

Corpus-based analysis of the usage of Korean markers *-(n)un* and *-i/ka* in editorial texts

Kyoung-Young Kim* †

University of Seoul

Kyoung-Young Kim. 2015. A Corpus-based analysis of the usage of Korean markers *-(n)un* and *-i/ka* in editorial texts. *Language and Information* 19.2, 19-36. The aim of this paper is to investigate the usage of Korean markers *-(n)un* and *-i/ka* in editorial texts focusing on information structure. Noun phrases ending with the markers *-(n)un* and *-i/ka* were annotated semi-automatically using a corpus obtained from an online newspaper. Two important factors to determine the choice of markers were examined with the annotated data: referential givenness/newness and position in a sentence. Referential givenness and newness were adopted as indicators of information structure, topic and focus respectively. In addition to quantitative analysis, qualitative analysis was conducted on the selected data. The results suggest that both the marker *-(n)un* and *-i/ka* could carry a topic and a focus reading. Sentence position also played a crucial role in determining the marker, and the marker *-i/ka* was used more frequently in a later position of a sentence than the marker *-(n)un*. (University of Seoul)

Key words: Korean marker *-(n)un*, *-i/ka*, topic, focus, corpus analysis, editorial texts

1. Introduction

Topic and focus are two main components in information structure which respectively characterize the information that we are already familiar with and that we are supposed to gain during a discourse. Erteschik-Shir (1997) posits a topic as "What the statement is about" and "What is used to invoke knowledge in the possession of an audience." In other words, new information – focus – is encoded about given information – topic. A great deal

* I would like to express my thanks to three anonymous reviewers for their comments and suggestions. Any remaining errors are solely mine.

† College English Center, University of Seoul. 163, Seoulsiripdae-ro, Dongdaemun-gu, Seoul, 02504 KOREA. E-mail: kyoungyoung@gmail.com

of work has suggested that the Korean marker *-(n)un* shows unique behavior among other markers. It has been generally termed as a topic marker which connotes that it leads a topic reading of the phrase it is marking. On the contrary, Korean marker *-i/ka* is known to serve as a case marker, which encodes new information within a discourse. Relatively recently, however, there have been studies pointing out that the marker *-(n)un* is not actually a topic marker (H-W Choi, 1997; C-H Han, 1998; J-S Jun, 2007; K-S Kim, 1990). Apart from topic reading, one other view which has been accepted as the usage of the marker *-(n)un* is a contrastive reading. There have been a wide range of arguments on the topichood, focality, and contrastiveness of the marker *-(n)un*. In addition, some researchers argue that the marker *-i/ka* also encodes a topic in some contexts.

This paper investigates information structure of Korean editorial texts in terms of the usage of markers *-(n)un* and *-i/ka* with a corpus-based approach. Editorial texts were collected from an online newspaper, and noun phrases ending with the markers were annotated for further analysis. Based on previous studies, referential givenness/newness and position within a sentence were hypothesized as decisive factors to determine the choice of markers. Significance of each factor was verified with statistical analysis. The rest of this paper is laid out as follows: Section 2 describes the terms, topic and focus, which are fundamental to the discussion of information structure, followed by section 3 which surveys previous studies on Korean markers *-(n)un* and *-i/ka*. Then in section 4, corpus annotations and analysis results are discussed. Finally, conclusion follows in section 5.

2. Topic and Focus

Topic and focus have long been recognized as complements to each other concerning information structure. Following Reinhart (1981), Erteschik-Shir (1997) characterizes a topic as "What the statement is about" and "What is used to invoke knowledge in the possession of an audience." This notion of aboutness is defined as one of the most important characteristics of topic (Gundel, 1988; Lambrecht, 1994; Vallduví, 1992). In Gendel and Fretheim (2004), two types of givenness/newness are presented: referential and relational. Referential givenness/newness is expressed regarding "A relation between a linguistic expression and a corresponding non-linguistic entity in the speaker/hearer's mind, the discourse or some real or possible world." Relational givenness/newness, on the other hand, involves two complementary parts within a sentence, "X: what the sentence is about" and "Y: what is predicated about." The part Y is new in relation to the part X in the sense that "It is new information that is asserted or questioned about X."

- (1) a. Who called?
b. Pat said SHE called.
- (2) a. Did you order the chicken or the pork?

b. It was the PORK that I ordered.

In the sentence (1b), SHE is relationally new in this conversation so that receives focus. PORK in the sentence (2b) as well is new in relation to the topic, ‘what is ordered.’ This relationally new component SHE in the sentence (1b) is referentially given in the sense that the referent is already familiar or activated in the hearer’s mind. Likewise, as PORK has already been mentioned by the speaker who asked the question, it is referentially given although it is relationally new in the sentence. Topic and focus have been distinguished with this relational givenness/newness concept in many studies with different terms: logical subject - predicate (Chao, 1968), presupposition - focus (Chomsky, 1971; Jackendoff, 1972), topic - comment (Gundel 1974, 1988), theme- rheme (Vallduví, 1992), and topic - predicate (Erteschik-Shir, 1997).

