Linguistic Research 31(2), 213-247

Semantic relations and multiple case constructions: an experimental approach*

Yong-hun Lee
(Chungnam National University)

Lee, Yong-hun. 2014. Semantic relations and multiple case constructions: an experimental approach. *Linguistic Research* 31(2), 213-247. Multiple Nominative Constructions (MNCs) and Multiple Accusative Constructions (MACs) have been some of the hottest and interesting topics in Korean syntax. Though there have been lots of previous studies on these constructions, most of them have provided theoretical accounts. Recently, Lee (2013) took an experimental approach and examined native speakers' grammaticality judgment of these constructions. This paper took the same approach to these constructions, but the experiment was performed with 100 native speakers. Ryu (2013) tried to unify MNCs and MACs into Multiple Case Constructions (MCCs) and to classify them into 16 types based on the semantic relations. This paper adopted his data sets and the experiment was performed based on these 16 semantic relations. The experiment was designed following Johnson (2008); and the native speakers' grammaticality judgments were measured with two scales, numerical estimates and line drawing, though the latter was adopted in the actual analyses. Through the experiment, the following facts were observed again: (i) The grammaticality of the MCCs does not constitute a homogeneous group, (ii) The grammaticality of the MCCs varies depending on which semantic relations hold between two NPs, (iii) MNCs had more grammaticality judgments than MACs if both constructions occurred in the similar contexts, and (iv) the sentences in some MAC types had much lower grammaticality than those in the others, as Ryu (2013) mentioned. **(Chungnam National University)**

Keywords Multiple Nominative Constructions, Multiple Accusative Constructions, Multiple Case Constructions, experimental approach, semantic relations

1. Introduction

Multiple Nominative Constructions (MNCs) and Multiple Accusative Constructions (MACs) are some of the hottest and interesting topics in Korean syntax. As Yoon (2004)

* I wish to thank two anonymous reviewers of this journal for their helpful comments and suggestions. All remaining errors, however, are mine.

pointed out, these constructions are some of the more puzzling phenomena in topic-prominent languages such as Korean.

For example, let's see the following two example sentences.[1,2]

(1) **N01**: integral object-component
Thokki-ka *kwi-ka* *kil-ta*.
rabbit.NOM ear.NOM be-long.DECL
'The ears of rabbits are long.'

(2) **N02**: collection-member
I *hamtay-ka* *camswuham-i* *manh-ta*.
this fleert.NOM submarine.NOM be-plenty.DECL
'There are plenty of submarines in this fleet.'

Both sentences contain MNCs, and they are examples extracted from Ryu (2013). Each sentence has two NPs: *Thokkili-ka* and *kwi-ka* in (1) and *I hamtay-ka* and *camswuham-i* in (2). Let's call the first NP NP1 and the second NP NP2 respectively.

There have been a lot of previous studies on these constructions, since MCCs are some of the most important phenomena in Korean syntax. However, most of the studies have focused on providing theoretical accounts for them, rather than empirical investigations. A few scholars pointed that there were some opinions that the grammaticality judgment of native speakers was not identical on these constructions, but there have been few studies on this topic.

Recently, Lee (2013) took an empirical approach to this problem and investigated how native speakers' grammaticality judgment varied depending on the semantic relations which hold between two NPs in these constructions. In this study, the

---

1 The nominative case markers *-ka* and *-i* and the accusative case markers *-lul* and *-ul* are allomorphs, respectively. The former is used in the post-vowel environments and the latter in the post-consonantal contexts. The Yale Romanization System is used for the romanization of the Korean words. The abbreviations for the glosses used in this paper are as follows: NOM (nominative), ACC (accusative), DAT (dative), PRES (present tense), PAST (past tense), DECL (declarative).

2 Here, **N01** and **N02** refer to the type of semantic relations. The sentence (1) has a integral object-component relation and (2) a collection-member relation.

grammaticality judgment of 20 native speakers was measured with the experiment, and some interesting findings were observed through the statistical examination of the collected data. Nevertheless, the study had shortcomings which were originated from a small number of informants. Though the study showed some interesting properties in the constructions, it was hard to generalize the findings in that paper into the general properties of MCCs.

This paper was designed to solve this problem. In this paper, the numbers of informants increased five times (100 native speakers), and more reliable experimental methods were used.[3] The goal of this paper is to check if the claims in Lee (2013) can be sustained with the extended data and if the findings can be generalized into the properties of these constructions.

In this paper, an experiment was designed and performed based on these 16 relation types, which were included in Ryu (2013) and Lee (2013). The experiment was designed following Johnson (2008); and the native speakers' intuition was measured with two scales, numerical estimates and line drawing, though the latter was adopted in the actual analyses.

This paper is organized as follows. Section 2 reviews previous studies on the empirical/experimental analyses of MCCs in syntactic research, especially focused on Ryu (2013) and Lee (2013). Section 3 mentions the research methods and procedure taken in this paper. Section 4 includes the analysis results, and Section contains discussions on the analysis results and compares the results with those of Lee (2013). Section 6 summarizes and concludes this paper.

2. Previous studies[4]

2.1 Grammaticality judgment task

A grammaticality judgment task (also known as native speakers' intuition test) is

---

3 For example, five different sets of questionnaires were used in the experiment in order to neutralize the effects of the sentence order which was given to the informants. In addition, unlike Lee (2013), the sentences in MNCs were reshuffled with those in MACs. For details, see Section 3.1.

4 Section 2.1 and Section 2.2 are also included in Lee (2013). They were included here again for explanatory convenience.

a psychological experiment which can be used to get the subconscious knowledge of native speakers in a given language. It involves asking native speakers to read a sentence and judge if it is well-formed (grammatical), marginally well-formed, or ill-formed (unacceptable or ungrammatical) (Carnie, 2012).

As Johnson (2008:218) mentioned, in syntactic research, an interval scale of grammaticality have commonly been used. There are usually five steps of scales (like Likert scales), and sentences are rated by native speakers as grammatical (no mark), questionable (? or ??), and ungrammatical (* or **). This is essentially five-point category rating scale, and the researcher could give people this rating scale and average the test results, where **=5, *=4, ??=3, ?=2, and no mark=1. However, it has been observed in the study of sensory impressions that raters are more consistent with an open-ended ratio scale than they are with category rating scale (Stevenson, 1975). Recently, researchers have had an interest in native speakers' intuition on syntactic data (Bard, Robertson, and Sorace, 1996; Schütze, 1996; Cowart, 1997; Keller, 2000). So, in recent years, various methods have been adapted into the study of sentence acceptability, from the study of psychophysics which studies the subjective impressions of physical properties of stimuli.

Johnson (2008) adopted a technique, so called *magnitude estimation*, using an open-ended ratio scale for reporting the impressions of native speakers. The experiment in his proposal starts with a demonstration of magnitude estimation by asking participants to judge the length of a few lines. These practice judgments provide a sanity check in which we can evaluate the participants' ability to use magnitude estimation to report their impressions. Stevenson (1975) found that numerical estimates of line length have a one-to-one relationship with actual line length (that is, the slope of the function relating them is close to 1). In the second session, the participants were presented with sample sentences. Some are grammatical and the others are not. Then, the participants were instructed to judge how good or bad each sentence is by drawing a line that has a length proportional to the grammaticality of the sentence. In the third session, the participants were provided the target sentences. Their job was to estimate the grammaticality of the target sentences by drawing lines, which indicated native speakers' impression of the grammaticality with the length of the line that they drew for the sentences. In the last session, the participants were provided the same target sentences. However, they were asked to estimate the grammaticality of the target sentences with numerical

estimations, which also indicate their impression on the acceptability of the target sentences.

Lodge (1981) mentioned that this *magnitude estimation* has three advantages over *category scaling*. First, the latter has limited resolution. For example, if native speakers may feel that a sentence is somewhere between 4 and 5 (something like 4.5), gradient ratings are not available in the latter method. However, the former permits as much resolution as the raters wish to employ. Second, the latter method uses an ordinal scale, and there is no guarantee that the interval between * and ** represent the same difference of impressions as that between ? and ??. The former method, on the other hand, provides judgments on an interval scale for which averages (mean value, *m*) and standard deviations (*sd*) can be more legitimately used. Third, the latter limits our ability to compare results across the experiments. The range of acceptability for a set of sentences has to be fitted to the scale, and what counts as ?? for one set of sentences may be quite different from what counts as ?? for another set of sentences.

However, *magnitude estimation* also has some shortcomings. In Johnson's proposal, for example, the participants in the example experiment were asked to judge sentences into two ways: (i) by giving a numeric estimate of acceptability for each phrase, as they did for the lengths of lines in the practice session; and (ii) by drawing lines to represent the acceptability of each line. Bard et al. (1996) found that the participants sometimes think of numeric estimates as something like academic test scores, and so they limit their responses to a somewhat categorical scale (e.g. 70, 80, 90, 100), rather than using a ratio scale as intended in the magnitude estimation. Consequently, the participants have no such preconceptions about using a line length to report their impressions, and we might expect more gradient unbounded responses by measuring the lengths of lines that participants draw to indicate their impressions of sentence grammaticality.

2.2 Previous studies on MCCs in Korean

Since Case markers are one of the typical syntactic phenomena in Korean, there have been lots of previous studies on this topic. These previous studies on Case markers are divided into roughly two groups.[5] One is syntactic approaches and the other is semantic approaches.

Syntactic approaches, once again, can be divided into two types: Constituent Approaches and Non-constituent Approaches. Constituent Approaches are based on the concept of possessor-raising or genitive NP. In this approach, NP has a structure [$_{NP}$ NP1 NP2] where NP1 becomes a possessor and NP2 is a possessee. Then, NP1 moves out from the NP, and the Case marker of NP1 changes into the Nominative marker *-ka* or the Accusative marker *-lul*. Many analyses including Choe (1987), Kitahara (1993), Ura (1996), and Cho (2000) took this approach.

