The Prosody of Contrastive Topic in Korean

SEUNG-EUN KIM
Cornell University

1 Introduction

This paper investigates how Contrastive Topic (CT) is prosodically realized in Korean, and also examines how the deletion of the postnominal particle =NUN interacts with prosody. I compare the prosodic pattern of CT with that of non-contrastive Topic (T) and Focus (F). I show that CT has a prosodic pattern distinct from T and F in the region where CT, T, or F occurs; yet, it is similar to F in the region following the CT as both show post-peak compression. In addition, I show that the particle =NUN is not necessary to represent the meaning of CT, and the deletion of =NUN affects the prosodic pattern of a CT phrase.

CT is defined as a phrase denoting what the question being addressed is about but implies questions about different topics (Constant 2014). Below is an example of CT adapted from Constant (2014).

(1) *A and B are at the potluck party. A showed up late, so he is asking B who brought what.*
A: What about Persephone and Antonio? What did *they* bring?
B: *Persephone* brought the gazpacho.

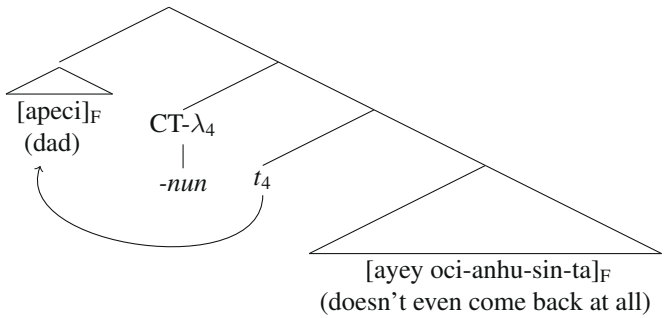
Japanese/Korean Linguistics 27.
Edited by Michael Barrie.

In this example, *Persephone* in B's response answers A's question, but at the same time it leaves another topic to be discussed (i.e., *Antonio*); *Persephone* is CT. In the semantic literature, CT has been discussed in relation to both T and F. Since CT answers the question being addressed (although partially), it is related to T. CT is also often compared to F because it is used to mark the presence of alternatives.

While researchers differ on how they view CT in relation to T and F, the present study follows the semantic perspective which views CT as distinct from T but similar to F (e.g., Constant 2014, Tomioka 2010). According to this view, CT and F are similar since they share the same semantic value as being F-marked; both mark the presence of alternatives. However, they undergo different semantic computations so that each achieves its own meaning. For instance, Constant (2014) argued that CT is associated with a semantic operator, CT-λ, which does not associate with F. (2) illustrates Constant (2014)'s analysis of CT in Mandarin Chinese, which is translated in Korean. CT *apeci* is marked as F, but it raises overtly to the position directly above the CT-λ, associates with it, and is thereby interpreted as CT. Although F phrases are also marked as F, they do not associate with CT-λ.

(2) *Every day, mom doesn't come home until late.*

Apeci=nun ayey oci-anhu-sin-ta
Dad=CT at all come back-NEG-HON-DCL
'(And) dad doesn't even come back at all.'



A semantic connection between CT, T, and F is reflected in their prosodic realization. Büring (2003), who argued that CT and F are semantically distinct, maintained that CT is realized as L+H*L-H% while F is realized as H*L-L%. This argument has its roots in Jackendoff (1972)'s distinction between A- (fall) and B-accent (fall-rise). Constant (2014), on the other hand, argued that the prosody of CT can be described only with a boundary tone (L-H%), and this is the realization of the semantic operator CT-λ. Contrary to

Büring (2003) and Constant (2014) who focused on the prosody only at the CT phrase, Tomioka (2010) examined prosodic patterns in phrases following the CT. He found that in Japanese, CT exhibits post-peak compression, which is also found in F. Tomioka takes this property to support his view that CT and F are semantically similar. As we see below, Korean and Japanese share this similarity between CT and F.