3. Earlier studies on Korean information structure

Topichood, focality, and contrastiveness of the marker *-(n)un* have been the focus of many studies on Korean linguistics. Although it has been called a topic marker predominately to be assumed to carry a topic reading (H-K Wee, 1998; M-K Kim, 2001; C-S Oh, 2009; C-M Lee, 2003; K-H Kim, 2007), there also have been arguments that it carries a focus reading instead of a topic reading (K-S Kim, 1990). About topicality and focality, H-K Wee (2010) argues that focality and topicality are not complementary and therefore could contain each other. As another different view, J-S Jun (2007) argues that the marker *-(n)un* is not a unique topic marker in Korean. That is, topics are not exclusively marked by *-(n)un*, and *-(n)un* does not always mark topics. He suggests that the marker *-(n)un* is morphosyntactically neutral and semantically empty. According to Szabolsci (1981), a sentence with a contrastive topic means that the claim in the sentence “need not be true of something else,” whereas the claim in the sentence with a contrastive focus is “not true of anything else.” Following Szabolsci’s (1981) notion of contrastiveness, H-W Choi (1997) states that a contrastive reading is carried with the use of the marker *-(n)un* either as a focus or a topic. She points out that the contrast marker *-(n)un* carries a presupposition of existence of other comparable entities in the discourse. She further suggests that topichood, which is previously believed to be encoded by the marker, is actually encoded by scrambling instead of by the marker *-(n)un* itself. Hence, the marker *-(n)un* in base position only carries contrastiveness irrespective of topichood according to her hypothesis, and the degree of topichood is increased continuously from base position to sentence-initial position as shown in sentences (3).

- (3) a. 메리가 어제 존은 만났다.
 meylika ecey conun mannassta.
 Mary-ka yesterday John-un meet-Pst-Dcl
 ‘Mary met John yesterday (but nobody else).’ [contrastive focus]

- b. 메리가 존은 어제 만났다.
 meylika conun ecey mannassta.
 Mary-ka John-un yesterday meet-Pst-Dcl
 'Mary met John yesterday (and Bill today).' [contrastive topic]
- c. 존은 메리가 어제 만났다.
 conun meylika ecey mannassta.
 John-un Mary-ka yesterday meet-Pst-Dcl
 'As for John, Mary met him yesterday (and as for Bill, Jane met him today).'
 [sentence topic]

C-H Han (1998) also categorizes the usage of the marker $-(n)un$ into topic, contrastive topic, and contrastive focus reading. She shows the difference in meaning syntactically: at S-structure, a VP-external $-(n)un$ -marked-NP carries a topic reading or a contrastive topic reading, while a VP-internal $-(n)un$ -marked-NP carries a contrastive focus reading. On her account, noun phrases *John-un* in both the sentences (3a) and (3b) carry a contrastive focus reading as they are in VP-internal position. A scrambled VP-internal $-(n)un$ -marked-NP is classified as a sentence topic as in H-W Choi (1997)'s study.

- (4) a. [IP 존이 [VP 메리는 좋아한다.]]
 coni meylinun cohahanta
 John-i (NOM) Mary-nun (CF) like-Pres-Dcl
 'John likes Mary, (but no others).' [contrastive focus]
- b. [IP 존은 [VP 메리를 좋아한다.]]
 conun meylilul cohahanta
 John-un (Top) Mary-Acc like-Pres-Dcl
 'John likes Mary.' [sentence topic]
 'John likes Mary, (Frank likes Susan, and Peter likes Laura.)' [contrastive topic]
- c. [IP 메리는 [VP 존이 좋아한다.]]
 meylinun coni cohahanta
 Mary-nun (Top) John-i (Nom) like-Pres-Dcl
 'Speaking of Mary, John likes her.' [sentence topic]

While most research working on topics has focused on the marker $-(n)un$, some studies suggest that the marker $-i/ka$ also carries a topic reading in addition to the focus reading which is generally assumed to be the main function of the marker. Y-S Hong (1985) suggests the topic reading of the marker $-i/ka$ with the following example sentences.

- (5) a. 네 차에 무슨일이 일어났니?
 ney chaey mwusunili ilenassni?
 your car-to whichevent-i happen-Pst-Int

‘What happened to your car?’

- b. 내 차는/차가 고장났어.
 nay chanun/chaka kocangnasse.
 my car-nun/car-ka break-Pst-Dcl
 ‘My car broke down.’

In the sentence (5b), the subject noun phrase *nay chanun/chaka* (my car) carries given information in contrast to the following comment *kocangnasse* (broke down), so that it is marked as a topic. Both *-(n)un* and *-i/ka* can serve as a topic marker in this sense. She further argues about contrastiveness and exclusiveness in information structure; in particular, the marker *-(n)un* expresses contrastiveness, while the marker *-i/ka* expresses exclusiveness. In H-P Im (2007), a similar argument of *-i/ka*'s exclusiveness is suggested.

- (6) a. 철수와 영희 중 누가 머리가 좋으냐?
 chelswuwa yenghuy cwung nwuka melika cohunya?
 Chelswu-and Yenghuy-between who smart-Int
 ‘Who is smart, Chelswu or Yenghuy?’
 b. 영희가 머리가 좋다.
 yenghuyka melika cohta.
 Yenghuy-ka smart-Dcl
 ‘Yenghuy is smart (not Chelswu).’

According to him, the subject noun phrase *Yenghuyka* in the sentence (6b) excludes the possibility that ‘Chelswu is also smart,’ so using the marker *-i/ka* leads to exclusiveness of the noun phrase. He especially argues that the noun phrase *Yenghuyka* marks a topic in the sense that it is followed by a comment about it although it is relationally new information. Y-C Jun (2009), on the other hand, disputes this argument based on the claim that topic and focus should be discussed with the relational givenness/newness concept, so relationally new information within a sentence cannot carry a topic reading.