Non-constituent Approaches have two different types of analyses. The first one is Major Subject Analyses. This approach assumes that Korean may have sentential predicates and that this language has a major subject in addition to the usual subject position. In this type of analysis, both NP1 and NP2 can be the subjects, and various notions of subjects are defined. In fact, this type of analysis started from Choe (1937), where he called them a *big subject* and a *small subject* respectively. Recently, Yoon (2003, 2009) and Lee (2007) took this approach. The second type is Topic/Focus Analyses. In these types of analyses, only NP2 is a subject and NP1 becomes a topic or a focus. Hong (1991), Rhee (1999), Yoon (1986), Schütze (2001), Kim (2000, 2001, 2004), Kim and Sells (2007), and Kim et al. (2007), Park (2005), Choi (2012) adopted this approach.

In contrast to syntactic approaches to MCCs, semantic approaches have focused on the licensing issues. That is, the semantic approaches to these constructions have tried to uncover what semantic relations hold between NP1 and NP2. Several semantic relations have been proposed to account for the MCCs, and they include the followings: the *macro-micro* relations (Yang, 1972), the *inalienable possession* (Kang 1987, Choe 1987, Kim 1989, Kim 1990, Yoon 1989, Maling and Kim 1992, Kitahara 1993, Yoon 1997, and Moon 2000), the *g*(*eneralized*)*-possession* in Park (2001), the (*thematically*) *subordinate* condition (Na & Huck 1993, Kim 2000, Kim 2001, Kim 2004, Kim and Sells 2007, and Kim et al. 2007), and the *aboutness* condition (Kang 1988, O'Grady 1991, Hong 1997, Yoon 2004, Choi and Lee 2008, Choi 2012).

---

5 As mentioned in Lee (2013), there is another type of syntactic approach to MCCs, though it has not been discussed much in a lot of literature. That is Case Spreading Analysis in Role and Reference Grammar (RRG; van Valin and Foley, 1980; van Valin and LaPolla, 1998). In the RRG account of MCCs, Nominative/Accusative Case markers can spread from one point to the other direction. Park (1995), Han (1999), and van Valin (2009) provided this type of account to Korean MNCs and MACs. For an example sentence, see the footnote 11.

Even though there are a lot of studies on the theoretical accounts for MCCs, only a few provided the classifications of MCCs in Korean, such as Yang (1972), Na and Huck (1993), and Park (2001). Recently, Ryu (2013) tried to unify MNCs and MACs into Multiple Case Constructions (MCCs) and to provide a unified account for them. He also classified the MCCs into 16 different types based on the *conceptual linking hierarchy*, which was constructed by collecting and classifying the semantic relations in the previous approaches. During the process, Ryu (2013: 17) summarized these classifications as follows.

Table 1. Types of multiple case marking constructions

Proposed type of MCCs	NOM-NOM	ACC-ACC	Yang (1972)	Na & Huck (1993)
Type 01 integral obj.-component	○	○	whole-part	meronomic rel.
Type 02 collection-member	○	○	×	×
Type 03 mass-portion	○	○	×	×
Type 04 object-stuff	○	○	×	×
Type 05 activity-feature	○	○	×	×
Type 06 area-place	○	○	×	×
Type 07 class-membership	○	○	class-member type-token	taxonomic rel.
Type 08 object-attachment	○	○	×	×
Type 09 object-quality	○	○	×	qualitative
Type 10 object-quantity	○	○	total-quantity	×
Type 11 space-object	○	*	×	×
Type 12 time-object	○	*	×	×
Type 13 possessor-object	○	*	×	×
Type 14 conventional relation	○	*	×	×
Type 15 object-predication	○	*	×	conventional
Type 16 conversive relation	○	*	affected-affector	conversive

The first column enumerates the 16 types of semantic relations, which holds between NP1 and NP2. Ryu (2013) re-organized the classifications based on previous studies such as Yang (1972), Na and Huck (1993), and Park (2001). Some of the type

names came from the previous studies, and others were made by him. The second and third column demonstrates if these types occur in the MNCs and MACs. Here, the symbol ○ refers to 'possible' and * to 'impossible'. The last two columns show us how each semantic relations were referred to in Yang (1972) and Na & Huck (1993) respectively. Here, the symbol × refers to 'not mentioned'. And, rel. and con. are abbreviations of *relation* and *constructions* respectively.

The criteria of these classifications are the semantic relations which hold between the two consequative NPs, i.e. the semantic relations between NP1 and NP2. He also provided example sentences for each type of construction.[6]

2.3 An empirical/experimental approach: Lee (2013)

Lee (2013) started the study from the following observation. When the example sentences in (1) and (2) were given to some native speakers in Korean, their grammaticality judgments were drastically different depending on which sentence were given to them. The above example sentences are shown here again for your convenience.[7]

(1) **N01**: integral object-component
Thokki-ka *kwi-ka* *kil-ta*.
rabbit.NOM ear.NOM be-long.DECL
'The ears of rabbits are long.'

(2) **N02**: collection-member
I *hamtay-ka* *camswuham-i* *manh-ta*.
this fleert.NOM submarine.NOM be-plenty.DECL
'There are plenty of submarines in this fleet.'

Each sentence has two NPs: *Thokkili-ka* and *kwi-ka* in (1) and *I hamtay-ka* and *camswuham-i* in (2). In spite of the structural similarity of these two sentences, most

6 These sentences were used as target sentences in Lee (2013) and this paper. They are provided in Appendix.

7 Here, **N01** and **N02** refer to the type of semantic relations which were shown in Table 1. For the example sentences, see Appendix.

of the native speakers answered that (1) was much better than (2), and more than half of them said that (2) was ungrammatical.

These examples demonstrate that native speakers' intuition is different even within the MNCs. Then, what makes these differences in MNCs and MACs? In order to answer this question, the paper took an experimental/empirical approach.

The paper basically followed the experimental design described in Johnson (2008). First, the 16 target sentences were made based on the 16 semantic relations which were proposed in Ryu (2013), and the distracting sentences of the same number (16 sentences) were also provided for MNCs and MACs respectively. Accordingly, a total of 64 sentences were included in the experiment. Then, the collected sentences were randomly ordered and provided to the informants. The experiments were conducted two times in the fall semester in 2013, for the purpose of consistence testing with the same informants. One is for MNCs and the other is for MACs. The informants were registered in university (freshmen and sophomores) at the time of the experiments, who were not linguistics majors. Each experiment was performed as follows. Each experiment consisted of four sections, following Johnson (2008). In the first section, the informants were given a sample line, and the numerical score of 100 was given to the line. Then, they were provided with 10 lines with different length, and they were instructed to judge the length of the lines. They were said to write the numerical estimates for each line, which they thought of as the lengths of the lines compared with the standard line with the numerical score of 100. In the second section, they were given a sample Korean sentence perfectly grammatical. The numerical estimate 157 was given to the sentence. This value was given to the informants in order to avoid the same problem that Bard et al. (1996) pointed out. Then, they were provided with 10 different Korean sentences. Some of them were grammatical, some others were ungrammatical, and the others are in-between. They were instructed to draw a line for each sentence which corresponded to their judgment on the acceptability, compared with that of the standard line. The possible length of the lines ranged from 0 mm to 170 mm.

Among the 27 students, only 20 data sets were available in the final step. Accordingly, the grammatical judgments of natives speakers were statistically analyzed based on these 20 data sets. Since line drawing showed a significant correlation with numerical estimates, the scores for line drawings were selected for statistical analysis.

In order to examine the distributions of data in each group (**N01-N16** and **A01-A16**), Shapiro-Wilk Normality Tests were performed and we found that only one set of data (**A04**) did not follow the normal distribution. Then, we examined each group of data, and the following graph summarizes the analysis results in the paper. Here, each point in the graph refers to the position of the mean values, and the I-shaped lines represent 95% confidence intervals (CIs).[8]

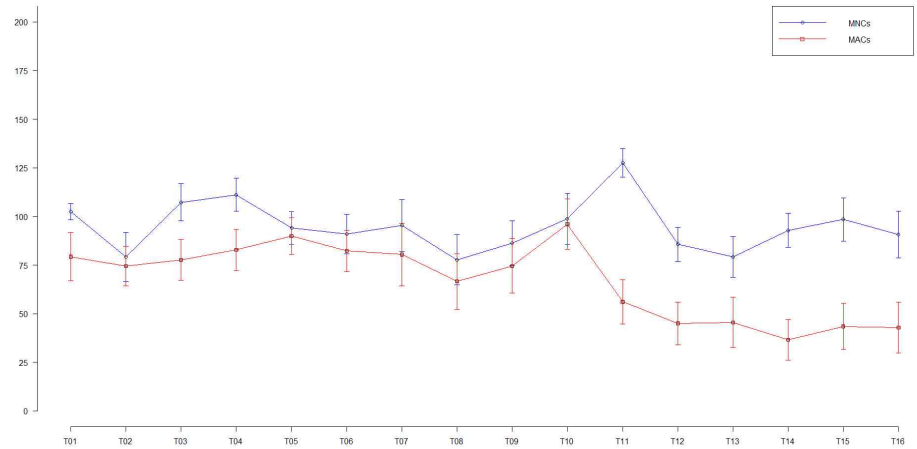


Figure 1. The statistical analysis results of MCCs in Lee (2013)

From the statistical analysis of these data, the following facts were observed.

