

"Acquisition of the subject and topic nominals and markers in the spontaneous speech of young children in Korean," in Chungmin Lee, Greg Simpson, and Youngjin Kim (eds.) *The Handbook of Korean Psycholinguistics* (General Editor: Ping Li), 23-33. Cambridge University Press.

# 1 Acquisition of the subject and topic nominals and markers in the spontaneous speech of young children in Korean

*Chungmin Lee and Sook Whan Cho*<sup>\*</sup>

## Purpose

The purpose of this paper is to examine the acquisition of the topic marker *-(n)un* and the subject (nominative) marker *-i/-ka* in Korean.<sup>1</sup> We will be concerned, in particular, with the way in which these case markers are produced over time, considering children's spontaneous speech as well as experimental data at several developmental time points. In this inquiry, we will be interested to find out how Korean children use or drop these markers on noun phrases, as well as how they produce or omit the entire topic/subject phrases over time, based on the pragmatic and syntactic properties governing them.<sup>2</sup> We show the following developmental sequence of the stages: **null topic/subject** → **bare nominal topic/subject** → **marked topic/subject**. Once these markers emerge, the developmental sequences for the functions of the topic and subject markers are: **contrastive topic** → **global/thematic topic** and **focus subject** → **neutral subject**. The latter two sequences are surprising in terms of markedness because they proceed from 'marked' to 'unmarked', unlike the sequence **null** → **bare** → **marked** for the use and marking of topic and subject noun phrases.

<sup>\*</sup> We thank Pat Clancy for making very helpful comments on the final draft of this joint paper.

<sup>1</sup> We take the Korean topic to be a sentence-level category marked by the affix *-un/-nun* and typically located in the initial position of an utterance, which is given or assumed to be given or easily elicited from the discourse or context. It is non-contrastive and non-focal. When it is anaphoric, it can be null. The same marker, typically with a high tone, functions as a contrastive topic, which, since it is contrastive, is focal as well as topical, as clarified recently (C. Lee, 2006). The position of a contrastive topic can be the initial or middle position of a sentence.

<sup>2</sup> Korean examples in this paper are transliterated on the basis of the Yale Romanization System in Martin and Lee (1969: 575). Hyphens are used to indicate morpheme boundaries.

**Korean***Word order and case markers*

As is well-known in the literature, Korean word order is, in general, regarded as fairly flexible; the preferred word order is either the SOV pattern, or a combination of SOV and OSV structures. Word order can vary for a variety of reasons, including topicality, focus, and emphasis, among others. The relatively free word order of Korean is facilitated by case markers that indicate the grammatical functions of the nouns used in a sentence. Korean is a discourse-prominent language, like Chinese and Japanese. Hence, sentences do not always require an overt subject or object when these elements are predictable from the discourse context. Korean differs in this respect from a pro-drop language such as Italian, in which object drop is impossible. Likewise, case markers can also be omitted optionally or pragmatically when recoverable from the context.

*Pragmatic constraint*

Korean is a discourse-prominent language and, therefore, nominals can drop for pragmatic reasons, if predictable in the immediate environment, as outlined in (1).

- (1) Pragmatic constraint:  
Drop nominals if predictable in the immediate discourse environment.

Overtly contrastive topic-marked nominals, on the other hand, often encode focal or contrastive information. As discussed in detail in C. Lee (2001), non-contrastive topics in Korean, and sometimes even contrastive topics, typically take sentence-initial position when the noun phrase conveys information that is shared by the speaker and the listener (e.g. discourse topic, physically present objects, implicitly understood facts, etc.); non-contrastive topics often occur at the end of an utterance. Consider sentence (2) for an illustration.

- (2) a. *Hankuk-uy kaul-un tanpwung-i coh-a.*  
Korean-of fall-TOP colored.leaves-NOM nice-DEC  
'As for Korean autumn, fallen leaves are nice.'  
b. *Tanpwung-i coh-a, ku-cuum-ey-nun.*  
leaves-NOM nice-DEC that-around-TIME-TOP  
'Around that time, fallen leaves are nice.'