Although previous studies on the prosodic realization of CT have argued that the semantics of CT is closely related to its prosody, several problems remain. First, the previous studies attempted to clarify the prosodic pattern of CT based solely on introspective evidence. No experimental data were provided to support the authors' claims. Second, studies differed as to whether they examined the prosody of the CT phrase itself or the phrases following the CT. This may cause different results in comparing the prosody of CT with T and F. Lastly, not many studies explained specifically how the prosody of CT is different from that of T. Researchers argue that CT and T are semantically distinct in that CT contains the additional meaning of contrast or marking alternatives, and accordingly, they argue that their prosodic patterns are different; however, except for Lee (2008), no studies actually explained *how* the prosodic patterns of CT and T are different nor showed their difference experimentally.

In order to resolve these problems, I conducted a production experiment to examine the prosody of CT, specifically in Korean. I compared the prosody of CT with that of T and F both at the region where CT, T, or F occurs and the region following them. CT is marked morphologically and prosodically in Korean: CT is followed by the postnominal particle =NUN, and it is argued to be realized with a high tone ((L)H*) or Jackendoff (1972)'s B-accent (Lee 1999, 2003, 2008). More specifically, Lee (2008) argued that the intonational rise starts from the L tone in the final syllable of a CT phrase and reaches H* at the particle =NUN. He further argued that this prosodic pattern of CT is distinct from that of T and F as F is realized as a falling accent (H*L) and T is produced without the high accent that is present in CT. Crucially, claims about the prosody of Korean CT have not been supported by experimental evidence, and in addition, no studies on Korean have examined the prosodic patterns in phrases following the CT.

The present study also examines the particle =NUN and its interaction with prosody. Lee (2008) argued that the CT marker =NUN can be omitted, but the high tone associated with CT has to remain in order to represent the meaning of CT. The current study thus examines whether it is acceptable to delete the particle =NUN and compares the prosody of CT *with*-NUN and *without*-NUN at a CT phrase. The particle =NUN is used to mark CT as well as T in Korean. Comparing the prosodic patterns of CT and T would thus show whether the semantic difference or morpho-syntactic similarity between the

two more strongly affects the prosodic pattern.

To examine these questions, an experiment with judgment and production tasks was conducted with five native speakers of Korean. Using a question-answer paradigm, CT, T, and F were elicited. Participants were asked to judge whether a given answer is felicitous to a given question, and their production of the question and answer was recorded. The experiment was designed to test the following hypotheses.

Hypothesis 1. The comparison between CT, T, and F: Although CT and T share the same particle in Korean, they will be produced differently as they are semantically distinct. Regarding CT and F, there are two possibilities: first, if the semantic value is important in prosodic realization, CT and F will be produced similarly as they are both F-marked; however, if semantic computation is important, CT and F will be produced differently since they undergo different semantic processes.

Hypothesis 2. The particle =NUN in CT: Following the argument of Lee (2008), the particle =NUN will not be necessary in marking CT; this means that the CT phrase without =NUN will be judged as acceptable. In addition, the deletion of the particle will not affect the prosodic pattern of the CT phrase.

2 Methods

2.1 Participants and Task

Five native speakers of Seoul Korean (three males, two females) participated in the experiment. There were nine blocks of 18 trials in each experimental session. On each trial, the context and question first appeared on the screen. After three seconds, the answer sentence appeared. All the answers started with a target word (i.e., CT, T, and F), but in CT, the target word was preceded by the phrase *"I don't know but. . . "*. This was to make the transition between the question and answer more natural in the CT context. All the questions and answers were provided in colloquial Korean. I consulted with two native speakers of Korean for the naturalness of the stimuli. Below are the example question-answer pairs each eliciting CT, T, and F (with context and question provided in English).

(3) **CT**

Context: Minsu, Ian, and Taehyun, who are friends of A and B, are going shopping this weekend.

A: What would Minsu buy?

B: (I don't know, but) Iani=(nun) kapang-ul sa-l kes kath-a
Iani=(CT) bag-ACC buy-might-DCL
'(I don't know but) Ian might buy a bag.'

(4) **T**

Context: A and B are talking about the trees planted in the school.

A: Where is the persimmon tree planted?

B: Kamnamwu=(nun) wuntongcang-ey sim-ecye-iss-e
Pers.tree=(TOP) playground-LOC plant-PASS-PRES-DCL
'The persimmon tree is planted at the playground.'

(5) **F**

Context: A is invited to B's place and they are having dinner together.

A: Who made this lotus dish?

