Linguistic Research 32(3), 671-691
DOI: 10.17250/khisli.32.3.201512.006

A multi-level analysis of subjecthood diagnostics in Korean*

Yong-hun Lee · Yeonkyung Park · Eunah Kim
(Chungnam National University · Hannam University · Seoul National University)

Lee, Yong-hun, Yeonkyung Park, and Eunah Kim. 2015. A multi-level analysis of subjecthood diagnostics in Korean. *Linguistic Research* 32(3), 671-691. There have been proposed several different diagnostics for identifying the subject in Korean. Among the previous studies, Hong (1991/1994) investigated 9 subjecthood diagnostics and classified them into a few groups. This paper is based on previous experimental studies and statistically examines how Hong's classifications affect each subjecthood diagnostics and the acceptability scores of native speakers. Among the 9 subjecthood diagnostics, the experimental data for 6 subjecthood diagnostics were taken and they were statistically examined. This paper organized 6 subjecthood diagnostics into 3 groups following Hong's classifications, and it statistically examined how each group and each diagnostics affected the acceptability scores of native speakers. A multi-level/hierarchical Generalized Linear Mixed-effects (Regression) Model was adopted to statistically examine the influence of each group of diagnostics. Through the analysis, the following two facts were observed: (i) Each group of classification did NOT play a significant role in the subjecthood diagnostics, and (ii) *Honorific Agreement* and *Plural Copying* demonstrated statistically different behaviors from the other 4 subjecthood diagnostics. **(Chungnam National University · Hannam University · Seoul National University)**

Keywords subject, subjecthood diagnostics, experimental, hierarchy, GLMM

1. Introduction

Subject is one of the important grammatical relations (GRs) which every natural language has. It is syntactically and semantically important to identify the subject in the languages, since subject is one of the arguments and the argument structure of sentences can be directly translated into predicate logic. There have been a lot of studies on the subjecthood diagnostics, and several different kinds of diagnostics

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

have been proposed for the identification of subject in Korean. However, there have been only a few studies which classified the subjecthood diagnostics and examined the validity of the classification. Hong (1991/1994) classified 9 subjecthood tests into 5 groups and examined the validity of each diagnostics in the identification of subject in Korean.

Nowadays, as experimental methods are developed in syntax, there are several attempts to examine the validity of subjecthood diagnostics through the experimental designs. As Sprouse and Hornstein (2013:3) clearly pointed out, “experimental syntax provides a set of tools that go beyond the traditional acceptability judgment experiments that have been used (to good success) in the existing literature.” Following this idea, Kim J. et al. (2015), Lee et al. (2015) and Kim E. et al. (in progress) experimentally examined the validity of 6 subjecthood diagnostics for both Single Subject Constructions (SSCs) and Multiple Subject Constructions (MSCs). These studies pointed out that some diagnostics were available for both SSCs and MSCs but that others were possible for only SSCs (not for MSCs). They proved these facts through experimental designs and statistical analyses.

This paper has taken these three experimental studies and examined the subjecthood diagnostics with a broad perspective. In this paper, the 6 subjecthood diagnostics were organized into a hierarchy based on Hong's classification (ibid.), and it was examined how the grouping factor and each diagnostics affected the acceptability scores of native speakers. Since the 6 subjecthood diagnostics were organized into a hierarchical structure, a multi-level/hierarchical Generalized Linear Mixed-effects (Regression) Model was applied. Through the analysis, the following two facts are observed: Through the analysis, the following two facts were observed: (i) Each group of classification did NOT play a significant role in the subjecthood diagnostics, and (ii) *Honorific Agreement* and *Plural Copying* demonstrated different behaviors from the others.

This paper is organized as follows. In Section 2, previous studies were reviewed focused on subjecthood tests in Korean, experimental approaches to the diagnostics, and Hong's classification. In Section 3, the 6 subjecthood diagnostics were organized into a hierarchy based on Hong's classification (ibid.), and it was examined how the grouping factor and each diagnostics affected the acceptability scores of native speakers. Section 4 showed us the analysis results, and Section 5 includes discussion. Section 6 summarizes this paper.

2. Previous studies

2.1 Subjecthood diagnostics in Korean

Though the subjects are often accompanied with the Nominative Case marker -*i*/*ka* in Korean, the Case marker is not enough to identify the subject of the Korean sentences. Let us consider the following examples.

