, Spec-Force position, as shown in (4):

- (4) [_{FP}[_{DP} John-ul]_j] [_F[_{TP}[_{NP} Mary]_i [_T[_{VP} t_j ...]]_T] F]

The derivation (4) is predicted to be ill-formed because the bare NP cannot occupy Spec-T, where the formal feature checking of Φ /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:

- (5) A: ecey Minswu-ka i cip-ul sa-re oa-ss-ta.
 yesterday Minsoo-Nom this house-Acc buy-to come-Pst-Ind
 haciman na-nun ku saram-hanthey nay cip an phan-ta.
 but I-Top that person-to my house(-Acc) not sell-Ind
 kurem phoki ha-l-keya.

then give up do-will
 ‘Minsoo came (here) yesterday to buy this house. But I won’t sell my house to him. (He) will give (it) up then.’

B: *i cip*(-ul) *ku saram*(-i) swipekey phoki an ha-l-keya.
 this house(-Acc) that person(-Nom) easily give up not won’t 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 (5), where *i cip* ‘this house’ is the topic, and the object *ku saram* ‘that person’, referring to *Minsoo*, is part of the comment (the information that is asserted about the topic). In this context, unlike in Ahn and Cho’s example (2b), case ellipsis for both the subject and the object is just as felicitous as case marking.

3 A New Probabilistic Account of Subject-Object Asymmetry in Case Ellipsis

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 case-marked form is preferred over the unmarked form as the suitable form for this less probable subject type, i.e., new information subject. As clearly demonstrated by the difference between (2a) and (2b), this is exactly what we find.

The acceptability of (1b) can be explained similarly. As noted above, the OSV order typically marks the object as prominent given information such as topic. As a referent that has been under discussion constitutes a highly predictable element in that it is associated with a high expectation that it is to be mentioned again in subsequent utterances (Arnold, 1998), and thus, it tends to be encoded with attenuated or reduced forms such as a zero pronoun or an unmarked overt NP. Thus, it is likely that speakers’ judgment of (1b) reflects their knowledge of the probabilistic association between high predictability and form reduction (Jaeger, 2006; Haspelmath, 2008). 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).

Our probability-based account further predicts that the sentence (2b), although unacceptable out of context, is judged natural by speakers only when the predictability of the subject referent increases. One such case is when the subject has a higher degree of givenness as in (5). In B’s utterance in (5), 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 (5B) is also felicitous because of the increased predictability of the subject referent.

The above case is in sharp contrast to the case in which the subject in the OSV sentence is both relationally and referentially new:

(6) 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 (6), *Minsoo* represents highly unpredictable information because this phrase is contrastively focused as well as discourse new. As predicted, there is a clear preference for case marking for this subject.

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 like 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.). Violations of such constraints, unlike violations of core syntactic principles, give rise to mild unacceptability and can be remedied by manipulation of context or 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, 2011). Explicit NPs occur more frequently as objects than as transitive subjects, whereas argument NPs are omitted more frequently when they are transitive subjects. The dominance of subject ellipsis and overtly realized object NPs is likely due to asymmetries in the usage probability of the properties associated with subjects and objects. 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.

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

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 (1 indicating completely unacceptable and 5, perfectly acceptable).

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). There were 30 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 60 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 2×2 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. 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: Mean judgments for subjects in OSV sentences

	High predictability	Low predictability	Means
Case-marked	3.87	4.51	4.19
Unmarked	3.90	2.22	3.06

The results of the ANOVA also indicate a significant main effect of subject form. 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 in the low predictability condition, both forms are judged acceptable showing acceptability values higher than 3 in the high predictability condition. This is expected under the current account: 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. 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, supporting the second major prediction tested in this experiment.

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.

5 Conclusion

We have argued that speakers' judgments of non-canonical OSV sentences with the unmarked subject or object reflect their knowledge of probabilistic properties of argument NPs. On this view, the unacceptability of OSV sentences in which the subject is unmarked can be accounted for as the result of the violation of the probabilistic correlation between the less probable argument type (i.e., new subject and explicit subject) and complex formal marking. Our probability-based account of variable case marking also explains why the same OSV sentence is acceptable only when the predictability of the subject referent increases in context. These results provide strong support for the view that native speakers' knowledge of grammar includes knowledge of probabilistic information (Bresnan and Ford, 2010; Jaeger, 2006, 2010; H. Lee, 2010).

References

- Ahn, H.-D. and S. Cho. 2006a. *Wh*-Topics and Unpronounced Cases on *Wh*-Phrases. *Studies in Generative Grammar*, 16(1), 61-90.
- Ahn, H.-D. and S. Cho. 2006b. Layered Nominal Structures: Implications for Caseless Nominals. *Korean Journal of Linguistics*, 31(2), 165-185.
- Ahn, H.-D. and S. Cho. 2007. Subject-Object Asymmetries of Morphological Case Realization. *Language and Information*, 11(1), 53-76.
- Arnold, J. 1998. *Reference Form and Discourse Patterns*. Ph.D. thesis, Stanford University.
- Bresnan, J. and M. Ford. 2010. Predicting Syntax: Processing Dative Constructions in American and Australian Varieties of English. *Language*, 86(1), 168-213.
- Choi, H.-W. 1999. *Optimizing Structure in Context: Scrambling and Information Structure*. Stanford: CSLI Publications.
- Haspelmath, M. 2008. Frequency vs. Iconicity in Explaining Grammatical Asymmetries. *Cognitive Linguistics*, 19(1), 1-33.
- Jaeger, T. F. 2006. *Redundancy and Syntactic Reduction in Spontaneous Speech*. Ph.D. thesis, Stanford University.
- Jaeger, T. F. 2010. Redundancy and Reduction: Speakers Manage Syntactic Information Density. *Cognitive Psychology*, 61(1), 23-62.

- Kim, T. 2008. *Subject and Object Markings in Conversational Korean*. Ph.D. thesis, State University of New York at Buffalo.
- Lee, H. 2010. Explaining Variation in Korean Case Ellipsis: Economy versus Iconicity. *Journal of East Asian Linguistics*, 19(4), 291-318.
- Lee, H. 2011. Contrastive Focus, Usage Probability and Gradients in Korean Case Ellipsis. *Discourse and Cognition*, 18(3), 219-244.
- Lee, S. 2009. The Characteristics of Case Omission in Spoken Korean. *Journal of the Humanities*, 43, 29-59. [written in Korean]