The noun phrase '*hankuk-uy kaul-un*' in (2a) is new information being introduced as a topic and is produced in sentence-initial position. It is hearer-new but is in the domain of common knowledge; the hearer cannot challenge it. The topical phrase '*ku-cuum-ey-NUN*' contains the anaphoric expression '*ku-cuum*' ('that-time'), indicating that the speaker assumes that the listener

knows what it may refer to. The shared information is sometimes placed in sentence-final position, as can be seen in (2b), where it is like an anti-topic in English, e.g. *She<sub>i</sub> is a real angel, my sister<sub>i</sub>* (Lambrecht 1994).

### *Morpho-syntactic constraint*

In addition to the pragmatic constraint in (1), the referents of null nominals can be predicted from morphemes such as the plural marker *-tul*, the honorific marker *-si*, and modals. These morphemes are required to agree with plurality, honorificity, and person/number of the topic/subject, and hence can provide information about the referent of the topic/subject, if it is missing, as exemplified in (3).

- (3) a. Ø ppali-tul ttwi-ess-ta.  
           fast-PL run-PAST-DEC  
           ‘(People) ran fast.’  
   b. Ø o-si-ess-ta.  
           come-HON-PAST-DEC  
           ‘(Someone respected) came.’

The grammatical morphemes *-tul* and *-si* encode plurality and honorificity, respectively, on the missing nominal in the subject position; the semantic content of its referent is recoverable in accordance with the morpho-syntactic constraint in (4) below.

- (4) Morpho-syntactic constraint:  
       Drop nominals if predictable from the verbal affixes in the same clause.

This constraint, however, is semantically or pragmatically motivated. A plural marker on a manner adverb may be a syntactic copy from the topic (from subject) or subject indicating its semantic plurality. Furthermore, an honorific marker on the verb, which shows the speaker’s respect for the topic (from subject) or subject referent, is crucially pragmatic. The constraints that a respected subject referent cannot be the speaker him/herself and that one does not show both a lack of respect (e.g. by using the speech level *-ta* in (3b) and respect (e.g. by using the honorific *-si*) for the same person at the same time (pragmatic contradiction) are again surely pragmatic. The referent of the subject in (3b) can therefore be either second or third person, if we assume that the subject honorific marker *-si* in (3b) shows such agreement; the failure to honor the hearer, however, which is conveyed by the use of *-ta* at the end, excludes the second person, leaving reference to a third person as the only possibility.

Other verbal affixes, such as modal morphemes, can also help in recovering the content of a null subject, as illustrated in (5).

26 *Chungmin Lee and Sook Whan Cho*

- (5) **John:** Swuni<sub>i</sub>-nun mwe ha-ni?  
Swuni-TOP what do-INTERROGATIVE  
'What is Swuni doing?'

**Mary:** Ø<sub>i</sub> no-n-tay.  
play-PRESENT-REPORTATIVE  
'(Swuni) is reported to be playing.'

The reportative *-tay* requires that the topic (from subject) co-occurring in the same local sentence be a third person noun phrase; first and second person noun phrases are not allowed due to subject-modal agreement, as in (6) below. The referent of the null subject in (5) can be recovered with the help of the morpho-syntactic constraint in (4).

- (6) Subject-modal agreement:  
The subject nominal and the modal affixed on the verb co-occurring in the same minimal clause must agree in person (and number).

This constraint again is pragmatically motivated; the reportative *-tay* comes from the complex underlying structure of 'is SAID to play' and the subject of the verb SAY, i.e. the source of the report, is epistemically constrained not to be identical to the speaker = subject or the hearer of the entire utterance (C. Lee, 1990).

## Methodology

### *The children*

The Korean data used for the purpose of this paper consist of tape-recorded speech samples from eight children, as follows.<sup>3</sup>

### *Data*

During each visit caregiver-child pairs were audio-taped for an hour by a fluent speaker of Korean.<sup>4</sup> The visits were spaced approximately one or two weeks apart (see Table 1.1) and the caregivers were asked to do whatever they would normally do at that time of day. In transcribing and coding the data, this study largely adopts the system developed in MacWhinney (1995).<sup>5</sup>

<sup>3</sup> JK's data were collected and transcribed by Seung-Bok Lee, JW's by Hae-Rhee Kim. We would like to thank the investigators and their subjects. CK's data were collected and transcribed by Chungmin Lee, AL's, AN's, and PL's by Sook Whan Cho. We would like to thank our subjects.