(1) a. *Cheli-ka* *musep-ta*.
Cheli.NOM scary.DECL
'Cheli is scary.'
b. *Cheli-nun* *musep-ta*.
Cheli.TOP scary.DECL
'Cheli is scary.'
c. *Cheli-ka* *emeni-ka* *musewu-si-ta*.
Cheli.NOM mother.NOM scary.HON.DECL
'Cheli's mother is scary.'
d. *Cheli-eykey* *ton-i* *manh-ta*.
Cheli.DAT money.NOM many.DECL
'Cheli has much money.'

Cheli has a Nominative Case marker, and the NP *Cheli-ka* becomes a subject in (1a). On the other hand, *Cheli* has a Topic marker -*nun* in (1b), but it becomes a subject. In (1c), both NPs *Cheli-ka* and *emeni-ka* have Nominative Case markers. Then, the question is which one is a subject? *Cheli-ka*, *emeni-ka*, or both? In (1d), the NP *ton-i* has a Nominative Case marker, but it is not a subject. Instead, the NP *Cheli-eykey* becomes a subject. These examples show us that the Nominative Case marker is not enough to identify the subject of Korean sentences. Here, (1a) and (1b) are called SSCs, (1c) is an MSC, and (1d) is named a Dative Subject Construction.

Accordingly, there have been a lot of studies on the status of subject in the syntactic literature of Korean. In order to identify the subject NP in a sentence, there have been proposed several different types of linguistic tests, which were called 'subject diagnostics.' The main goal of these diagnostics was to pick up a subject NP through the examining the subject properties of the candidate NPs. The following list summarizes the proposed subjecthood diagnostics (Yoon, 2009:14).[1]

(2) Proposed Subject Diagnostics in Korean
  a. Nominative Case marking
  b. Controller of optional plural-marking
  c. Controller of subject honorification
  d. Target of subject-to-object raising
  e. Target of control
  f. Controller of PRO in complement (obligatory) control
  g. Controller of PRO in adjunct control
  h. Controller of coordinate deletion
  i. Antecedent of (subject-oriented) anaphors
  j. Exhaustive-listing interpretation of '-*ka*/-*i*'

Though these diagnostics may behave similarly in the SSC environments, it is known that they behave differently in the MSC environments. For example, if two NPs in MSCs were called NP1 and NP2 respectively, Yoon (2004, 2007, 2009) claimed that some of these diagnostics preferred to pick up the Major Subject (MS; NP1 in MSCs) and others picked up the Grammatical Subject (GS; NP2 in MSCs).[2] These studies also claimed that, among the above diagnostics, four tests ((2b), (2c), (2f), and (2i)) were the typical diagnostics for GS, and another four tests ((2a), (2d), (2g), and (2h)) were those for MS.

2.2 Experimental approaches

Although there have been several studies on the subjecthood diagnostics in the theoretical syntax in Korean, there have been only a few studies which examined the subjecthood diagnostics through the experimental designs. Recently, following the tradition of experimental syntax, three studies (Kim, J. et al., 2015; Lee et al., 2015; Kim E. et al., in progress) closely examined how the subjecthood diagnostics in (2) behave differently in the SSC and MSC environments through the experimental designs.

The three experimental studies chose two subjecthood diagnostics in each paper and examined their behaviors in the SSC and MSC environments respectively.

1 This list is constructed based on Yoon (1986), Hong (1991/1994), Youn (1990), etc.

2 In the MSC example of (1c), *Cheli-ka* corresponds to NP1 and *ememni-ka* to NP2.

Among the 10 diagnostics in (2), only 6 subjecthood diagnostics were examined in the three studies. They were (2b), (2c), (2f), (2g), (2h), and (2i).

The three studies took the same identical design. First, they made the following template for the target sentences.

(3) Template for Target Sentences
a. Type 1: $[NP1]_{gen}$ $[NP2]_{nom}$ [SSC, NP1]
b. Type 2: $[NP1]_{gen}$ $[NP2]_{nom}$ [SSC, NP2]
c. Type 3: $[NP1]_{nom}$ $[NP2]_{nom}$ [MSC, NP1]
d. Type 4: $[NP1]_{nom}$ $[NP2]_{nom}$ [MSC, NP2]

For example, the following sentences were used for the diagnostics (2c) *Honorific Agreement*.