B: Myenuli=(ka) yenkuncolim-ul mantul-ess-eyo
Daughter-in-law=(NOM) lotus dish-ACC make-PST-DCL
'Daughter-in-law made the lotus dish.'

Participants were asked to judge whether the answer is acceptable or not considering the context and question. After this judgment was given, the recording was conducted. They were asked to read out loud the question and answer, but not the context. Although the target of the analyses was the answer sentence, both question and answer were read. In addition, the recording was conducted for all trials regardless of the subjects' judgments in order to prevent them from skipping recordings by simply judging the trials as unacceptable.

There were three target conditions in the experiment: CT, T, and F. All of these conditions had the corresponding information structure (IS) category in the subject of the answer sentence (See (3)-(5)). CT and T were followed by F (i.e., *kapang* in (3) and *wuntongcang* in (4)), while F was followed by a given element already mentioned in the question (i.e., *yenkuncolim* in (5)). (6) shows the schematic structure of the answer sentence in each target condition; X represents the given element already mentioned in the discourse.

(6)

CT	+	F	+	X
T	+	F	+	X
F	+	X	+	X

In order to elicit F+F structure, the question must have two *wh*-words. An answer to that question can be understood as F+F, but alternatively as CT+F; thus, in order to avoid the semantic ambiguity, F was not followed by another F but instead was followed by the given word. Fillers looked similar to the target stimuli; for instance, they had F in the verb position or used pronouns instead of the given element.

To examine the question of the postnominal particle =NUN, all of the stimuli appeared in two variants, one *with* the particle and one *without* it. In the *with*-particle condition, the topic marker =NUN followed the IS word in condition CT and T, while the subject marker =I/KA followed the IS word in condition F. These markers were omitted in the *without*-particle condition.

2.2 Data Processing and Analysis

A total of 162 trials were recorded for each of the five participants; among them, 72 trials were CT, 18 were T, and another 18 were F, half of which had a particle, while the other half did not. Out of the 810 trials in total, 47 trials (5.8%) were discarded due to disfluencies, which left 763 trials. Among these trials, only the trials that were judged as acceptable were passed for acoustic analyses. 53 trials in CT (15.5%), 1 trial in T (1.2%), and 25 trials in F (31.3%) were excluded. A forced alignment was conducted for the remaining trials using the Korean Phonetic Aligner (Yoon and Kang 2013).

The F0 information was extracted via Matlab based on the pitch tracking algorithm in Talkin (1995). For each speaker, F0 values whose standardized values were > 2.807 were excluded (0.5% of a normal distribution). After outlier removal, F0 contours were interpolated and smoothed.

For the target word analysis, the number of F0 peaks and values of those peaks as well as the mean F0, F0 range, and the duration of the target word were measured. The F0 peaks were automatically detected and went through manual inspection. The duration of the target word was measured separately for the IS word and the particle. Thus, there were three duration measurements in the *with*-particle condition, total target word (IS word + particle), IS word, and particle, while only the duration of the IS word was measured in the *without*-particle condition (which is also the duration of the total target word in this case). To characterize the global F0 contour, the max F0 difference was calculated between prosodic word (pwd) 1 and pwd2 and between pwd2 and pwd3.

For all of the acoustic measurements (except the number of peaks), a mixed effects regression with IS category (i.e., CT, T, and F) and particle (i.e., *with* and *without*) as fixed effects and speaker as a random effect was conducted. If there was a significant main effect, two-sample t-tests were conducted for post-hoc comparisons. The t-tests were conducted with the raw residuals regressed on the random effect of speakers.

3 Results

The results of the judgment task found that the *without*-NUN CTs are acceptable although some speaker variation exists. Acoustic analyses showed that the prosodic pattern of CT is significantly different from that of T and F in the target word. However, analyses of the F0 contour of the entire sen-