<sup>4</sup> JK was 0;1 when he began to be observed and tape-recorded while interacting with his mother, grandmother, and nanny. His mother was a developmental psychologist and took notes about details concerning JK's verbal and non-verbal behavior during each session of taping.

<sup>5</sup> MLU was not calculated in the present study, since the transcripts are in the Korean alphabet, making it impossible to use CLAN (MacWhinney, 1995) to compute MLU.

Table 1.1. *The children: age and frequency of visits*

Name	Age	Frequency of visits
AL	2:2–2:9	bi-weekly
AN	2:10–3:5	bi-weekly
C	2:0–2:2	weekly
HS	1:8–2:11	weekly
JK	0:1–3:0	weekly and bi-weekly
CK	1:3–3:11	every day in principle
JW	2:0–3:3	bi-weekly
PL	2:7–3:2	bi-weekly
Y	1:3–3:11	every day in principle

## Results

### *Development of pragmatic constraints*

*Topic-marked nominals and topic markers* Topic recoverability has been noted in numerous studies for decades. Gruber (1967: 39, 64–5), for example, once argued that a topic–comment pattern is likely to underlie the subject–predicate structure. Along similar lines, Givón (1979: 294–5) noted that at the very start of language development, communication is “exclusively about the immediate environment,” which is “largely predictable.” Givón goes on to argue that early child discourse is characterized by the gradual learning of strategies for making the topic obvious to others (1979: 300). More recently, Clancy (1993) reported that Korean-speaking subjects (aged 1:8–2:1 and 1:10–2:3) used overt subjects in queries about physically visible, or newly introduced information, while dropping them in referring to highly predictable information. These previous studies suggest that development reflects sensitivity to topicality.

Topic prominence seems to be supported by Korean child data. Korean children have previously been observed to be initially motivated by the universal topic–comment relation and to drop nominals that refer to old, topical information (e.g. S.W. Cho, 1994, 2001). In her study of four children (1:10–2:2, 2:0–2:2, 2:3–2:6, 3:0–3:4), Cho (1994) found that the majority of missing topics/subjects (91%, 159/174 utterances) expressed old information while overt topics/subjects referred to newly introduced information (87%, 204/292 utterances). This result clearly demonstrates Korean children’s sensitivity to the function of old vs. newly introduced information in association with the production of arguments.

As discussed above, the topic marker *-nun* in Korean can also refer to a contrastive focal entity as well as to old, given information. Interestingly

enough, younger Korean children have been observed to overtly produce the topic marker in encoding contrastive information.<sup>6</sup> C. Lee (2001) reports that among marked nominals contrastive topics are acquired at the earliest phase of development. Younger children would start out, for example, by producing an utterance like (7).

- (7) Emma, ike-nun chamwey-ya?  
mom this-TOP (contrastive) melon-DEC  
'Mom, is this a melon (as contrasted with a picture of a chair next to it.)?' (CK 2:0.10)

It was also observed that the contrastive topic marker (CT) began to be produced at around 1;7 in an extraordinarily early case, as illustrated in (8). A **thematic topic** use of the marker *-nun* is exemplified in (8'a) for a newly introduced referent (not discourse-old) and another exclusive, contrastive **focus** use readily follows in (8'b) below to exclude the fish from the contrastive set created in the context {fish, CK}, which comes via the alternative question 'Is the fish 2 years old or is CK 2 years old?' (see C. Lee 2003):

- (8) Appa hweysa, emma-nun yeki iss-ney.  
Dad company Mom-CT here is-DEC (to my surprise)  
'Dad (is at his) company, Mom is here (to my surprise).' (Y)
- (8') [Pointing to a fish]  
a. I sayngsen-un myet sal-i-ya?  
this fish-TOP how many years be-old-DEC  
'How old is this fish?'  
(Mom: twu sal 'Two years')
- b. Anya. Chwunkyu-ka twu sal-i-ya  
no CK-NOM two years-be-DEC  
'No, I (= CK) am two years old.' (C 1:7.8 Lee's 1986 diary)

Furthermore, it was also noted with surprise that the **focal** additive marker *-to* ('-also') as in (9a) and the **focused** nominative marker *-ka* also begin to emerge at around the same stage, as demonstrated in (9b) and (9d).