(4) a. *Kim Kyouswunim-uy cengwen-i alumtawu-si-ta.*
Kim professor.GEN garden.NOM beautiful.HON.DECL
'Professor Kim's garden is beautiful.'
b. *Cheli-uy emeni-ka musewu-si-ta.*
Cheli.GEN mother.NOM scary.HON.DECL
'Cheli's mother is scary.'
c. *Kim Kyouswunim-i cengwen-i alumtawu-si-ta.*
Kim professor.NOM garden.NOM beautiful.HON.DECL
'Professor Kim's garden is beautiful.'
d. *Cheli-ka emeni-ka musewu-si-ta.*
Cheli.NOM mother.NOM scary.HON.DECL
'Cheli's mother is scary.'

As you can observe in these sentences, (4a) and (4b) contain SSCs but (4c) and (4d) have MSCs. On the other hand, in (4a) and (4c), the predicate *alumtawu-si-ta* agrees with the NP1 *Kim Kyouswunim*. However, in (4b) and (4d), the predicate *musewu-si-ta* agrees with the NP1 *emeni*. These example sentences illustrate the fact that the template in (3) is fully reflected in four sentences in (4).

After the base sentences were made as in (4), the studies made 4 more different sets of sentences where the sentences had the same format but they were constructed with different lexical items. The purpose of this strategy is to make the choice of lexical items neutral to the acceptability scores of sentences. For example, if the choice of *Cheli* or *cengwon* 'garden' affected the acceptability scores, it would not be possible to say that the acceptability scores of sentences became different depending on the choice of constructions. Therefore, 4 more different sets of sentences were constructed in the target sentences so that the lexical choices cannot affected the acceptability scores of sentences. Through the process, a total of 20 target sentences were made for the experiment. Then, another 20 filler sentences were also made, where the relevant factor (the presence and absence of the diagnostics) was eliminated. The target and filler sentences were prepared for each diagnostics, and a total of 240 sentences (40 sentences per each diagnostics × 6 diagnostics) were prepared in the experiment. The 240 sentences were randomly ordered and they were given to the participants.

The data for 70 native speakers were collected through the experiment. Their intuition was measured with the Magnitude Estimation (Stevenson, 1975; Lodge, 1981; Bard et al., 1996; Schütze, 1996; Cowart, 1997; Keller, 2000; Johnson, 2008). Among the Magnitude Estimation methods, the line drawing was taken, where the participants drew the line which corresponded to the acceptability of the sentences.

After all the data were collected, the data were statistically analyzed. Since most of the data sets did not follow the normal distributions, generalized linear regression analyses were applied (Bayeen, 2008; Gries, 2013). The goal of the analyses was to examine if and how the six subjecthood diagnostics behaved in the SSC and MSC environments. In order to achieve this goal, these studies compared the acceptability scores of two types of sentences. The predicate showed the agreement with the NP1 in one group of sentences, and the same predicate showed the agreement with the NP2 in the other group. Then, they calculated the *p*-values for the differences on two groups of sentences. The following table illustrates the analysis results.

Table 1. Analysis results of six subjecthood diagnostics

Diagnostics	SSC	MSC
Honorific Agreement	p<.001	p=.323
Plural Copying	p=.058	p=.152
Obligatory Control	p<.001	p<.001
Reflexive Binding	p<.001	p<.01
Coordinated Deletion	p<.001	p=.113
Adjunct Control	p<.05	p=.615

This analysis results demonstrated the following facts. First, the 5 diagnostics (except *Plural Copying*) showed a statistically significant difference in SSCs (p<.05). This means that these 5 tests can be used as subjecthood diagnostics for SSCs. That is, these 5 tests correctly pick up the subject for the given sentences. Second, even though *Plural Copying* is not statistically significant, it is *marginally* significant (p=.058). Third, the 2 diagnostics (*Obligatory Control* and *Reflexive Binding*) showed a statistically significant difference in MSCs (p<.05). It means that these 2 tests can be used as subjecthood diagnostics for MSCs. Fourth, the other 4 tests are not statistically significant. This means that these 4 diagnostics (*Honorific Agreement*, *Plural Copying*, *Coordinated Deletion*, and *Adjunct Control*) cannot be used as a subjecthood diagnostics in the MSC environments, since they can pick up a wrong NP as a subject. Overall, the analysis results show that 5 tests can be used as subjecthood diagnostics for SSCs, but only 2 tests can be used as subjecthood diagnostics for MSCs.

2.3 Hong's classification

Hong (1991/1994) investigated various subjecthood diagnostics and classified into a few groups, based on the properties of each diagnostics. Hong (ibid.) scrutinized 9 subjecthood diagnostics with the relevant examples and classified them into 5 groups as follows. For the comparison with the terminology in this paper, another column was provided for the terminology of this paper.