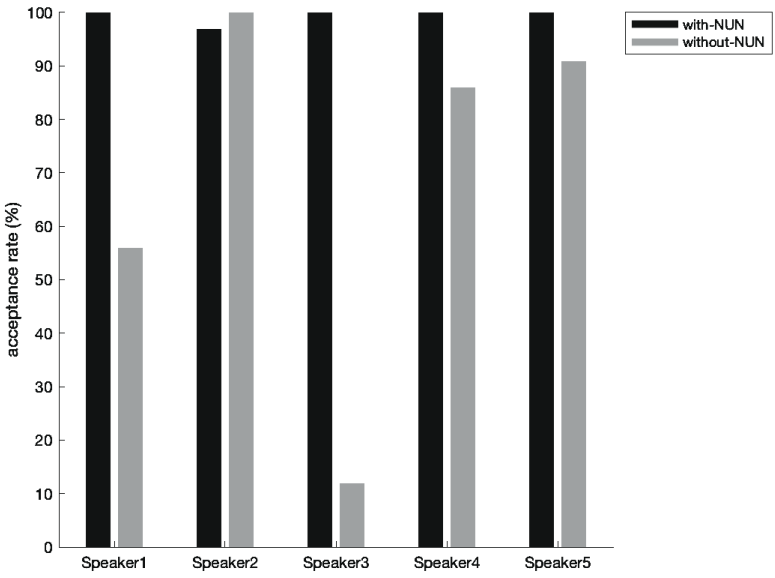


FIGURE 1 Judgment results of CT. The vertical axis is the percentage of the CT trials that were judged as acceptable. The black bar shows the acceptance rate of CTs *without*-NUN while the gray bar shows the acceptance rate of CTs *with*-NUN.

tence showed that CT and F are similar in that both lower the F0 peak of the following prosodic word. The comparison between the prosodic pattern of CT *with*-NUN and CT *without*-NUN at a CT phrase showed that they are significantly different in all acoustic measurements.

3.1 Judgment Results

The results of the judgment task showed that the particle =NUN can be omitted in CT, which supports Hypothesis 2. Figure 1 shows the acceptance rate of CTs *without*-NUN in five speakers. For comparison, the judgment results of CTs *with*-NUN are provided as well.

Although CTs *without*-NUN were not highly accepted as CTs *with*-NUN, they were judged as fairly acceptable. The CTs *with*-NUN were almost always judged as acceptable; the mean acceptance rate of the CTs *with*-NUN was 99.4%. With respect to CTs *without*-NUN, except for Speaker 3, participants judged them as fairly acceptable. Speaker 2, 4, and 5 showed over 80% of acceptance rate for CTs *without*-NUN (Speaker 2: 100%, Speaker 4: 86%, Speaker 5: 91%). In addition, Speaker 1 showed 56% of acceptance rate which was above the chance level. For Speaker 3, he moved to the United States much earlier than other speakers, and this may have affected his judg-

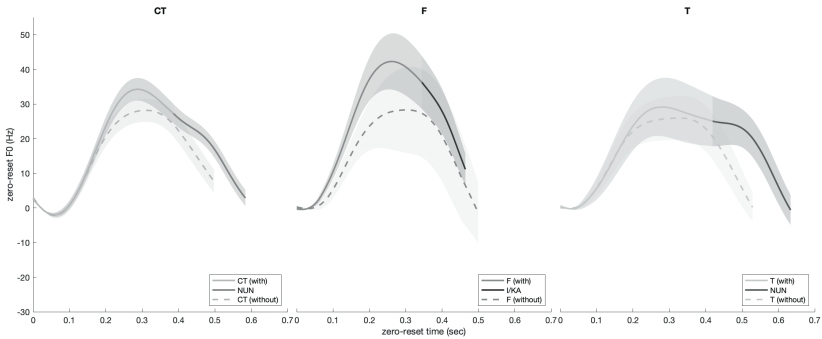


FIGURE 2 F0 contour of the target word. Time-warping was conducted separately for *with*- and *without*-particle condition and separately for the IS word and the particle. The middle lines represent the mean F0 value, and the bands indicate mean $\pm 1.96 \times$ standard error. F0 is shifted to 0 at the beginning of the target word.

ment of the less common type of CT which lacks the particle. Thus, although variation exists among speakers, the particle =NUN can be omitted in CT which comes at the subject of a sentence.

3.2 Acoustic Analyses

Regarding Hypothesis 1, it was found that CT and T are produced differently both in the target word and in the following phrases. This shows that although CT and T share the same particle, they have distinct prosodic patterns, consistent with the view that they are semantically different. With CT and F, they were produced differently at the target word, but they were similar following the target. This reflects the semantic relation between CT and F, where they share the same semantic value but undergo different semantic computations.