4. Usage of Korean Markers *-(n)un* and *-i/ka* in editorial texts

In this study, the behavior of Korean markers *-(n)un* and *-i/ka* is investigated in view of information structure via corpus-based analysis. In consideration of overall previous studies on Korean markers, information structure of either marker is not predetermined in this study although it is hypothesized that there is an inclination of each marker to be a specific information status in a neutral condition. More specifically, the marker *-(n)un* is more likely to carry a topic reading, while the marker *-i/ka* has a tendency to carry a focus reading. As pointed out in Y-C Jun (2009), indication of topic and focus within a sentence should be

relational givenness/newness which determines what is new information about given information. In the example sentences in section 3, the noun *Yenghuy* in the sentence (6b) is referentially given information in that it is previously mentioned although it is relationally new in its own sentence. However, due to the fundamental limitation of a corpus study that precludes the identification of exact relational givenness/newness throughout the corpus, referential givenness/newness is adopted as the second best property to imply information status. In addition, difference in givenness/newness between referential and relational frequently occurs in a particular conversation like sentences (6), which contain a question asking a choice. As the corpus used in this work includes editorial texts instead of a conversational discourse, it is more reasonable to take referential givenness/newness as an alternative. The relationship between relational givenness/newness and referential givenness/newness is illustrated as below (Y-C Jun, 2009):

(7)

relationally given	relationally new
referentially given	referentially new

As shown, referentially given information is not always relationally given. Referentially new information, on the contrary, could quite safely be categorized as relationally new information, which is a focus within a sentence. In addition to the referential givenness/newness, sentence position of the information with each marker is examined as another factor to determine the use of markers. This factor is based on H-W Choi (1997)'s argument that the degree of topichood increases continuously from base position to sentence-initial position. Overall, following research questions are studied to investigate the usage of markers in editorial texts:

- a. Whether the marker *-(n)un* carries a focus reading in addition to the topic reading
- b. Whether the marker *-i/ka* carries a topic reading in addition to the focus reading
- c. Do referential givenness/newness and position within a sentence play a role as factors to determine the choice of markers between *-(n)un* and *-i/ka*?

4.1 Corpus and annotation

As the present study is based on corpus analysis, comparable amount of data should be collected automatically from websites. Among available data, personal blogs were excluded as they frequently contain ungrammatical sentences and expressions. Newspapers are one of the popular sources of corpus analysis as the texts used in newspapers are representative in grammar and usage. In this study, editorial texts, which present an opinion of the writer on a specific issue, were collected for further analysis as they are more comparable to real conversations than news articles in the sense that the markers are deliberately chosen to express opinions.

Corpora of editorial texts were collected from an online newspaper via web crawling with the use of ‘urllib’ library in Python. The collected corpus contains a total of 1,483 articles of editorial texts dated from Aug. 01 2013 to Mar. 23 2015. Detailed corpus statistics are presented in Table 1.

[Table 1] Corpus Statistics

Documents	Sentences	Word Types	Word Tokens
1,483	34,489	103,346	437,680

To study the research questions posited above, noun phrases whose endings are either *-(n)un* or *-i/ka* within the corpus were annotated as the first step. The annotation was performed semi-automatically following the steps presented below.

A. Collect Unigram Noun Phrases:

1. Unigram words ending with *-(n)un* or *-i/ka* were extracted from the whole corpus using a Python program.
2. Extracted unigram word list was trimmed to exclude the cases with ending *-(n)un* or *-i/ka* which are not the markers focused in this study:
 - a. Extracted words were morphologically analyzed using “Sejong Korean Morphological Analyzer” developed by the Sejong Project (<http://www.sejong.or.kr>).
 - b. Only nouns ending with *(n)un/JX*, *i/JKS*, and *ka/JKS* were collected and passed to the next step. After that, seemingly irrelevant words were screened with the following criteria.
 - Words containing a morpheme with tags /VV, /VA, /VX, /MA*, /E*
 - Examples which were excluded: *쌍/VV+아/EC+가/VX+아는/EC*, *절충/NNG+하/XSV+려는/EC*, *이대로는/MAG*, *물려받/VV+은/ETM*
 - c. Although the performance of the morphological analyzer was quite successful, there were some mistakes in analysis which led to incorrect results (e.g. *생똥맛/NF+은/JX*). In order to make the extracted list cleaner, manual correction was additionally performed as a supplementary step of the automatic extraction.

[Table 2] Part of POS tags generated by Sejong Morphological Analyzer

JX	Auxiliary Particle	EP	Prefinal EM	NNG	General Noun
JKS	Subjective Case Particle	EF	Final EM	NF	Noun-like word
VV	Verb	EC	Conjunctive EM	XSV	Verb DS
VA	Adjective	ETN	Nominalizing EM	MAJ	Conjunctive Adverb

VX	Auxiliary Predicate	EPM	Adnominalizing EM	MAG	General Adverb
----	---------------------	-----	-------------------	-----	----------------

EM: Ending Marker; DS: Derivational Suffix

B. Collect Bigram Noun Phrases:

- As many reasonable noun phrases in Korean are composed of more than a word, the unigrams obtained from the previous step were expanded to bigrams to cover more data.
 - With 2,990 unigram word types ending with the marker *-(n)un* or *-i/ka* from the previous step, bigrams (b1 b2) were extracted from the whole corpus.
- After morphological analysis, irrelevant bigram phrases were excluded.
 - Tags /S* (Punctuations) were added to the criteria used in the unigram screening process.
 - Picked out examples which are not noun phrases:
 - b1 불리/VV+였/EP+지만/EC+./SP b2 국민/NNG+들/XSN+은/JX
 - b1 없/VA+다/EF+./SF b2 체육/NNG+계/XSN+는/JX

C. Collect Trigram Noun Phrases: Step B was repeated with extracted bigram noun phrases to expand the data to cover trigrams.