First, the grammaticality of the MCCs does not constitutes a homogeneous group. Since only one group of data did not follow the normal distribution, a repeated-measures ANOVA was performed for both MNCs and MACs. The result was that the mean values became significantly differentiated depending on which semantic relations both MNCs and MACs had (F=6.818, p<0.001 for MNCs; F=11.01, p<0.001 for MACs)

Second, the grammaticality of the MCCs varies depending on their semantic relations. Since the mean values became significantly differentiated depending on the types of semantic relations both in MNCs and MACs, a Tukey's HSD test (the parametric post-hoc test) was performed. In MNCs, 19 pairs (15.83%) among the 120 pairs had statistically significant differences. Likewise, 45 pairs (37.50%) had

8 The 95% CIs were not included in Lee (2013). They were shown here for easy comparison.

the significant differences in MACs.

Third, MNCs were more grammatical than MACs if both constructions occurred in the same environments. In order to check the differences between MNCs and MACs, paired *t*-tests were performed. The results showed that the distributions of MNCs were significantly higher than those of MACs (t=11.99, p<0.001). In addition, the CIs in Figure 1 demonstrates, 9 groups of data have significant differences between MNCs and MACs.[9]

Fourth, the sentences in some MAC types had much lower grammaticality than those in the others, as Ryu (2013) mentioned. Ryu (2013) claimed that the grammaticality judgment of the sentences from **A11** to **A16** is significantly different from those of the others. He said that the former sentences were ungrammatical. In order to examine if there are statistically significant differences between the two groups, the data were collected into two groups separately. Group 1 was composed of the data from **A01** to **A10**, and Group 2 was from **A11** to **A16**. Since the Shapiro-Wilk Normality tests said that both groups did not follow the normal distributions (Group 1: p<0.001; Group 2: p<0.001), a Mann-Whitney's U test (the non-parametric counterpart of an independent sample *t*-test) was performed. The results was that there are statistically significant differences between the two groups (W=19796.5, p<0.001). This results indirectly supports Ryu's claim (2013) that the grammaticality of the sentences from **A11** to **A16** are different from those in the other groups.

Though the analysis results in Lee (2013) demonstrated some crucial properties of MCCs in Korean, they have also shortcomings.

First, the most critical problem is that the number of informants is too small. The data for only 20 native speakers' intuition were considered in the analysis. Though it is possible to observe the overall tendencies of MCCs, it is hard to generalize the analysis results to the grammaticality judgment of the Korean people. In addition, the analysis results can significantly be affected by the extreme data, if the number of data is small. Therefore, it is necessary to re-test the grammaticality judgment of native speakers with more extended number of informants.

---

9 When two data sets have significant differences, there is no overlap in CIs. Accordingly, the following 9 pairs have significant differences between the distributions of MNCs and those of MACs: **N01-A01**, **N03-A03**, **N04-A04**, **N11-A11**, **N12-A12**, **N13-A13**, **N14-A14**, **N15-A15**, and **N16-A16**.

Second, the range of age of the informants was too restricted in the experiment. The majority of the students were 19 and 20, and the mean (*m*) and standard deviation (*sd*) were 20.15 and 0.93 respectively. Therefore, it is necessary to collect the data from more various age groups.

Third, there may be some problems in randomizing. In the last experiment, since the number of informants was too small, only one set of questionnaire was used. However, in this case, the answer might be affected by the order of the sentences which was given to them. It is necessary to remove this unsatisfactory factor in the experiment.

3. Research methods

3.1 Experimental design

As in Lee (2013), this paper basically followed the experimental design described in Johnson (2008). The target sentences were basically the same that were used in Lee (2013) and they originally came from Ryu (2013). There are two reasons to use the sentences again. First, Ryu (2013) contained almost all of the semantic relations which occurred in MCCs. Therefore, it was possible to cover most types of semantic relations in MCCs. Second, Ryu (2013) provided the sentences which belonged to both MNCs and MACs. Accordingly, it was easy to get the target sentences for both types of constructions. The experiment in this paper used the target sentences without any modification in order to avoid any irrelevant distortion when the lexical items were changed.

Example sentences for MNCs were given in (1) and (2). Likewise, the target sentences for MACs were also extracted from Ryu (2013). The following sentences in (3) and (4) are the counterparts of (1) and (2) (Ryu, 2013:9).[10,11]

---

10 As in Lee (2013), the experiment in this paper includes the sentences in type **A11-A16** to check how much their grammaticality was bad to native speakers.

11 As mentioned in Lee (2013), it doesn't imply that MACs have parallel structures with MNCs and that they have to be analyzed with the same mechanism with MNCs. Let's see the following sentences (Han, 1999).

(i) a.	*Chelswu-ka*	*Yenghi-eykey*	*kkoch-ul*	*cwu-ess-ta.*
	Chelswu.NOM	Yenghi.DAT	flower.ACC	give.PAST.DECL

(3) **A01**: integral object-component

Hans-ka	*thokki-lul*	*kwi-lul*	*cap-ass-ta.*
Hans.NOM	rabbit.ACC	ear.ACC	grab.PAST.DECL

'Hans grabbed the ears of rabbits.'

(4) **A02**: collection-member

Cekkwun-i	*i*	*hamtay-lul*	*camswuham-ul*	*paksalnay-ss-ta.*
enemy.NOM	this	fleert.ACC	submarine.ACC	destroy.PAST.DECL

'The enemy destroyed the submarines of this fleet.'

As in Lee (2013), because both MNCs and MACs had 16 types of semantic relations, a total of 32 target sentences were included in the experiment. Along with these target sentences, distracting sentences of the double number (32 sentences) were also provided for MNCs and MACs respectively, unlike Lee (2013).[12] Among the distracting sentences, the half of them were constructed by replacing the Case markers with the topic marker -(*n*)*un*, and the other half were constructed from the combination of grammatical and ungrammatical sentences which have no relations with the target sentences. Accordingly, a total of 96 sentences were included in the experiment. Then, the collected sentences were randomized using the sampling function in R (a statistics software). However, unlike Lee (2013), the randomizing processes proceeded as follows. The randomizing function was run five times, and the different order of numbers (from 1 to 96) was generated each time. Then, the generated numbers were given to each sentence, and the sentences were sorted by the assigned numbers. At last, the final questionnaire was made by sorting the sentences with the assigned numbers. Accordingly, five different sets of

---

'Chelswu gave a flower to Yenghi.'

b. *Chelswu-ka* *Yenghi-lul* *kkoch-ul* *cwu-ess-ta.*
Chelswu.NOM Yenghi.ACC flower.ACC give.PAST.DECL
'Chelswu gave Yenghi a flower.'

In the RRG account, this sentence can be explained with Case Spreading. That is, the Accusative marker -*lul* spreads to the left, and the Dative Case marker -*eykey* in (ia) is changed into an Accusative in (ib). Though both (3)/(4) and (ib) contain MACs, their sources are different. In our terminology, the two NPs have different semantic relations in (3) and (4). It may be impossible to improvise the MNC counterpart of the sentence in (ib). As this sentence illustrates, MACs may have different syntactic structures and semantic relations from MNCs.

12 In Lee (2013), the same numbers of extracting sentences were given to the informants.

questionnaires were generated through the randomizing processes, and these questionnaires were randomly provided to the informants.[13]

The experiments were conducted for the 5 different groups of students in March of the 2014 spring semester. The experiment was performed as follows. The questionnaire consisted of four sections, following Johnson (2008). In the first section, the informants were given a sample line, and the numerical score of 130 was given to the line. Then, they were provided with 10 lines with different length, and they were instructed to judge the length of the lines. They were said to write the numerical estimates for each line, which they thought of as the lengths of the lines compared with the standard line with the numerical score of 130. In the second section, the informants were given a sample Korean sentence which is perfectly grammatical. Unlike Johnson (2008), both the line drawing and the numerical estimate 183 was given to the sentence. This numerical value was given to them in order to avoid the same problem that Bard et al. (1996) pointed out. Then, they were provided with 10 different Korean sentences. Some of them were grammatical, some others were ungrammatical, and the others are in-between. They were instructed to draw a line for each sentence which corresponded to their judgment of the acceptability, compared with that of the standard line with the numerical score of 183, and they were also instructed to provide the numerical estimate for the given sentence. The possible length of the lines ranged from 0 mm to 170 mm, and the possible range of numerical scores was from 0 to 200.[14] In the third section, the target sentences were given. The informants were instructed to estimate the grammaticality of the target sentences by drawing lines. The possible length of the lines ranged from 0 to 170 mm, as in Lee (2013). In the last session, the informants were provided with the same target sentences. Now, they were to estimate the grammaticality of the target sentences with numerical estimates. The possible range of numerical scores was from 0 to 200.

After the experiment, all the data for the 32 sets of target sentences were

---

13 Also note that the randomizing function was performed over all of the 96 sentences. In Lee (2013), since the experiments were performed two times (one for MNCs and the other for MACs), the randomizing functions were performed twice. However, the randomizing function was performed over all of the 96 sentences in this experiment, whether the sentence belonged to MNCs or MACs.

14 The possible lengths of the lines ranged from 0 to 190 mm if the participants used them up to the right margin. However, the maximum value was 163 mm in all of the experiments.

extracted: 16 for MNCs and 16 for MACs. A total of 132 students participated in the experiments. Among the 132 students, only the data for 117 informants were available.[15] However, among the answers of these 117 students, some answers were missing.[16] That is, there were some students who answered to some sentences but provided no answer to some others. A total of 14 students answered in this fashion, and the data sets for these informants were excluded. Finally, the data sets of the remaining 103 participants were extracted. However, among those students, one belonged to the outlier in terms of their ages and two are very close to it. Accordingly, the data sets for these three students were also excluded. Consequently, the data sets for only a total of 100 students were included in the statistical analyses. The age distribution of those 100 students was as follows (m=22.21, sd=1.909).

For each informant, 32 target sentences were collected (16 for the MNCs and 16 for MACs). For each of the data sets, two different kinds of data were extracted: one for *numerical estimates* and the other for *line drawing*. Since two different kinds of scales were used in the experiment, it was necessary to check the correlation between these two scores. Figure 2 shows the correlations of the first data set **N01**.