With respect to the prosodic aspect of Hypothesis 2, the deletion of the particle changed the prosodic pattern of the CT phrase. In other words, the prosodic pattern of CT phrase that does not have =NUN was different from the prosodic pattern of CT that has =NUN. All the acoustic measurements were significantly different between *with* and *without* condition.

CT vs. T vs. F

The target word analyses found a significant difference between CT and F in the height and number of peaks, F0 mean, and all duration measurements (i.e., total target word, IS word, particle), while CT and T were significantly different in the number of peaks, F0 mean, and the durations of the total word and IS word. Figure 2 shows the time-warped F0 contours based on the average of five speakers' data.

Analyses of the height of the peak found a significant difference between CT/T and F, but no difference was found between CT and T. A mixed-effects

linear regression found a significant main effect of IS category (χ^2(2, N=427) = 28.74, $p < 0.001$). Post-hoc comparisons showed that the peak height is higher in F than in CT and T (all ps < 0.001), but there was no significant difference between the two Topic conditions.

Regarding the number of peaks, almost all target words had either one or two peaks, and the proportion having one or two peaks was different according to IS category. (There were four trials in CT which had three peaks in the target word, but these trials were not included in the analysis.) Table 1 compares the number of trials with one peak versus two peaks in each IS condition. Among trials *with* the particle, F exclusively had one peak in the target word. In CT, there were more trials with two peaks, but still a majority of them had a single peak. The proportion of trials with two peaks increased in condition T as almost half of the trials had two peaks and the other half had a single peak. If a trial had two peaks, the second peak mostly occurred at the particle (See Figure 2). In this sense, although the same particle =NUN is used to mark CT and T, they are different in that =NUN in T is more likely to have the peak than the =NUN in CT. In sum, it can be argued that in the *with*-particle condition, F and T pattern differently in that F exclusively has one peak while T has a balanced number of trials with one peak and two peaks. CT is in between these cases, having usually one peak but some two peak trials. In the *without*-particle condition, almost 80% of the total data had a single peak although T still had some trials with two peaks. CT patterned more similarly to F in this case since it mostly had a single peak.

TABLE 1 Number of peaks in the target word

with-	1pk	2pks	*without-*	1pk	2pks
CT	131(77.5%)	38(22.5%)	CT	101(86%)	16(14%)
F	36(95%)	2(5%)	F	15(88%)	2(12%)
T	21(49%)	22(51%)	T	30(77%)	9(23%)

The F0 mean was significantly different in all IS conditions, while the F0 range did not show a difference. A mixed-effects linear regression found a significant main effect of IS category (χ^2(2, N=427) = 46.34, $p < 0.001$), and the post-hoc comparisons showed that the F0 mean is highest in F, second highest in T, and lowest in CT (all ps < 0.01). Analyses of the F0 range however did not show a significant main effect of IS category.

The duration measurements showed a significant difference in all IS categories in the total target word (IS word + particle) and IS word, while a significant difference was found only between F and CT/T in the duration of the particle. The duration of the total target word in the *with*-particle condition was significantly longer in T than in CT, which was significantly longer than F (all ps < 0.001). The same result was found in the duration of the IS

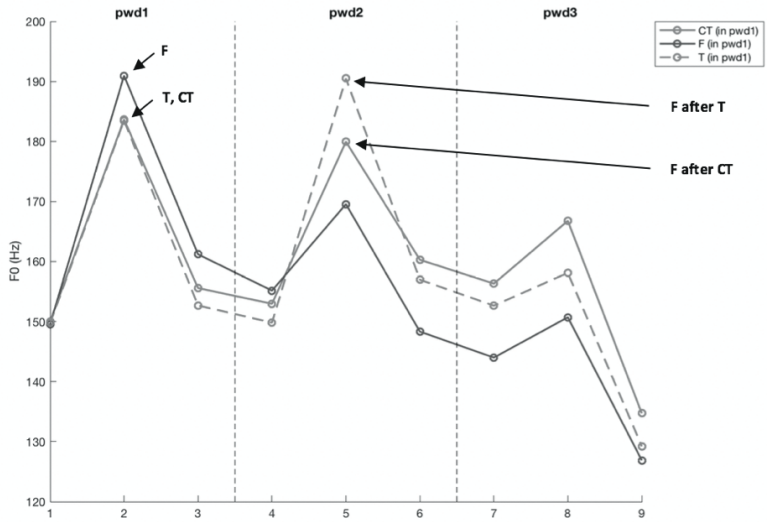


FIGURE 3 Schematic representation of the global F0 contours of the three target conditions. For each pwd, the initial, max, and end F0 values were measured. The mean F0 of the initial five samples was used as the start value, and the mean F0 of the last five samples was used as the end value in each pwd. The figure is based on all speakers' data and includes both *with-* and *without-* conditions.