D. Annotation: The collected unigram, bigram and trigram list were automatically annotated to the whole corpus using a Python program. An excerpt of the annotated corpus is presented in (8). The underlined noun phrases are annotated unigrams and bigrams: unigrams are marked with <u>, and bigrams are marked with <b1> and <b2>.

Table 3 presents statistics of target noun phrases in the whole corpus. Numbers of types and tokens irrespective of the markers are presented in the first row (AM) for each ngram. Then statistics of each marker are presented in the next rows separately.

(8) Annotated corpus example

인천 강화도 캠핑장에서 22일 불이 나 텐트에서 잠자던 가족 등 5명이 숨졌다. <b1>문제의 <b2>캠핑장은 텐트 안에 난방기구와 침구 등 편의시설을 갖췄다. 간편하게 몸만 가면 즐길 수 있는 새로운 야외활동으로 요즘 인기를 끄는 형태다. <u>텐트는 <b1>방염 <b2>처리가 안 된 가연성 소재였다. <u>불꽃이 튀면 확 옮겨붙을 수 있었고, 실제로 2~3분 만에 <u>텐트가 전소됐다고 한다. <b1>텐트 <b2>출입구는 <u>어른이 허리를 숙이고 다녀야 할 정도로 작아, 어두운 밤에 대피로를 찾기 어려웠을 것으로 보인다. <b1>캠핑장의 <b2>확산이 <u>비밀이 아니고 갑작스러운 것도 아닌데 이해하기 어렵다. <b1>이번 <b2>사고는 예견된 것이다. 당국의 대응 실패를 지적하지 않을 수 없다.

[Table 3] Statistics of target noun phrases in the corpus

	unigrams		bigrams		trigrams	
	types	tokens	types	tokens	types	tokens
AM	2,628	11,503	5,511	8,117	259	326
<i>-(n)un</i>	1,840	6,057	3,382	4,980	240	299
<i>-i/ka</i>	1,150	5,446	2,393	3,137	22	27

4.2 Factors to determine the usage of markers

For each noun phrase ending with the marker *-(n)un* or *-i/ka*, candidate factors to determine the marker were extracted using a Python program for further analysis. Examples of extracted data and detailed notations are presented in Table 4. Referential givenness/newness and position in a sentence are hypothesized to be crucial factors to choose each marker as previously discussed. Binary distribution of Initial/Non-Initial and proportional position were marked as factors indicating position of the noun phrase.

To evaluate the overall effect of factors, multiple logistic regression was performed using R program as the response variable (Y) is categorical: *-(n)un* or *-i/ka*. Predictor variables, referential givenness/newness- gn (X1)- and Initial/Non-Initial - ini (X2)-, are also categorical in this test, while proportional position - ppos (X3)- is continuous. Proportional position was calculated by normalizing word position with sentence length (pos/LenS). The formulation of the regression model is shown in (8):

$$(8) \quad \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

where p=probability that Y=1 (*(n)un*), and 1-p=probability that Y=0 (*i/ka*)

[Table 4] Examples of Extracted factors for each noun phrase

date	Marker	gn	SNo	pos	LenS	ini	ppos	word
2015.03.23_1.txt	0	1	S11	1	8	1	0.125	홍 지사
2015.03.23_1.txt	1	1	S12	1	22	1	0.045454545	경남도
2015.03.23_1.txt	0	0	S12	21	22	0	0.954545455	필요
2015.03.23_1.txt	1	1	S14	1	17	1	0.058823529	홍 지사
2015.03.23_1.txt	1	0	S16	5	18	0	0.277777778	아이들의 밥그릇
2015.03.23_1.txt	1	0	S16	8	18	0	0.444444444	자기
2015.03.23_2.txt	0	0	S0	10	13	0	0.769230769	대북전단 갈등

2015.03.23_2.txt	1	0	S2	8	15	0	0.533333333	최악의 사태
2015.03.23_2.txt	1	0	S3	17	35	0	0.485714286	북쪽
2015.03.23_2.txt	0	0	S4	4	6	0	0.666666667	상황
2015.03.23_2.txt	0	0	S5	3	19	0	0.157894737	기간
2015.03.23_2.txt	1	0	S5	14	19	0	0.736842105	대화 통로

Marker: $(n)un$ -1, i/ka -0; **gn:** given-1, new-0; **SNo:** sentence number;
pos: word position in a sentence; **LenS:** No. of words in each sentence;
ini: initial-1, non-initial-0; **ppos:** proportional position (pos/LenS)

[Table 5] R output of multiple logistic regression

	Coefficients:			
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.64425	0.03881	16.599	<2e-16 ***
gn1	0.28426	0.03109	8.911	<2e-16 ***
ini1	0.44606	0.04576	9.749	<2e-16 ***
ppos	-1.40735	0.06658	-21.139	<2e-16 ***

gn1: referential givenness; **ini1:** sentence-initial position; **ppos:** proportional position

As presented in Table 5, relations between all four factors and the choice of the marker are verified to be statistically significant (Pr <0.05). Specifically, referential givenness (gn1) increases the log odds of choosing the marker $(n)un$ by the coefficient estimate 0.28426. Likewise, sentence-initial position (ini1) increases the chance to choose the marker $(n)un$ by 0.44606. On the other hand, as proportional position in a sentence increases, the log odds of choosing the marker $(n)un$ decreases; in other words, it is more likely to use the marker i/ka in the later position in a sentence. More detailed discussion and examples of each factor follow in the next section.