- (9) a. Chos-pwul, ikes-to chos-pwul  
Candlelight this-too candlelight  
'A candlelight, this, too, a candlelight.' (Y 1:6.23)

<sup>6</sup> This contrastive usage of the topic marker *-nun* in Korean is locally 'focal' as well as topical and a few syntacticians label it 'contrastive focus.' Contrastive focus is a sharply distinct independent phenomenon, as shown in C. Lee (2003). For the sake of theoretical consistency and convenience, we will adopt the term 'contrastive topic' throughout this paper.

- b. Nay-ka ha-kkey.  
 I-NOM do-PROMISE (intentional)  
 'I will do it (i.e. Me, not you).' (Y 1:9.4)

*Subjects and nominative markers* As is well known in the literature, at around two years of age children across languages typically produce short utterances in which they frequently drop the subject noun phrase.<sup>7</sup> None of the subjects in this study produced overt topics/subjects (with no markers) until around 1:7. JK, for example, was observed to start using topics/subjects productively at 1:7.16, although they were frequently associated with just a few referents such as his caretakers (*appa* 'daddy,' *emma* 'mommy') or physically present objects expressed by a demonstrative (e.g. *i-ke* 'this,' *ce-ke* 'that'). An analysis of JK's and JW's speech collected at 1:8, 2:0, 2:4, 2:8 and 3:0 indicates that they produced overt topics/subjects with a higher rate as they got older, JK from 13% to 37%–53%, and JW from 28% to 32%–52%.

It was observed, overall, that the frequency of overt subjects in the speech of our subjects did not increase with age. AL and AN, for example, produced a subject overtly 65%–81% of the time at 2:3–2:6 and 56%–66% on average at 3:0–3:4. This rate of overt topics/subjects in AL and AN is higher than the rate of overt topics/subjects observed in Clancy's subjects H (1:11–2:9) and W (2:0–2:7), which gradually increased over time from 38.8% up to 51.2% of the time (Clancy, 1984). This developmental trend was not observed in our subjects, although an increase would seem to be expected. Note that AL, for example, produced an overt topic/subject consistently more than 65% of the time at 2:3–2:6.

It was observed in Zoh (1982c) and Kim (2000) that the subject marker was produced at around 1:8–2:0, earlier by a month than the topic marker. But the topic marker was produced one month earlier in S. Kim's (1990) data, and was also a little earlier (Y 1:7) in the data reported in C. Lee (2001). One of our subjects, AL (2:2–2:9), however, produced no nominative or topic markers in her spontaneous speech, although she was able to imitate ones that appeared in her mother's speech. There were twenty instances of imitations in her recorded speech, five of which involved case markers. She imitated an accusative marker twice and nominative markers three times. Regarding the issue of how early topic and nominative markers are acquired and which is acquired first, there seems to be individual variation.

<sup>7</sup> Although it is not so clear what the grammatical status of missing and overt arguments at the early stage is, and it is far more likely to be a topic, we are, for the sake of discussion and convenience, going to tentatively employ the terms subject or topic/subject and object to refer to overt and null (implicit) nominals in pre- and post-verbal position.

Both PL (2:7–3:2) and AN (2:10–3:6) were observed to use the nominative marker from the beginning of the study. Out of a total of 345 multi-word utterances containing a subject, PL produced nominative markers, some of which were not correctly used, 107 times (31%). Ninety-eight occurrences involved only spontaneous marker usage, excluding imitated markers (98/345=28.4%, 98/107=91.6%). Out of a total of 670 multi-word utterances containing a subject, AN produced the nominative marker 238 times (35.5%), of which 218 uses were spontaneous (218/670=32.5%, 218/238=91.6%).