word, showing T > CT > F, both in *with-* and *without-* particle conditions (all *p*s < 0.001). However, the duration of the particle was only significantly different between F and CT/T (all *p*s < 0.001), and no significant difference was found between CT and T. Unlike the analyses in the number of peaks, the =NUN in CT and T was similar in terms of duration.

While the target word analyses showed that the prosodic pattern of CT is distinct from both T and F, analyses of the F0 contour following the target word showed that CT patterns differently from T but similarly to F. Figure 3 is a schematic representation of the global F0 contours of the three IS conditions.

Analyses of the global F0 contour found that CT constrains the peak of the following pwd like F. The max F0 difference between the target word (pwd1) and the following word (pwd2) was largest in F, second largest in CT, and smallest in T (all *p*s < 0.001). In condition F, F (pwd1) was followed by a given element (pwd2), and the peak of the pwd2 was significantly lowered compared to the peak of the pwd1. In condition T and CT, both were identically followed by F (pwd2), but this F was realized differently; F following T was realized with a high peak, while F following CT did not show this property. It can be argued that the F-marked property of CT in pwd1 constrains

the prosodic realization of F in pwd2, which leads to a difference between CT and T.

Further comparison between the max F0 in pwd2 and pwd3 showed a significant difference between condition T and CT ($p < 0.001$). In condition T, the peak of the given element in pwd3 was lowered following the F in pwd2; yet in condition CT, since the F in the pwd2 was not realized with a high peak, the given element in pwd3 did not show compression. This shows that in condition CT, since F in the pwd2 was not realized as F due to the influence of CT in pwd1, the peak of the pwd3 is not lowered.

Overall, the comparison of the prosodic patterns between CT, T, and F over the entire sentence showed that CT is similar to F as they both constrain the F0 peak of the following pwd. However, CT, T, and F each showed a distinct prosodic pattern in the target word.

CT *with*-NUN vs. CT *without*-NUN

Given that the particle =NUN can be deleted in CT, it was examined whether CTs *without*-NUN have the same prosodic pattern as that of CTs *with*-NUN at the target word. A comparison between the two showed that they are significantly different in all acoustic measurements (See CT in Figure 2).

First, regarding the height of the peak, CTs *with*-NUN had a higher F0 peak than CTs *without*-NUN ($p < 0.001$). Second, the number of peaks was also different in that CTs *without*-NUN had more trials with a single peak than CTs *with*-NUN. While 77.5% of CT trials *with*-NUN had a single peak, 86% of trials had a single peak in CT *without*-NUN. Additionally, both F0 mean and range were significantly different between the two conditions (all ps < 0.01). It was found that the F0 mean was higher and the F0 range was wider in CTs *with*-NUN than CTs *without*-NUN. Lastly, the duration of the IS word, which was the only duration measurement that could be compared between the two conditions, was also significantly different; the duration of the IS word in *without*-NUN condition was significantly longer than that in *with*-NUN condition ($p < 0.001$). It is possible that speakers compensate for the absence of the postnominal particle by lengthening the IS word.

In sum, the deletion of the particle significantly changed the prosodic pattern of a CT phrase. As the second peak usually occurred at the particle, the deletion of =NUN resulted more trials with a single peak. Furthermore, while the F0 measurements (i.e., peak height, F0 mean, F0 range) were found to be higher in CTs *with* =NUN, the duration of the CT phrase was longer when the particle was dropped. These results are different from what is argued in Lee (2008), who maintained that CTs *without*-NUN will exhibit the same prosodic pattern as CTs *with*-NUN.