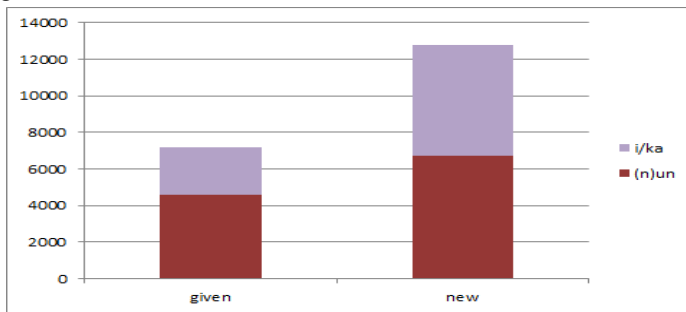
4.2.1 Referential givenness/newness. R output of logistic regression in Table 5 implies that there is a significant relation between referential givenness (gn1) and the marker $(n)un$. Referential newness, on the contrary, doesn't show a significantly different relation with either of the markers $(n)un$ or i/ka . The frequency of choosing each marker depending on givenness/newness is presented in Table 6 and Figure 1. As illustrated, previously discussed statistical interpretation from R output is correspondent to the frequency distribution. In particular, no significant difference in choosing markers is shown when the noun phrase is new information. As referentially new information could quite safely be assumed as relationally new information, this result suggests that both markers $(n)un$ and i/ka can carry a focus reading with no significant preference. This suggestion is in line with recent studies

arguing that the marker *-(n)un* is not a topic marker but morphosyntactically neutral marker carrying either a topic or a focus reading. In the case of topic reading, the distribution shows a remarkable difference between the two markers when the noun phrase is referentially given information; almost two times more frequently use the marker *-(n)un* than *-i/ka*. However, the result still suggests that the marker *-(n)un* is not an exclusive topic marker as *-i/ka* as well could carry a topic reading with less frequency.

[Table 6] Distribution of each marker occurrences depending on the referential givenness/newness

	given	new
<i>(n)un</i>	4,618	6,718
<i>i/ka</i>	2,540	6,070

[Figure 1] Schematic representation of each marker occurrences depending on the referential givenness/newness



Below are selected examples from the corpus to further analyze the usage difference of the markers in view of information structure. English translations were provided from the English edition of the corpus. Sentences (9) and (10) contain examples of noun phrases ending with the marker *-(n)un*. The noun phrases in sentences (9) are referentially given information so that they are sentence topics followed by comments about the topics, which has long been admitted as one of the typical usages of the marker *-(n)un*. On the contrary, the noun phrases in sentences (10) are referentially new information although used with the marker *-(n)un*. Therefore, the sentences as a whole could be read as new information, which is parallel to the presentational reading induced by the marker *-i/ka* in the sentence (11). According to H-Y Choi (1997), the sentence (11) containing the marker *-i/ka* could be read either as presentational or focus, and all the information in the presentational reading is focus which is new. Examples from sentences (10) suggest that the marker *-(n)un* as well as the marker *-i/ka* could carry this presentational reading.

- (9) a. 일본은 역사 문제가 과거형이 아니라 현재형임을 직시해야 한다.
 ilponun yeksa mwunceyka kwakehyengi anila hyencayhyengimul ciksihayya hanta
 'Japan (un) needs to face up to the fact that these historical issues are not just part of
 the past, but continue to affect the present.'
- b. 원래 김영란법은 '부정청탁 및 금품수수 금지'
 wenlay kimyenglanpepun pwucengchengthak mich kumphwumswuswu kumci'
 부분과 '이해충돌 방지' 부분이 함께 시행되도록
 pwupwunkwa 'ihaychwungtol pangci' pwupwuni hamkey sihayngtoytolok
 설계됐다.
 selkyeytwayssta.
 'In its original conception, the KimYoungran Act (un) was supposed to not only ban
 illicit requests and bribes but also to block conflict of interest.'
- (10) a. 류길재 통일부 장관은 마지막 회담을 제의하면서 협상을
 lyukilcay thongilpwu cangkwanun macimak hoytamul ceyuyhamyense hyepsangul
 하자는 건지 최후통첩인지를 분간하기 힘든 강경 자세를
 hacanun kenci choyhwuthongchepincilul pwunkanhaki himtun kangkyeng caseylul
 보였다.
 poyessta
 'In making this "final offer" for talks, Unification Minister Ryoo Kihl-jae (un) was so
 rigid that it was tough to tell whether he was suggesting discussions or issuing an
 ultimatum.'
- b. 실제로 금강산관광 재개 문제는 정부가 마음만
 silceylo kumkangsankwankwang caykay mwunceynun cengpwuka maumman
 먹으면 쉽게 타결할 수 있다.
 mekumyen swipkey thakyelhal swu issta
 'The Mt. Keumgang tourism issue (nun) would be easy enough to resolve if the
 government put its collective mind to it.'
- c. 이번 사건의 진위는 앞으로 법정에서 가려질 것이다.
 ipen sakenuy cinwinun aphulo pepcengeyse kalyecil kesita
 'The truth of the case (nun) will be determined in court.'
- (11) 메리가 존을 만나고 있다.
 meylika conul mannako issta.
 Mary-Nom John-Acc meet-Inf be-Dcl
 'Mary is meeting John. [presentational]
 'It is Mary who is meeting John' [focus]