An overall examination of four children (B 1:10–2:2, C 2:0–2:2, AL 2:3–2:6, AN 3:0–3:4) indicates that the production of overt and null topics/subjects largely conforms to the pragmatic condition that old information is likely to be omitted while newly introduced information is required to be given in speech. On the whole, these children produced overt topics/subjects in conveying newly introduced information (70% of overt topics/subjects were new, 30% old), while largely dropping topics/subjects in encoding old information (91%). Sensitivity to topicality was also discovered in the speech of JK (2:0–3:0) and JW (2:0–3:0), who expressed old referents more frequently with null topics/subjects than with overt topics/subjects (67%–90% vs. 10%–33% in JK's speech and 58%–84% vs. 16%–42% in JW's). (For details on these figures, see Cho, 1994, 2001.)

As shown earlier in (9b), as well as in (9d) below, the exclusive **focus** use, a marked use, of the nominative marker is earlier in development than its neutral (or even unaccusative) subject use, as in (9c) below:

- (9) c. Papo-i-ya, appa  
fool-be-DEC dad  
'You are a fool, Dad.' (SK 1:7.21 Lee, 1993)
- d. Appa-ka papo-i-ya  
dad-NOM fool-be-DEC  
'Dad, YOU are a fool.' (SK 1:11.10 in Lee, 1993)
- e. Emma, an pi-ka w-a  
mom not rain-NOM come-DEC  
'Mommy, it is not raining.' (C 2:1.4 in Lee, 1993)

In (9c), the bare nominal *appa* 'Dad' is a typical use of a nominal as a topic, the predicate being an individual level one, requiring a thematic topic. A kinship term applied to the hearer is equivalent to the second person in Korean. Three months later, in (9d), an exclusive or contrastive **focus** use of the nominative marker *-ka* is mobilized to exclude the alternatives, e.g. the speaker SK herself, from the contrastive set, and is equivalent to 'YOU are a fool, NOT ME.' If she had used the NOM marker in the topic sense, it would have been an error; *-nun* must be used for that purpose, and would not have fitted the context. If she had used the NOM marker in the neutral sense, that would also have been an error;



the individual level predicate requires a topic and the only possible context for a neutral NOM-marked individual-level predicate sentence is a complex sentence in which (9d) is embedded, e.g. *na-nun appa-ka papo-i-ra-nun kes-ul kkaytal-ass-ta* 'I realized that Dad was/is a fool' (see C. Lee, 1996).

### *Development of morpho-syntactic constraints*

At around 2;0, Korean children begin to produce modal suffixes incorrectly, as exemplified in (10).

- (10) a. \*Na kki. (stem *kki*- 'insert' not possible without TENSE-DEC)  
           I put  
           'I (will) put (something) (somewhere).' (AL 2:2)
- b. \*Mas-iss-ni-ta. (*Mas-iss-ni*?)  
           taste-exist-QUE-DEC  
           'Is (something) tasty?' (B 1:10)
- c. \*Mary-ka ha-l-lay. (*-ul-lay* not possible with third person)  
           Mary-NOM do-GLOSS?-DEC (intention)  
           'Mary will do (it).' (P 2:6)

Sentence (10a) involves an under-generalization in which a declarative mood (e.g. *-e*, *-lay*, etc.) is absent. In (10b), the child over-generalized the modal suffix by erroneously using both the interrogative (*'-ni'*) and the declarative suffix (*'-ta'*). In (10c), the third person subject ('Mary') co-occurs with the first person modal suffix (*'-ul-ay'*), violating subject-mood agreement (see C. Lee, 1993 for further examples).