4 Discussion and Conclusion

The current study investigated how the semantic category of CT is realized in Korean by conducting judgment-production experiments. The prosodic pattern of CT was compared with that of T and F, both at the target word and the entire sentence. It was found that CT is produced differently from T and F in the target word, but it is similar to F in the global F0 contour. This result supports Hypothesis 1. Although CT and T are morpho-syntactically similar, they are produced differently both locally and globally. This indicates that semantics affects prosodic realization more strongly than morpho-syntax. Regarding CT and F, it was found that they are prosodically different at the target word, but similar in the global F0 contour. This prosodic pattern can be interpreted to reflect the semantic relation between CT and F, whereby they share the same semantic value but undergo different semantic computations. This also shows that semantics and prosody are closely related to each other.

The results on the target word are consistent with Lee (2008), who argued that CT has a distinct prosody. However, CT in the current experiment was not produced with high tone or rising accent as argued in Lee (2008). On the other hand, the results for the global contour matched what Tomioka (2010) found for Japanese. As in Japanese, Korean CT also showed post-peak compression similarly to F. The results in the current study highlight the importance of looking at experimental data and examining the prosody of CT both at the target word and the entire sentence.

It should be noted that in the target word, although CT had a prosodic pattern distinct from both T and F, it was more similar to T than to F. Some might argue that the prosodic similarity between the two is due to givenness, as T appears in the previous question while CT appears in the context (See (3) and (4) above; CT *Iani* appears in the context in (3), while T *Kamnamwu* appears in the question in (4).) This however is not quite true as there were CT stimuli where CT was not previously mentioned, as in the following example.

(7) **CT**

Context: A and B are talking about pets.

A: What do dogs like to eat?

B: (I don't know, but) Koyangi=(nun) yene-lul cal mek-e
Cat=(CT) salmon-ACC well eat-INT
'(I don't know but) Cats like to eat salmon.'

In (7), CT *koyangi* is not explicitly mentioned in the previous context or in A's question; this rules out the possibility that 'givenness' drives the prosodic similarity between CT and T. However, although the CT word was not explicitly mentioned, it is not too difficult for speakers to think of the word *koyangi*

since the context delimits the possible alternatives to 'pets'. We can therefore explain that CT is one of the *limited* alternatives and T is the given word, and this causes them to have a similar prosodic pattern. F is different from both CT and T as it does not have a *limited* set of alternatives. Further semantic work needs to be conducted to examine this claim.

The current study also examined the deletion of the particle =NUN and its interaction with prosody. It was found that =NUN can be omitted in CT which comes at the subject of a sentence. This result supports Hypothesis 2 regarding the deletion of =NUN. However, the prosodic aspect of the hypothesis was not supported, as CTs *without*-NUN had a significantly different prosodic pattern compared to CTs *with*-NUN.

In sum, the current study examined the prosodic properties of Korean CT using experimental data. The unique semantic properties of CT were manifested in its prosodic realization. These findings are important since they demonstrate the interaction between semantics and prosody. Furthermore, by examining the particle =NUN together with its prosody, this study contributes to the understanding of CTs in Korean. Future studies should be conducted by varying the location of CT (e.g., CTs in object position or the predicate CT) or examining CT in more complex structures such as embedded clauses. It would also be instructive to conduct a perception experiment to examine whether the acoustic differences found in the present study help speakers distinguish the different IS categories properly.

References

Büring, Daniel. 2003. On d-trees, beans, and b-accents. *Linguistics and philosophy* 26(5):511–545.

Constant, Noah. 2014. Contrastive topic: Meanings and realizations .

Jackendoff, Ray S. 1972. Semantic interpretation in generative grammar .

Lee, Chungmin. 1999. Contrastive topic: a locus of the interface–evidence from korean and english. *The semantics/pragmatics interface from different points of view* 1:317–342.

Lee, Chungmin. 2003. Contrastive topic and/or contrastive focus. *Japanese/Korean Linguistics* 12:352–364.

Lee, Chungmin. 2008. Contrastive (predicate) topic, intonation, and scalar meanings. In *Topic and focus*, pages 151–175. Springer.

Talkin, David. 1995. A robust algorithm for pitch tracking (rapt). *Speech coding and synthesis* 495:518.

Tomioka, Satoshi. 2010. Contrastive topics operate on speech acts. *Information structure: Theoretical, typological, and experimental perspectives* 115138.

Yoon, T and Y Kang. 2013. The korean phonetic aligner program suite.