Further examples of using $-(n)un$ for new information are related to the contrastiveness carried by the marker. Both noun phrases ending with $-(n)un$ in the sentence (12a) are referentially new information, and they contrast with each other in that each noun phrase emphasizes the subject of each clause which acts in a specific way. While the noun phrases

in the sentence (12a) are in VP-external position, those in the sentence (12b) are in VP-internal position. C-H Han (1998) argues that *-(n)un* carries a topic reading when it marks a VP-external noun phrase, while it is a focus marker when it marks a VP-internal noun phrase. Although the examples in this study suggest not all VP-external noun phrases ending with *-(n)un* carry a topic reading, it is remarkable that VP-internal noun phrases are highly likely to carry a focus reading as shown in the sentence (12b).

(12) a. 이런 탓에 김대중 대통령은 베트남 방문 당시 호찌민
 ilen thasey kimtaycwung taythonglyengun peythunam pangmwun tangsi hoccimin
 묘소에 헌화하고 정상회담에서 “베트남 국민에게 본의 아니게
 myosoey henhwahako cengsanghoytameyse “peythunam kwukmineykey ponuy anikey
 고통을 준 데 미안하게 생각한다”고 사과했고, 노 대통령은
 kothonggul cwun tey mianhakey sayngkakhanta”ko sakwahayssko, no taythonglyengun
 호찌민 묘소를 헌화·참배하고 “우리 국민에게 마음의 빛이 있다”고
 hoccimin myosolul henhwa champayhako “wuli kwukmineykey maumuy pici issa”ko
 더욱 수위를 높였다.
 tewuk swuwilul nophyessta.

‘The situation prompted Kim Dae-jung (un) to give an offering of flowers at Ho Chi Minh Cemetery during his visit and say that he felt “sorry for the unintended suffering caused to the Vietnamese people.” Roh Moo-hyun (un) went further, offering flowers and paying respects at the same cemetery and saying that a “debt exists in the hearts of the South Korean people.”

b. 이 발언은 야당 대표와의 회담 의제는 ‘민생 법안 처리’ 등으로
 i palenun yatang tayphyowauy hoytam uyceynun ‘minsayng pepan cheli’ tungulo
 한정하고, 회담 형식도 야당 대표와의 단독 회동은
 hancenghako, hoytam hyengsikto yatang tayphyowauy tantok hoytongun
 거부한다는 뜻으로 읽힌다.
 kepwuhantanun ttusulo ilkhinta.

‘This can be seen as limiting the agenda of the talks (nun) to legislation related to public welfare and as rejecting one-on-one talks (un) with the leader of the opposition party.’

Although two-thirds of given information in this corpus are carried by the marker *-(n)un*, the marker *-i/ka* also carries given information. The underlined noun phrases ending with the marker *-i/ka* in sentences (13) are referentially given so that carry a topic reading. By using the marker *-i/ka* instead of *-(n)un*, focality is added to the noun phrases in addition to the topicality it has already contained.

- (13) a. 그 가운데에서도 위안부 문제는 해결이 시급하다.
 ku kawunteyeyseto wianpwu mwunceynun haykyeli sikuphata.
 'Addressing (i) the comfort women issue is an urgent part of that.'
- b. 통령은 저 높은 데서 '국정'을 논하고 야당은 여당과만
 taythonglyengun ce nophun teyse 'kwukceng'ul nonhako yatangun yetangkwan
 경쟁하는 식이 돼선 야당이 대안 정치세력으로 설 수가 없다.
 kyengcaynghanun siki twaysen yatangi tayan cengchiseylyekulo sel swuka epsta
 'The opposition (i) is never going to be viable political alternative when it's squabbling
 with the ruling party while the President perches up above the fray talking about
 "governance."'

Sentences (14) present examples of using the marker *-i/ka* for given information which is part of new information. The sentence (14a) contains a relative clause modifying the noun phrase *kim silcangi* (Kim) so that the given noun phrase is changed into new information. In sentences (14b) and (14c), the noun phrases ending with *-i/ka* are subjects of embedded clauses, and the whole embedded clauses are new information carrying a focus. Likewise, the sentence (14d) contains the noun phrase *nocoka* (union), which is the subject of the embedded clause. The VP-internal noun phrase of the relative clause *haycikcauy cohapwen ciwinun* (what status terminated workers should hold) is also new information, which makes the whole sentence new information.