---

15 Also note that the randomizing function was performed over all of the 96 sentences. In Lee (2013), since the experiments were performed two times (one for MNCs and the other for MACs), the randomizing functions were performed twice. However, the randomizing function was performed over all of the 96 sentences in this experiment, whether the sentence belonged to MNCs or MACs.

16 In fact, there were two more experiments except the one described in this paper. Accordingly, a total of three different experiments were performed to the same informants. The goal of the first experiment, which was described in this paper, was to investigate how the semantic relations affected the grammaticality judgment of native speakers. The second and the third experiment were to examine how the inalienable/alienable possessions and the number of NPs affected the grammaticality judgment of the MCCs respectively. The experiments were performed at the beginning of the 2014 spring semester, just before the midterm exam, and just before the final exam. In order to examine how each factor affected the grammaticality judgment of the students, the private information of the informants was also controlled. In the questionnaires, the informants were asked to add their personal information (name and student number) so that the three sets of data could be correctly aligned per each person after the experiment. Then, only the data were selected for the person who answered all of the three times of questionnaires. That's why only 117 sets of data were chosen among the 132 data sets.

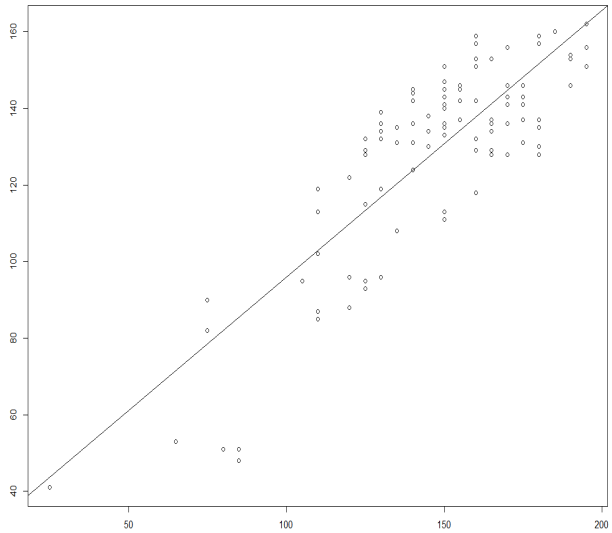


Figure 2. Correlation between numerical score and line length for the set **N01**

Here, *r* was 0.835. Since it is said that the two variables have correlations if the *r* value is over 0.5, it will be safe to say that the *line drawing* and *numerical estimate* are highly correlated in this data set. The mean and standard deviation of the whole data sets were 0.897 and 0.028 respectively.

In the actual statistical analyses below, the scores for the *line drawing* were used. The reason was that the problem of category scaling can be avoided in the scores for line drawing. Even though the informants were given the 0-200 numerical ranges, they used only some of them, i.e., the multiple numbers of 5 or 10. In the scores for the line drawing, since they were instructed to draw a line without a ruler, it was possible to avoid such kind of subconscious tendency. Because the line lengths were highly correlated with the numerical estimates, it was possible to use only the scores for the line drawing in the analysis.

3.2 Normality test

After the scores for the *line drawing* were chosen for each target sentence, the first thing that we had to do was a normality test. The reason was that the types of statistical tests were determined by the results of the normality tests. If the

distributions of data followed the normal distribution, we could apply parametric tests such as a *t*-test or an ANOVA. If not, non-parametric tests had to be applied, including Wilcoxon tests or Friedman tests. Therefore, it was important to check if the distributions of data sets followed the normal distribution or not.

There are a few different sorts of normality tests. One is to use a Normal Quantile Plot (Baayen, 2008). For example, the 100 data for the set **N02** can be represented in the Normal Quantile Plot as follows:

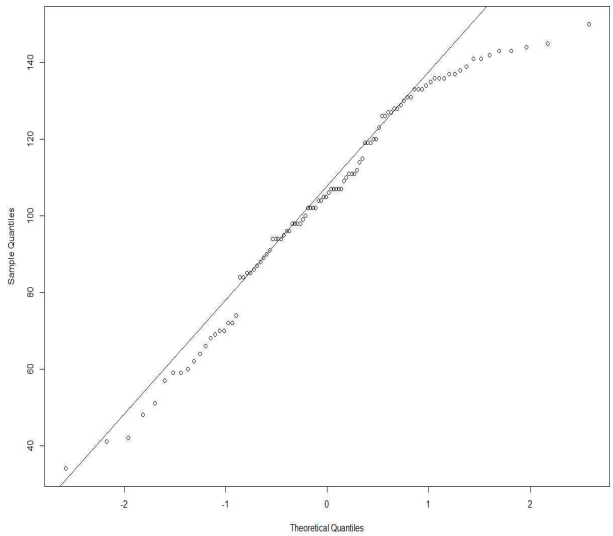


Figure 3. Normal quantile plot for the set **N02**

In this plot, the closer the points get to the Q-Q line, the closer they are to the normal distribution. As you can see, most of the points, especially those in the middle, are attached very close to the Q-Q line. Accordingly, we may guess that these data follow the normal distribution. However, see the upper right part of the plot. Most of the points are very far from the Q-Q line. Consequently, you cannot be sure the normality of the distribution.

One of the disadvantages using the Normal Quantile Plot is that we cannot numerically decide if the given data follows the normal distribution or not. The normality test that solves this problem is a Shapiro-Wilk Normality Test. For example, if we perform the test with the scores for the set **N01**, we have a *p*-value

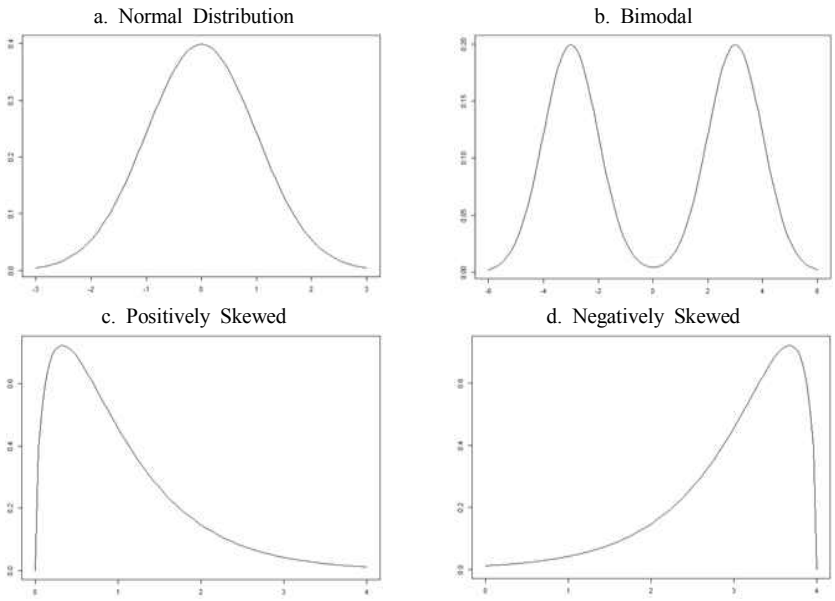
0.005. Since this *p*-value is much smaller than the *a*-value of 0.05, we can reject the Null Hypothesis that this data follows the normal distribution. That is, we cannot say that this data follows the normal distribution.

In the actual statistical analyses, Shapiro-Wilk Normality Tests were used. If the *p*-value is bigger than the *a*-value of 0.05, the data is said to follow the normal distribution. If the *p*-value is smaller than the *a*-value of 0.05, the data is said not to follow the normal distribution. In the data sets of our experiment, only one set of data (**A06** in Table 4, *p*=0.081) followed the normal distribution. Accordingly, non-parametric tests were frequently used such as Wilcoxon tests or Friedman tests.[17]

---

17 Unlike Lee (2013), only one data set followed the normal distribution in this experiment. Each data set was closely examined again, since it was known that the data followed the normal distribution as the number of data increased (Gries, 2013:34-35, for example). After the close examination, it was found that there were some patterns in the distributions of our data sets. Johnson (2008:14) mentioned some distributions of data as follows.

(i) Types of distributions



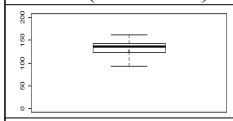
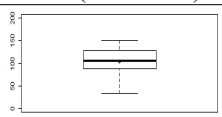
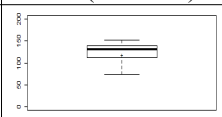
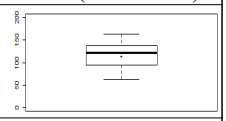
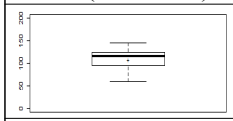
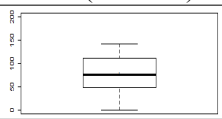
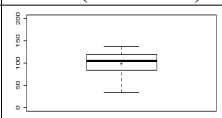
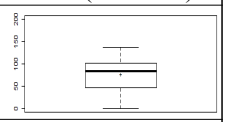
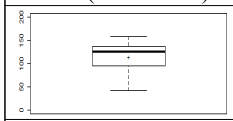
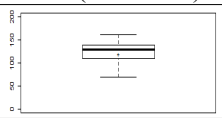
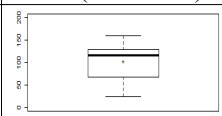
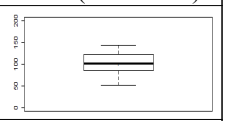
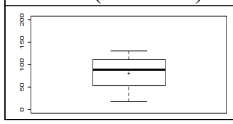
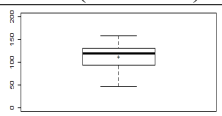
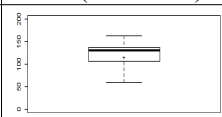
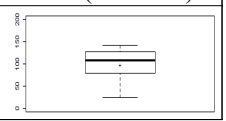
As many statistical books pointed out, the distribution becomes close to (a) as the number of data increase. However, the distributions of grammaticality judgment of native speakers in the experiment

4. Analysis results

4.1 MNCs

Table 2 illustrates the results of the grammaticality judgment task for the 16 types of MNCs. For each type, the mean values are provided in addition to the box plots, where little plus signs (+) were used for the (arithmetic) means in the box plots.