As briefly mentioned earlier, a subject can optionally drop in Korean in cases where its person (e.g. first (*'I'*), third (*'Mary'*)) is predictable from the modals affixed on the verb in the same clause. This suggests that subject drop in Korean may be motivated by a grammatical condition such as subject-modal agreement. In this section we are concerned particularly with whether or not the production of null topics/subjects by Korean children relates to changes in their marking of verb modality. We examined two children, JK (2;0–3;0) and JW (2;0–3;0), to investigate whether they were likely to omit topics/subjects in contexts where the verbs contain a person-specific modal suffix at age 2;0 and thereafter.

It was found that the frequency of person-specific modals in null topic/subject utterances in the speech of both JK and JW increased over the first three time points from 2;0 to 2;8 (from 14% to 39%) and then decreased somewhat (to 28%) from 2;8 to 3;0. It was also observed that both children began to use person-specific modals encoding intention, request, and proposal more frequently at age 2;8 than at earlier time points (3% vs. 18%–30% in JK and 5%–12% vs. 19% in JW). Moreover, it was discovered that the rate of person-sensitive suffixes was,

on the whole, higher in null subject utterances than in overt subject utterances across time points in the speech of both children (29% vs. 14%). The chi-square statistic was computed for each child to determine whether the frequencies of person-neutral and person-specific modal suffixes were the same regardless of the type of utterance they occurred in (with overt or null topics/subjects). It was found that the frequencies of verb modality were likely to be related to the type of topics/subjects for JK ( $X^2 = 7.58$ ,  $df = 1$ ,  $p < .01$ ) and JW ( $X^2 = 36.92$ ,  $df = 1$ ,  $p < .001$ ).

### Summary and conclusions

We have seen in this paper that Korean children start acquiring the topic marker *-(n)un* conveying contrastive meaning and the subject marker *-ka* conveying focus at around 20 months of age. The rare occurrences of the thematic/topical use of these topic markers in the children's spontaneous speech are likely to be related to the observation that children are extremely sensitive to topicality as seen earlier, correctly and easily dropping nominals that convey old information from the earliest phase on. The most unmarked topic, realized as a null topic with no nominal and no marker, is acquired earliest, as is probably the case in all languages. Then, bare nominals with no markers are acquired to draw attention to the referents that the child has in mind. These bare nominals still function as topics/subjects, which must be why the thematic/topical use of *-nun* is acquired somewhat late. If a topic marker or subject/nominative marker is to be used, it must be distinct from the bare nominals in function; children are sensitive to the markers' marked functions, i.e. the contrastive topic function for the topic marker and the exclusive focus function for the nominative marker. This phenomenon must fit an optimality-theoretic account. Given children's precocious awareness of topicality, it would not be unreasonable to conclude that Korean children are learning to understand and produce both topical and focal or contrastive functions of the topic markers and both the focus and neutral uses of the subject marker before age 2:0.

In our data there is also support for the possibility that 2-year-old Korean children gradually become sensitive to the connection between modal suffixes and the subjects that agree with them. As reported in C. Lee (1993), modality markers generally appear right after two years of age, whereas basic moods or sentence types occur earlier. The primacy of mood over modality is evident and the imperative (request) is earliest. The relation between the imperative mood and various deontic modal markers explains erroneous uses of mixed modality and misapplication of modality to the wrong person in early child Korean. One important feature of the acquisition of modal suffixes in younger children seems to be early production of imperative, request, and desiderative modal suffixes, in particular, with sensitivity to subject-modal agreement emerging gradually at later time points.

In conclusion, we can speculate that, along the lines suggested in Rispoli (1995: 345), pragmatics and speech act morphemes are possible sources of information for construction of the relation between the meaning of predicates (e.g. pushing, being affected, etc.) and the discourse-pragmatic and morpho-syntactic rules encoded by a subject (e.g. agent, patient, etc.). As seen throughout this paper, Korean involves both pragmatic and grammatical constraints on dropping nominals. Since children begin by talking about various real-life scenes primarily socio-pragmatically (Rispoli, 1991b; Tomasello and Brooks, 1999; Snow, 1999), we propose that the effects of language typology are both pragmatic and grammatical (though pragmatically motivated, as in modal–subject agreement), and that the role of pragmatics must be explored in the emergence of topic and subject nominals and markers.