- (14) a. 남북정상회담과 엔엘엘 문제의 진실을 누구보다 잘 아는
 nampwukcengsanghoytamkwa eyneyleyl mwunceyuy cinsilul nwukwupota cal anun
 김 실장이 오랜 침묵을 깨고 중요한 증언을 하고 나선 것이다.
 kim silcangi olayn chimmwukul kkayko cwungyohan cungenul hako nasen kesita
 'Kim (i), who knows the truth of the inter-Korean summit meeting and the NLL issue
 better than anyone, broke his long silence to offer this important testimony.'
- b. 민주당은 원내외 병행투쟁을 강조했지만 국회가
 mincwutangun wennayoy pyenghayngthwucayngul kangcohayssciman kwukhoyka
 열리면 정국의 무게중심은 원내로 쏠릴 수밖에 없다.
 yellimyen cengkwukuy mwukeycwungsimun wennaylo ssollil swupakkey epsta.
 'While the Democratic Party (DP) insisted that it would continue its efforts both inside
 and outside of the National Assembly, the focus is bound to shift in that direction once
 the assembly (ka) is in session.'
- c. 야당이 지금처럼 여당 앞에서 맥을 못 추던 시절은 별로 없었다.
 yatangi cikumchelem yetang apheyse maykul mos chwuten sicehun pyello epsessta
 'It is indeed rare for an opposition party (i) to be as powerless in checking the ruling
 party as the DP is today.'
- d. 게다가 해직자의 조합원 지위는 노조가 재량껏 자율적으로 결정하는
 keytaka haycikcauy cohapwen ciwinun nocoka caylyangkkes cayulcekulo kyelcenghanun

것이 세계 여러 나라의 관행이다.
kesi seykyey yele nalaury kwanhayngita
'The practice around the world has been to allow unions (ka) to decide for themselves
what status terminated workers should hold (nun).'

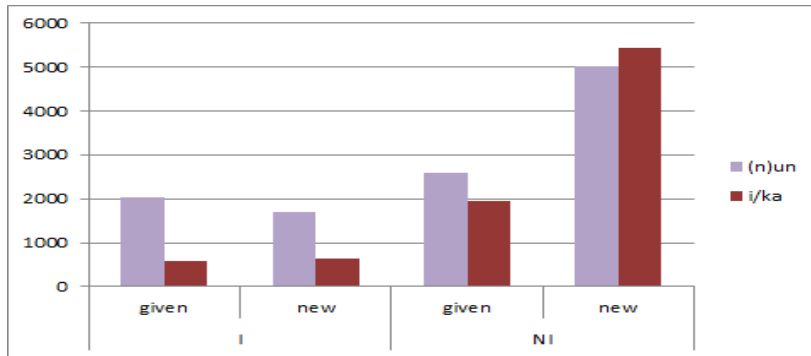
4.2.2 Position within a sentence. Previous studies suggest that position within a sentence has a significant effect on determining information structure of the words/phrases: the degree of topichood increases continuously from base position to sentence-initial position (H-W Choi, 1997). From this argument, it is hypothesized that the usage of markers is also dependent on the position within a sentence as the markers and information structure are hypothesized to be significantly related. Most of all, the initial position is expected to be highly inclined to the marker *-(n)un* rather than *-i/ka*. With this assumption, the frequency distribution of each marker depending on position was analyzed in two ways: binary distribution of Initial/Non-Initial and proportional position.

[Table 7] Distribution of each marker occurrences depending on the position (Initial/Non-Initial)

	Initial		Non-Initial	
<i>(n)un</i>	3,730		7,606	
<i>i/ka</i>	1,236		7,374	
	given	new	given	New
<i>(n)un</i>	2,031	1,699	2,587	5,019
<i>i/ka</i>	591	645	1,949	5,425

As illustrated in Table 7, the marker *-(n)un* occurs three times more frequently than *-i/ka* in sentence-initial position irrespective of givenness/newness of the noun phrase. In non-initial position, on the other hand, the overall frequencies of *-(n)un* and *-i/ka* are not significantly different. What is remarkable in non-initial position is that the marker *-(n)un* is more frequent when the noun phrase is given information, while the marker *-i/ka* is more frequent in case of new information. Pearson's chi-squared test was performed to see if this frequency difference is statistically significant, and the output confirms the significance (X-squared = 101.5917, df = 1, p-value < 2.2e-16).

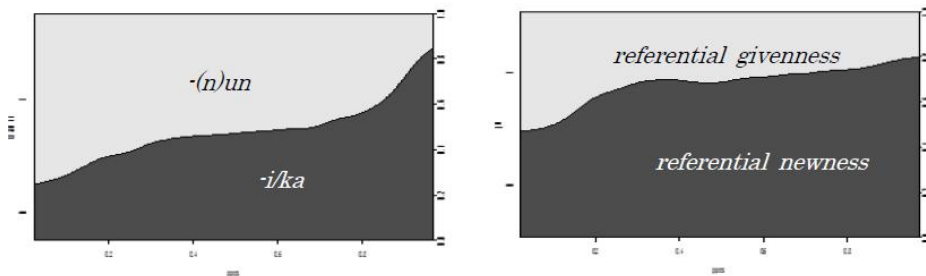
[Figure 2] Schematic representation of each marker occurrences depending on the position and givenness/newness



Proportional position of each word/phrase was calculated based on the number of words within a sentence: $\text{proportional position} = \text{word position} / \text{sentence length}$ (the whole number of words)

Figure 3a illustrates the conditional density plot of choice of markers given proportional position. The probability of the marker 0 (-i/ka) generally increases as proportional position increases. The tendency is more significant in the initial and final position in a sentence. Overall, the plot suggests that the marker -i/ka is more likely to be chosen as a noun phrase occurs in a later position of a sentence. In comparison to Figure 3a, conditional densities of referential givenness/newness given proportional position are presented in Figure 3b. There shows no significant difference between referential givenness and newness in the initial position of a sentence, while the frequency of referentially new information grows in the later position. These results suggest that the sentence position influence on the choice of markers more significantly than the topicality of the noun phrase.