Table 2. Results of the grammaticality judgment task for the 16 types of MNCs

N01 (*m*=128.20)	**N02** (*m*=104.00)	**N03** (*m*=118.4)	**N04** (*m*=116.15)
			
N05 (*m*=109.00)	**N06** (*m*=79.31)	**N07** (*m*=100.70)	**N08** (*m*=77.94)
			
N09 (*m*=113.90)	**N10** (*m*=119.60)	**N11** (*m*=103.70)	**N12** (*m*=101.50)
			
N13 (*m*=81.83)	**N14** (*m*=112.06)	**N15** (*m*=115.00)	**N16** (*m*=98.00)
			

In order to examine if the mean values of each type became different depending on the semantic relations of two consequative NPs, a statistical test had to be performed. Since all the types in MNCs did not follow the normal distribution, a Friedman test (the non-parametric counterpart of a repeated-measures ANOVA) was

didn't followed the distributions in (a), even though the same controlling conditions that Lee (2013) adopted were given in the experiment. Rather, the distributions had the form (b), (c), or (d). The reason seems to be that the native speakers' intuitions were measured in the experiment. The strong positive or negative tendency toward the target sentences seems to be reflected in the distributions of each data set.

performed, and the result was that the mean values became significantly differentiated depending on the semantic relations that these two NPs had ($X2$=417.437, p<0.001).

Next, in order to examine the mean value of which type was significantly different from that of which type, a pairwise Wilcoxon test (the non-parametric counterpart of a Tukey's HSD test) was performed, and its results are shown in Table 3. Here, '×' is used when 0.1<p, *ms* (marginally significant) when p<0.1, '*' (significant) when p<0.05, '**' (very significant) when p<0.01, and '***' (highly significant) when p<0.001.

Table 3. Results of the pairwise Wilcoxon test for the 16 types of MNCs

	N01	**N02**	**N03**	**N04**	**N05**	**N06**	**N07**	**N08**	**N9**	**N10**	**N11**	**N12**	**N13**	**N14**	**N15**
N02	***														
N03	×	*ms*													
N04	×	**	×												
N05	***	×	×	×											
N06	***	***	***	***	***										
N07	***	×	***	**	×	***									
N08	***	***	***	***	***	×	***								
N09	***	×	×	×	×	***	**	***							
N10	*	*	×	×	×	***	***	***	×						
N11	***	×	*	×	×	***	×	***	***	***					
N12	***	×	**	**	×	***	×	***	×	***	×				
N13	***	***	***	***	***	×	*	×	***	***	***	×			
N14	***	×	×	×	×	***	***	***	×	×	×	**	***		
N15	×	×	×	×	×	***	**	***	×	×	×	×	***	×	
N16	***	**	***	***	**	**	×	***	×	**	×	×	**	×	***

Among the 120 pairs (=16×(16-1)/2), 68 pairs (except *ms*; 56.67%) had statistically significant differences.[18]

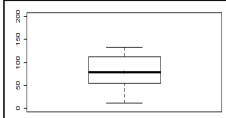
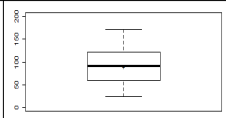
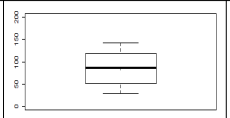
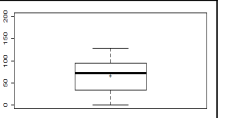
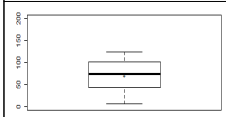
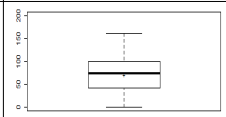
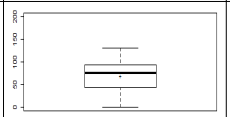
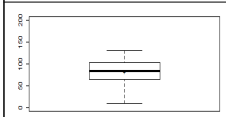
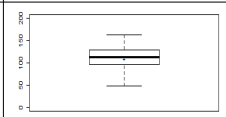
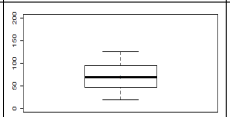
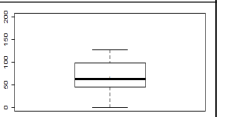
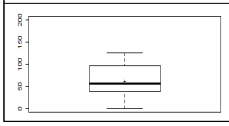
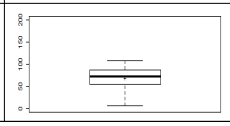
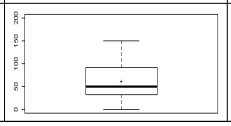
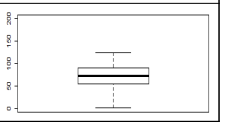
[18] In Lee (2013), only 19 pairs (15.83%) had statistically significant differences. Note that the number of pairs with significant differences increased in this paper. The reason seems to be that the number of informants drastically increased in this paper, which made the CIs narrower. It made little possibility to let the CIs overlap between two groups. For the relation between CIs and the number of data, see footnote 22.

Here, let's take the **N01-N02** pair as an example. In Lee (2013), it was found that the grammaticality of **N01** (*m*=102.50) was higher than that of **N02** (*m*=79.30) and that the difference was marginally significant (*p*=0.083). In this experiment, it was also found that the grammaticality of **N01** (*m*=128.20) was higher than that of **N02** (*m*=104.00) and that the difference was highly significant (*p*<0.001). Accordingly, we can say that the grammaticality of these two sentences are significantly different.

4.2 MACs

Table 4 illustrates the results of the grammaticality judgment task for the 16 types of MACs. As in MNCs, the mean value and little plus sign were provided in addition to the box plot for each type.

Table 4. Results of the grammaticality judgment task for the 16 types of MACs

A01 (*m*=80.28)	**A02** (*m*=89.77)	**A03** (*m*=85.60)	**A04** (*m*=66.38)
			
A05 (*m*=71.58)	**A06** (*m*=71.81)	**A07** (*m*=70.56)	**A08** (*m*=68.16)
			
A09 (*m*=82.07)	**A10** (*m*=109.76)	**A11** (*m*=70.45)	**A12** (*m*=65.93)
			
A13 (*m*=61.97)	**A14** (*m*=68.32)	**A15** (*m*=62.17)	**A16** (*m*=71.40)
			

As you could observe, the overall mean values of MACs were lower than those of

MNCs. Also note that the minimum scores of some types are very close to 0, especially from **A12** to **A16**.

In order to examine if the mean values became different depending on the semantic relations, a statistical test had to be performed. Since only one type of MAC follow the normal distribution, a Friedman test (the parametric counterpart of a repeated- measures ANOVA) was performed and the result was that the mean values became significantly differentiated depending on what semantic relations held between NP1 and NP2 in the MACs ($X2$=456.996, p<0.001).

Next, in order to examine the mean value of which type was significantly different from that of which type, a pairwise Wilcoxon test (the non-parametric counterpart of a Tukey's HSD test) was performed, and its results are shown in Table 5.

Table 5. Results of the pairwise Wilcoxon test for the 16 types of MACs

	A01	**A02**	**A03**	**A04**	**A05**	**A06**	**A07**	**A08**	**A09**	**A10**	**A11**	**A12**	**A13**	**A14**	**A15**
A02	×														
A03	×	×													
A04	***	***	***												
A05	*	***	***	***											
A06	***	***	***	×	×										
A07	***	**	***	*ms*	×	×									
A08	***	***	***	×	×	×	×								
A09	×	×	×	***	*ms*	×	×	**							
A10	***	***	***	***	***	***	***	***	***						
A11	***	***	***	×	×	×	×	×	**	***					
A12	***	***	***	×	×	×	×	×	***	***	×				
A13	***	***	***	×	***	*	×	×	***	***	**	×			
A14	*	***	×	×	×	×	×	×	***	***	×	×	×		
A15	***	***	***	*ms*	***	*	**	*	***	***	***	***	×	×	
A16	**	***	**	×	×	×	×	×	***	***	×	×	**	×	***

Among the 120 pairs, 67 pairs (except *ms*; 55.83%) had the significant differences.[19]

[19] In Lee (2013), only 45 pairs (37.50%) had statistically significant differences. Note that the

In Ryu (2013), it was pointed out that there were significant differences in the grammaticality judgment between two groups. One is from **A01** to **A10**, and the other is from **A11** to **A16**. In order to examine if there are statistically significant differences between the two groups, the data were collected into two groups separately. Group 1 was composed of the data from **A01** to **A10**, and Group 2 was from **A11** to **A16**. When the Shapiro-Wilk Normality Tests were performed for these two groups of data, both groups did not follow the normal distributions (Group 1: $p<0.001$; Group 2: $p<0.001$). Since both groups did not follow the normal distributions, a Mann-Whitney's U test (the non-parametric counterpart of an independent sample *t*-test) was performed, which was also known as a Wilcoxon Rank Sum Test. The results was that there are statistically significant differences between the two groups ($W=363898.5$, $p<0.001$). This result again supports Ryu's claim (2013) that the sentences from **A11** to **A16** are different those in other groups.

4.3 MNCs vs. MACs

Now, let's see how the semantic relations affected the grammaticality of MNCs and MACs. Figure 4 shows the comparison of the scores with 95% CIs in MNCs and MACs.

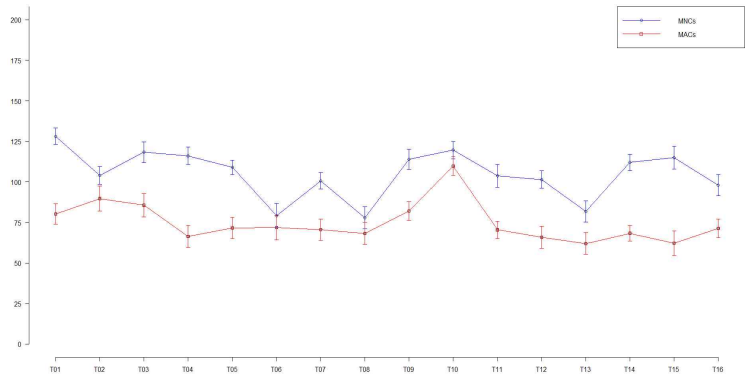


Figure 4. MNCs vs. MACs

number of pairs with significant differences increased also in the MACs. The reason seems to be the growth of informants, as in MNCs. For the relation between CIs and the number of data, see footnote 22.

As you can observe, there are some differences between each pair of type.

In order to examine if the distributions of MNCs were significantly different from those of MACs, a statistical test was performed. Since both data sets did not follow the normal distributions except **A06**, Wilcoxon signed-rank tests (the non-parametric counterpart of paired *t*-tests) were performed. The results showed that the distributions of MNCs were significantly higher than those of MACs (V=1093892, p<0.001). The analysis results for each pair are shown in Table 6.

Table 6. MNCs vs. MACs

	T01	**T02**	**T03**	**T04**	**T05**	**T06**	**T07**	**T08**
V	4738	3443	4282	4933	4729	2825	4262	3099
p	**.000**	**.000**	**.000**	**.000**	**.000**	.069	**.000**	**.000**
	T09	**T10**	**T11**	**T12**	**T13**	**T14**	**T15**	**T16**
V	4649	3149	4467	4863	3937	4913	4850	4411
p	**.000**	**.010**	**.000**	**.000**	**.000**	**.000**	**.000**	**.000**

Here, the *p*-value is bold-faced when it is less than 0.05, which means significant differences. As you can observe, only the **N06-A06** pair shows a marginal significance. Accordingly, we can say that the grammaticality of MNCs is significantly higher than those of MACs.

5. Discussions

5.1 General discussions on MCCs in Korean

In this paper, it was examined how native speakers' grammaticality judgment varies depending on the semantic relations of two consequative NPs in MCCs. Do these semantic relations affect the grammaticality of MCCs in Korean? If so, how? To answer this question, let's examine the analysis results in the experiment more closely.

Let's see the MNCs first. As mentioned in Section 4.1, all of MCCs examples in the experiment came from Ryu (2013). These data contained the typical semantic relation types which were frequently mentioned in previous studies including Yang (1972), Na and Huck (1993), and Park (2001). Notwithstanding, the grammaticality

judgment on these typical examples in MNCs did not constitute a homogeneous group, as you observed in the statistical tests in Section 4.1. This implies that the semantic relations of two consequative NPs surely affect the grammaticality of MNCs.

Let's see the distributions of MNCs more closely. Figure 5 demonstrates the grammaticality judgments of MNCs in the experiment.[20]

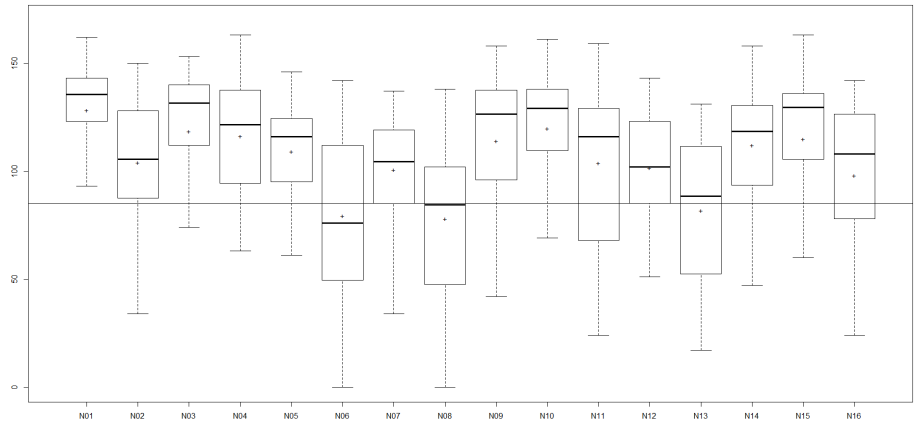


Figure 5. Grammaticality judgments of MNCs depending on the semantic relations

As in Lee (2013), the line in the middle is the one for the line length 85 mm. As mentioned in Section 3.1, the maximum length of lines allowed for the informants to draw was 170 mm. Therefore, we can divides the grammaticality judgment into just two parts based on this line: the positive zone (85 mm - 170 mm) and the negative zone (0 mm - 84 mm).

As you can see, the mean values of only the 3 types (**N06**, **N08**, and **N13**) are located in the negative zone, while the values for the other 13 types are in the positive zone. In addition, the median values (the second quantile which was identified by a thick black lines in the box plots) of only the 2 types (**N06** and **N08**)

20 From the box plots in Figure 5, we are able to observe that all of the MNCs don't follow the normal distribution. First, note that the median values (the black lines in the middle of box plots) are higher than the (arithmetic) means (the little + signs). This fact implies that the distributions are negatively skewed. Second, the lengths of the whiskers in the box plots are different. Usually, the lower whiskers are longer than the upper whiskers. These two facts demonstrate that more people are positive toward the MNCs.

are located in the negative zone, while the values for the other 14 types are in the positive zone. This implies that native speakers have positive positions toward MNCs, even though they did not feel the sentences are perfect.

Then, how about the MACs? Figure 6 demonstrates the grammaticality judgments of MACs in the experiment.[21]

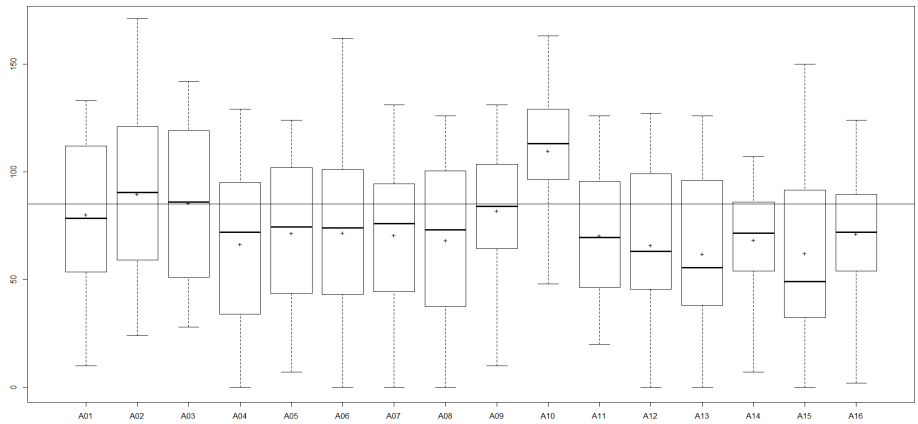


Figure 6. Grammaticality judgments of MACs depending on the semantic relations

As you can see, the mean values of most types are located in the negative zone. Even though the sentences from **A11** to **A16** are excluded from the discussion, the mean values of 7 types (**A01**, **A04**, **A05**, **A06**, **A07**, **A08**, and **A09**) are located in the positive zone. In addition, the median values of only the 3 types (**A02**, **A03**, and **N10**) are located in the positive zone, while the values for the other 13 types are in the negative zone. Also note that 7 types (**A04**, **A05**, **A06**, **A07**, **A08**, **A12**, **A13**, and **A15**) of MACs included the value 0, which means that the sentences are completely ungrammatical. This tendency was surely different from those of MNCs,

21 As in Figure 5, we are able to observe that almost all of the MACs do not follow the normal distribution from the box plots in Figure 6. First, note that the median values (the black line in the middle of box plots) are higher than the (arithmetic) means (the little + signs) in most types. This fact implies that the distributions are negatively skewed. However, the median values are lower than the means in 4 types (**A01**, **A12**, **A13** and **A15**). This fact implies that the distributions are positively skewed. Second, the lengths of the whiskers in the box plots are different. The lower whiskers are longer than the upper whiskers in some types, and the opposite tendencies appear in other types. These two facts demonstrate that the distributions of MACs don't follow the normal distribution.

where only 2 types had the value 0 (**N06** and **N08**). This implies that native speakers have negative positions toward MACs, even though they didn't feel the sentences are not ungrammatical.

Then, the next question is what the analysis results imply in the experiments. The box plots in Figure 5 and Figure 6 have the following implications in the studies of MCCs.

First, the distributions of data in Figure 5 and Figure 6 indicate that detailed examinations of syntactic phenomena have to come before their theoretical explanations of MCCs. Most previous approaches to these constructions have been theoretical in nature and have focused on how these constructions were made or what the semantic relations licensed their constructions. However, as you can see in Figure 5 and Figure 6, the native speakers' grammaticality was drastically different depending on the type of semantic relations which the MCCs had. For example, even though most previous studies implicitly assumed that both (1) and (2) are grammatical, the experimental results demonstrated that the grammaticality of (1) is much higher than that of (2). In addition, the ranges/variances of the grammaticality were various from type to type. For example, **N01** and **N10** have small variances in MNCs, whereas **N02**, **N06**, **N08**, **N11**, **N13**, and **N16** have large variances. In MACs, however, most types had large variances, especially in **A02**, **A06**, and **A15**. This means that some extent of agreements can be drawn from the former groups but that those kinds of agreements cannot be drawn from the latter groups. Accordingly, more studies are necessary on which factors would make these discrepancies in the native speakers' grammaticality to MCCs.