[Figure 3] Conditional density plots given proportional position (ppos)



a. Conditional densities of choice of marker b. Conditional densities of referential givenness/newness

5. Conclusion

The Korean markers *-(n)un* and *-i/ka* selected from editorial texts were analyzed in this paper. The marker *-(n)un* has long been recognized as a topic marker in Korean linguistics, but objection to this idea has been raised by many researchers. In addition, there have been disputes about the information structure of the marker *-i/ka* and whether it can carry a topic reading as well. Statistical results with marker-annotated corpus verify that both *-(n)un* and *-i/ka* carry a topic and a focus reading. However, detailed contexts will influence the choice of each marker in the same information structure. As another factor, noun phrase's position in a sentence is shown to be significant in choosing the marker irrespective of its givenness/newness. Sentence-initial position is much more inclined to the marker *-(n)un* than *-i/ka*. Moreover, proportional position and the usage of marker are related linearly; the later in a sentence, the higher tendency to choose *-i/ka* as a marker. In addition to statistical verification, further analyses of contexts with selected examples were provided. The qualitative analyses show that referential givenness/newness and position in a sentence are fundamental starting points for further study with detailed grammatical and pragmatic contexts decisive factors in determining the markers.

<References>

- Choi, Hye-Won. 1997. Topic and Focus in Korean: The information partition by phrase structure and morphology. *Japanese/Korean Linguistics* 6: 545-561
- Gundel, Jeanette K. 1974. *The Role of Topic and Comment in Linguistic Theory*. Ph.D. Dissertation, University of Texas at Austin. Published by Garland, 1989.
- Gundel, Jeanette K. 1988. Universals of topic-comment structure. In M. Hammond, E. Moravczik and J. Wirth (eds.), *Studies in Syntactic Typology*: 209-239.
- Gundel, Jeanette K., Nancy Jedberg, and Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69: 274-307.
- Gundel, Jeanette K. and Thorstein Fretheim. 2004. Topic and Focus. In *The Handbook of Pragmatic Theory*: 174-196 Laurence Horn and Gregory Ward (eds.). Oxford/Cambridge, MA: Blackwell.
- Han, Chung-Hye. 1998. Asymmetry in the interpretation of *-(n)un* in Korean. *Japanese/Korean Linguistics* 7: 1-15. CSLI Stanford.
- Hong, Yun-Sook. 1985. A quantitative analysis of KA/I and (N)UN as subject markers in spoken and written Korean. In S. Kuno, et al. (eds.), *Harvard Studies in Korean linguistics-proceedings of the 1985 Harvard workshop on Korean linguistics (Harvard WOKL-1985)*: 145-153. Hanshin.
- Im, Hong-Pin. 2007. *Discourse-pragmatic notion of topic and syntactic analysis in Korea : a new*

- approach to the concept of topic and its applications to Korean syntax. Seoul National University Press.
- Jun, Jong-Sup. 2007. Korean Has No Unique Topic Marker. *Language and Linguistics* 39: 199-218.
- Jun, Young-Chul. 2009. On 'i/ka' as a Topic Marker. *Discourse and Cognition* 16.3: 217-238
- Kim, Kwang-Sup. 1990. Where Do Contrastive and Focus Readings Come From? *Japanese/Korean Linguistics* 1: 395-412.
- Kim, Kyu-Hyun. 2007. The Korean Topic Marker nun Revisited: nun as a Tying Device. *Japanese/Korean Linguistics* 15: 81-92.
- Kim, Mi-Kyung. 2001. Informational Status of Contrastive nun in Korean. *Korean Journal of Linguistics* 26.1: 21-38.
- Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus and the mental representations of discourse referents*. Cambridge, U.K.: Cambridge Press.
- Lee, Chung-Min. 2003. "Contrastive (predicate) topic vs. contrastive focus: with special reference to scale semantics and alternative semantics." In Gregory K. Iverson and Sang-Cheol Ahn (eds.), *Explorations in Korean Language and Linguistics: 277-291*. Seoul: Hankook Publishing Co.
- Oh, Chi-Sung. 2009. -Un/Nun as an Unratified Topic Marker in Korean. *Korean Journal of Linguistics* 34(3): 603-624
- Reinhart, Tanya. 1981. Pragmatics and Linguistics: An analysis of sentence topics. *Philosophica* 27: 53-94
- Szabolcsi, Anna. 1981. The semantics of topic-focus articulation. In *Formal Methods in the Study of Language: 513-540*. ed. by T.M.V. Hanssen & M.B.J. Stokhoj. Amsterdam: Mathematisch Centrum.
- Wee, Hae-Kyung. 1998. Semantics and Pragmatics of Contrastive Topic in Korean and English. *Japanese/Korean Linguistics* 7: 549-565
- Wee, Hae-Kyung. 2010. Topicality and Focality of Contrastive Topic. *Language and Information* 14.2: 47-70.

Submitted on: May 29, 2015

Revised on: November 15, 2015

Accepted on: November 16, 2015