Second, the box plots in Figure 5 and Figure 6 imply that *magnitude estimation* has advantages over *category scaling* in the grammaticality judgment tasks. For example, as you can observe in Figure 5, the range values of the grammaticality were various from type to type. **N08** had the maximum range 138, and **N04** had the minimum range 100. The range value of **N08** is almost one and half times as big as that of **N04**. Likewise, **A06** had the maximum range 162, and **A14** had the minimum range 107. The range value of **A06** is almost one and half times as big as that of **N14**. The differences in the range values were able to be noticed here since *magnitude estimation* was adopted in the experiments. If *category scaling* with 5 or 7 steps had been used instead, these range differences could not be observed correctly or the differences could be smaller than the values gauged with *magnitude*

estimation. This fact demonstrates that *magnitude estimation* has advantages over *category scaling* in the grammaticality judgment tasks.

5.2 Comparison with Lee (2013)

As mentioned in Section 1, this paper is an extension of the empirical approach to MCCs, but with extended number of informants. Then, let's examine what the differences exist between two experiments, what kinds of effects were resulted in by the changes, and what are the common findings in these two studies.

What were the differences in the design of experiments? First, there were differences in the number of informants. Only 20 sets of data were used in Lee (2013), whereas this paper used 100 sets of data. The numbers of informants increased five times. Second, only one type of questionnaire was used in Lee (2013), while 5 different types of questionnaires were used in this paper. The questionnaires contained 5 distinguished orders of sentences. Third, there were differences in the randomization. The sentences were randomized only within MNCs or MACs in Lee (2013), but the randomization was performed across the distinctions in this paper.

Then, how did these changes affect the analysis results? Let's see Figure 7 and Figure 8. Figure 7 compares the analysis results in MNCs between two studies, and Figure 8 provides the comparison of the analysis results in MACs. In both figures, MNCs020 and MACs020 refer to the analysis results in Lee (2013), whereas MNCs100 and MACs100 to the analysis results in this paper.

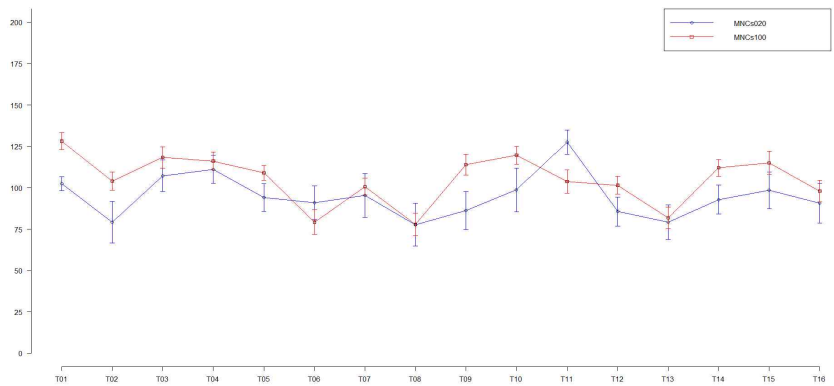


Figure 7. MNCs020 vs. MNCs100

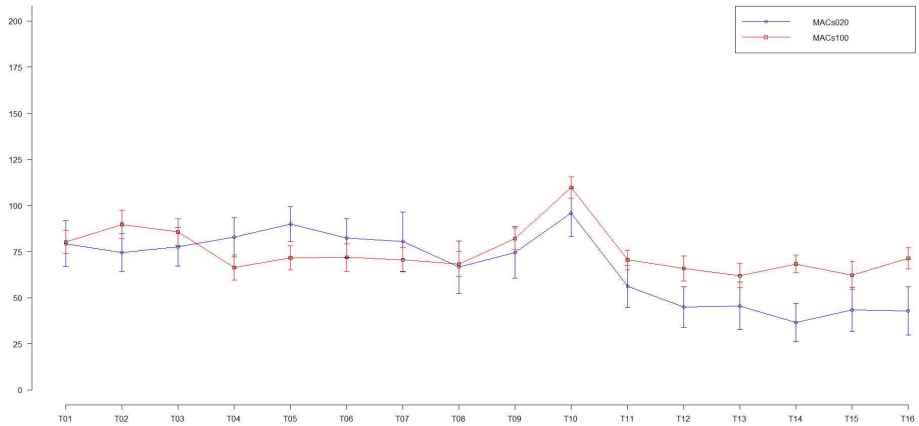


Figure 8. MACs020 vs. MACs100

All the data were provided with the mean values and their CIs.

As you can see, there are some differences between the analysis results in Lee (2013) and those in this paper. First, note that CIs became narrower in the latter analysis. It seems that the growth of data sets made the CIs narrower than the previous study.[22] Second, there are some cases where the mean values significantly higher or lower than those of previous studies. In both figures, if two CIs don't overlap, it means that there are significant differences in the mean values. In MNCs, 9 pairs of data sets (**N01**, **N02**, **N05**, **N09**, **N10**, **N11**, **N12**, **N14**, and **N15**) had no overlap in CIs. In MACs, 5 pairs of data sets (**A02**, **A05**, **A12**, **A14**, and **A16**) had no overlap in CIs. It seems that the growth of data sets made the changes. It is necessary to examine which analysis is correct with more data. Third, as you observed in Table 3 and Table 5, there were big differences in the pairswise comparison of MCCs. In MNCs, the number of significant differences increased from 19 pairs (15.83%) to 68 (56.67%). Likewise, In MACs, the number of significant differences increased from 45 (37.50%) to 67 pairs (55.83%). It seems that the growth of data sets also made these changes, since more data make less possibility of overlap in the CIs.

However, there are common findings in spite of these differences. First, we

22 Since *CI* is calculated as $CI=\bar{x}\pm SE$ (Gries, 2013:132) and *SE* is calculated as $SE=\sqrt{var/n}=sd/\sqrt{n}$ (Gries, 2013:129), *CI* is calculated as $CI=\bar{x}\pm sd/\sqrt{n}$. Accordingly, *CI* becomes narrower as the number of data *n* increases.

found that the distributions of grammaticality judgment in MCCs did not make a homogeneous and that it was significantly affected by the semantic relations (p<0.001 for both MNCs and MACs). Second, the grammaticality of MNCs is significantly higher than that of MACs (p<0.001). Third, native speakers have positive positions toward MNCs, while they have negative positions toward MACs. Fourth, the grammaticality of **A11-A16** is significantly different from those of the other groups (p<0.001). These properties did not change even in the growth of data sets. Therefore, we can say that these tendencies can be general properties of MCCs in Korean.

6. Conclusion

In this paper, we took an empirical approach and examined how the grammaticality of MNCs and MACs in Korean varies depending on the semantic relations which hold between two consequative NPs.

In order to examine how semantic relations affect the grammaticality of MCCs, an experiment was designed and performed. Unlike Lee (2013), the data for 100 informants were collected in the experiment. The grammaticality judgment tasks were designed following the guidelines in Johnson (2008), and native speakers' intuition was measured with two scales: *numeric estimates* and *line drawing*. After the intuition tests, the normality tests were performed on each set of the collected data. Since most data sets didn't follow the normal distribution, non-parametric tests were conducted.

Through the analysis, we found that the following facts. First, the distributions of grammaticality judgments in MCCs did not make a homogeneous group and that it was significantly affected by the semantic relations. Second, the grammaticality of MNCs is significantly higher than that of MACs. Third, native speakers have positive positions toward MNCs, while they have negative positions toward MACs. Fourth, the grammaticality of **A11-A16** is significantly different from those of the other groups.

These analysis results have some implications that the examples should have been given after careful investigations in these constructions, because not all the sentences got positive answers from the native speakers. The results also showed that

there must be systematic and scientific studies on the factors to decide the grammaticality of these sentences. Though more studies are necessary, the experimental design and the analysis methods are pre-requisite for the theoretical studies of MCCs.

References

Baayen, Harald. 2008. *Analyzing Linguistic Data*: *A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Bard, Ellen, Dan Robertson, and Antonella Sorace, 1996. Magnitude Estimation of Linguistic Acceptability. *Language* 72: 32-68.

Carnie, Andrew. 2012. *Syntax*: *A Generative Introduction*. 3rd Edition. Oxford: Blackwell.

Cho, Seongeun. 2000. *Three Forms of Case Agreement in Korean*. Ph.D. thesis, State University of New York, Stony Brook, New York.

Choe, Hyun-Bae. 1937. *Wulimalpon* (*The Korean Grammar*). Seoul: Top Publishing Company.

Choe, Hyun-Sook. 1987. Syntactic Adjunction, A-chains, and the ECP: Multiple Identical Case Construction in Korean. In Joyce McDonough and Bernadette Plunkett (eds.), *Proceedings of the 17th North East Linguistic Society* (*NELS-17*): 100-121.

Choi, Incheol. 2012. Sentential Specifiers in the Korean Clause Structure. In Steffan Müller (eds.), *Proceedings of the 19th International Conference on Head-Driven Phrase Structure Grammar* (*HPSG-19*): 75-85.

Choi, Incheol and Eunsuk Lee. 2008. The Partial Parallelism between Double Nominative Construction and ECM Construction in Korean. *Studies in Modern Grammar* 53: 75-101.

Cowart, Wayne. 1997. *Experimental Syntax*: *Applying Objective Methods to Sentence Judgments*. Thousands Oaks, CA: Sage Publications.

Han, Jeonghan. 1999. *On Grammatical Coding of Information Structure in Korean*: *A Role and Reference Grammar Account*. Ph.D. thesis. The State University of New York at Buffalo.

Hong, Ki-Sun. 1991. *Argument Selection and Case-Marking in Korean*. Ph.D. thesis. Stanford University.

Hong, Ki-Sun. 1997. Eynge-wa Kwuke-uy Insang Kwumwun Pikyo Pwunsek (Subject-to-object raising constructions in English and Korean). *Language Research* 33: 409-434.

Johnson, Keith. 2008. *Quantitative Methods in Linguistics*. Oxford: Blackwell.

Kang, Beom-Mo. 1988. *Functional Inheritance, Anaphora and Semantic Interpretation*. Ph.D.

thesis. Brown University.

Kang, Myung-Yoon. 1987. 'Possessor Raising' in Korean. In Susumo Kuno (eds.) *Harvard Studies in Korean Linguistics* 2: 80-88.

Keller, Frank. 2000. *Gradient in Grammar*: *Experimental and Computational Aspects of Degrees of Grammaticality*. Ph.D. thesis. University of Edinburgh.

Kim, Jong-Bok and Peter Sells. 2007. Two Types of Multiple Nominative Construction: A Constructional Approach. In Stefan Müller (ed.), *Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar* (*HPSG-14*): 364-372.

Kim, Jong-Bok, Peter Sells, and Jaehyung Yang. 2007. Parsing Two Types of Multiple Nominative Constructions: A Constructional Approach. *Language and Information* 11(1): 25-37.

Kim, Jong-Bok. 2000. A Constraint-based and Head-driven Approach to Multiple Nominative Constructions. In Dan Flickinger and Andreas Kathol (eds.), *Proceedings of the 7th International Conference on Head-Driven Phrase Structure Grammar* (*HPSG-7*): 166-181.

Kim, Jong-Bok. 2001. A Constraint-based Approach to Some Multiple Nominative Constructions in Korean. In Akira Ikeya and Masahito Kawamori (eds.), *Proceedings of the 14th Pacific Asia Conference on Language, Information, and Computation* (*PACLIC-14*): 165-176.

Kim, Jong-Bok. 2004. *Hankwuke Kwukwuco Mwunpep* (written in Korean, *Korean Phrase Structure Grammar*). Seoul: HankukMunhwasa.

Kim, Young-Joo. 1989. Inalienable Possession as a Semantic Relationship Underlying Predication: The Case of Multiple-Accusative Constructions. In Susumo Kuno (eds.) *Harvard Studies in Korean Linguistics* 3: 445-467.

Kim, Young-Joo. 1990. *The Syntax and Semantics of Korean Case*: *The Interaction between Lexical and Syntactic Levels of Representation*. Ph.D. thesis, Harvard University.

Kitahara, Hisatsugu. 1993. Inalienable Possession Constructions in Korean: Scrambling, the Proper Binding Condition, and Case-Percolation. In Hajime Hoji and Patrica Clancy (eds.) *Japanese/Korean Linguistics* 2: 394-408.

Lee, Seong-yong. 2007. Two Subject Positions in Multiple Nominative Constructions. *Journal of Language Sciences* 14(2): 239-262.

Lee, Yong-hun. 2013. An Experimental Approach to Multiple Case Constructions in Korean. *Language and Information* 17(2): 29-50.

Lodge, Milton. 1981. *Magnitude Scaling*: *Quantitative Measurement of Opinions*. Beverley Hills, CA: Sage Publications.

Maling, Joan and Soowon Kim. 1992. Case Assignment in the Inalienable Possession Construction in Korean. *Journal of East Asian Linguistics* 1: 37-68.

Moon, Gui-Sun. 2000 The Predication Operation and Multiple Subject Constructions in Korean: Focusing on Inalienable Possessive Constructions. *Studies in Generative*

Grammar 10(1): 239-263.

Na, Younghee. and Geoffrey. Huck. 1993. On the Status of Certain Island Violations in Korean. *Linguistics and Philosophy* 16(2): 181-229.

O'Grady, William. 1991. *Categories and Case*. Amsterdam: John Benjamins Publishing.

Park, Byung-Soo. 2001. Constraints on Multiple Nominative Constructions in Korean: A Constraint-based Lexicalist Approach. *The Journal of Linguistic Science* 20: 147-190.

Park, Hee-Moon. 2005. Constraint-based Analyses on the Korean Double Nominative Constructions. *The Journal of Studies in Language* 21: 87-111.

Park, Ki-Seong. 1995. *The Semantics and Pragmatics of Case Marking in Korean*: *A Role and Reference Grammar Account*. Ph.D. thesis. The State University of New York at Buffalo.

Rhee, Seongha. 1999. On the Multiple Nominative Constructions in Korean. In Young-hwa Kim Il-kon Kim, and Jeong-won Park (eds.) *Linguistic Investigations*, 398-490. Seoul: Hankuk Publishing Co.

Ryu, Byong-Rae. 2013. Multiple Case Marking Constructions in Korean Revisited. *Language and Information* 17(2): 1-27.

Schütze, Carson. 1996. Korean "Case stacking" Isn't: Unifying Non-case Uses of Case Particles. In Kiyumi Kusumoto (ed.), *Proceedings of the 26th North East Linguistic Society* (*NELS-26*): 351-365.

Schütze, Carson. 1996. *The Empirical Base of Linguistics*: *Grammaticality Judgments and Linguistic Methodology*. Chicago, IL: University of Chicago Press.

Schütze, Carson. 2001. On Korean "Case Stacking": The Varied Functions of the Particles *-ka* and *-lul*. *The Linguistic Review* 18: 193-232.

Stevenson, Stanley. 1975. *Psycholinguistics*: *Introduction to Its Perceptual, Neural, and Social Prospects*. New York: John Wiley.

Ura, Hiroyuki. 1996. *Multiple Feature-Checking*: *A Theory of Grammatical Function Splitting*. Ph. D. thesis. MIT.

Van Valin, Robert and Randy LaPolla. 1998. *Syntax*: *Structure, Meaning, and Function*. Cambridge: Cambridge University Press.

Van Valin, Robert and William. Foley. 1980. Role and Reference Grammar. In Eeith Moravcsik and Jesica Wirth (eds.) *Current Approaches to Syntax*: 329-352. New York: Academic Press.

Van Valin, Robert. 2009. Case in Role and Reference Grammar. In Andrej Malchukov and Andrew Spencer (eds.) *The Oxford Handbook of Case*: 102-120. Oxford: Oxford University Press.

Yang, In-Seok. 1972. *Korean Syntax*: *Case Markers, Delimiters, Complementation, and Relativization*. Ph.D. thesis. University of Hawaii.

Yoon, James. 1986. Some Queries Concerning the Syntax of Multiple Subject Constructions in Korean. *Studies in the Linguistic Sciences* 16: 215-236.

Yoon, James. 1989. The Grammar of Inalienable Possession Constructions in Korean, Mandarin and French. In Susumu Kuno (eds.) *Harvard Studies in Korean Linguistics* 3: 357-368.

Yoon, James. 2003. Raising Specifiers: A Macroparametric Account of SOR in Some Altaic Languages. A Paper Presented at the Workshop on Formal Approaches to Altaic Languages, MIT.

Yoon, James. 2004. Non-nominative (Major) Subjects in Korean. In Peri Bhaskararao and Karumuri Subbarao (eds.) *Non-nominative Subjects*: 265-314. Berlin: Mouton deGruyter.

Yoon, James. 2009. The Distribution of Subject Properties in Multiple Subject Constructions. In Yukinori Takubo, Tomohide Kinuhata, Szymon Grzelak, and Kayo Nagai (eds.) *Japanese/Korean Linguistics* 16: 64-83.

Yoon, Jeong-Me. 1997. The Argument Structure of Relational Nouns and Inalienable Possessor Constructions in Korean. *Language Research* 33(2): 231-64.

Appendix

※ The sample sentences were enumerated with their type numbers. They were provided in Korean, since Yale Romanization and adding the interpretations with glosses to each sentence take too much space.

N01. 토끼가 귀가 길다.
A01. 철수가 토끼를 귀를 잡았다.
N02. 이 함대가 잠수함이 많다.
A02. 적군이 이 함대를 잠수함을 박살냈다.
N03. 소금이 알갱이가 크다.
A03. 철수가 소금을 알갱이를 녹였다.
N04. 기아차가 강판이 두껍다.
A04. 철수가 기아차를 강판을 좋아한다.
N05. 골프가 퍼팅이 어렵다.
A05. 철수가 골프를 퍼팅을 좋아한다.
N06. 캘리포니아가 실리콘벨리가 따스하다.
A06. 철수가 캘리포니아를 실리콘벨리를 방문했다.
N07. 비행기가 에어버스가 크다.
A07. 철수가 비행기를 에어버스를 탔다.

N08. 귀가 귀고리가 너무 크다.
A08. 철수가 귀를 귀고리를 잡았다.
N09. 바지가 길이가 짧다.
A09. 철수가 바지를 길이를 잘랐다.
N10. 학생들이 두 명이 왔다.
A10. 철수가 학생을 두 명을 보내었다.
N11. 그 해변이 미인이 많다.
A11. 나는 그 해변을 미인을 좋아한다.
N12. 여름이 맥주가 맛있다.
A12. 나는 여름을 맥주를 좋아한다.
N13. 그 여자가 가방이 멋있다.
A13. 나는 그 여자를 가방을 좋아한다.
N14. 독일이 자동차가 튼튼하다.
A14. 나는 독일을 자동차를 좋아한다.
N15. 딸이 불평이 대단하다.
A15. 나는 딸을 불평을 미워한다.
N16. 그 의사가 환자가 많다.
A16. 나는 그 의사를 환자를 좋아한다.

Yong-hun Lee
Department of English Language & Literature
Chungnam National University
220 Gung-dong, Yuseng-gu, Daejeon 305-764, Korea
E-mail: yleeuiuc@hanmail.net

Received: 2014. 06. 29
Revised: 2014. 08. 19
Accepted: 2014. 08. 19