Table 2. Hong's classification of subjecthood diagnostics

Diagnostics Group	Hong (ibid.)	This Paper
Grammatical Subject	Honorification	Honorific Agreement
	Equi Control	Obligatory Control
Consciousness and Predication Subject	*Caki* Binding	Reflexive Binding
	Plural Marker Copying	Plural Copying
Discourse Topichood	Coordination	Coordinated Deletion
	Control in -Myense(to) 'Although' Adverbial Clauses	Adjunct Control
	Subject-to-Object Raising	
Word Order	Plain Topicalization	
Anti-Redundancy of Semantic Information	Quantifier Float	

Hong (ibid.) did not provided the exact reason why the 9 diagnostics can be classified into the above 5 groups. However, the basis of this grouping can be inferred from the following phrases: “These groups are related to the notion of grammatical subject in various ways: first, grammatical subject takes first priority to be interpreted as a discourse topic; second, grammatical subject is placed sentence-initially in unmarked word order; third, grammatical subject takes nominative case in many instances. Based on these results, I will claim that honorification and the equi control construction are what we can rely on as necessary and sufficient subjecthood tests: they pick out all and only grammatical subjects. *Caki* binding and plural marker copying pick out all the grammatical subjects and something else. Accordingly, if something is a grammatical subject, it should be picked out by these tests; but not vice versa. Coordination, control in -*myense* clauses, subject-to-object- raising, plain topicalization, and Quantifier float do not bear any direct relationship to subjecthood, although they tend to pick up subjects in many cases, for different reasons as mentioned above.” (Hong, 1994:99-100)

That is, the grouping structure in Table 2 can be explained as follows. First, Honorification and Equi Control can form a group since they are necessary and sufficient condition in the subjecthood status. That is, they pick out all and only grammatical subjects. Second, *Caki* Binding and Plural Marker Copying can be grouped, since they sometimes pick out the grammatical subjects (but not always). Coordination, Control in -*Myense*(*to*) 'Although' Adverbial Clauses, Subject-to-Object

Raising can be a group, since they are sufficient condition for subjecthood. Plain Topicalization and Quantifier Float constitute separate groups respectively, since they are related with word order variation and semantic information respectively.

3. A statistical analysis

3.1 Toward a multi-level generalized linear mixed-effects model

Since the goal of the statistical analysis in this paper was to examine whether the classifications in Hong (ibid.) were valid, the first thing to do was to organize the hierarchy of subjecthood diagnostics. This paper took the 6 tests in Section 2.2 and constructed a hierarchy as follows.

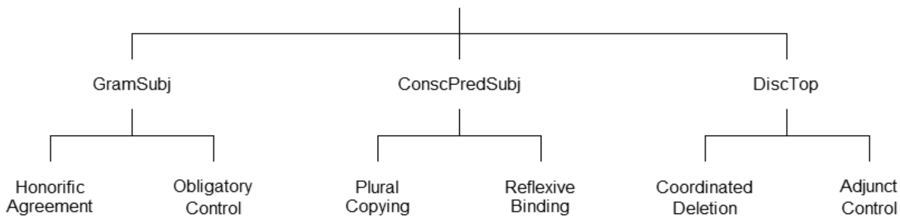

Figure 1. Hierarchical structure of 6 subjecthood tests

A statistical analysis was performed to examine if the grouping factor of diagnostics affected the members of each group and the acceptability scores of target sentences. Then, why is it necessary? If every member in each group shows similar behaviors, it implies that there is another factor that is located above the diagnostics and affects the behaviors of diagnostics. If not, it implies that there is no such factor and that each diagnostics behaves independently. If such factor exists and it is linguistically meaningful, the factor can be used in the subjecthood tests in Korean.

The statistical model which was adopted in this paper is a multi-level/hierarchical Generalized Linear Mixed Effect Model (Multi-level GLMM). This model was chosen through the following steps.

First, since the acceptability scores were collected for 6 different diagnostics, a Linear Regression analysis had to be adopted. Second, because the collected data sets did not follow the normal distributions, a Generalized Linear Regression Model had to be used with a Gaussian distribution. Third, now that some factors (such as SSCs vs. MSCs) were applied to all of the data but others were applied to only some part of the data (such as each diagnostics and the groupings), a Generalized Linear Mixed Effect (Regression) Model (GLMM) had to be used. Fourth, as 6 diagnostics were organized into a hierarchy in Figure 1, a multi-level/hierarchical Generalized Linear Mixed Effect (Regression) Model (multi-level GLMM) had to be used (Gelman and Hill, 2006).

In this study, we followed a stepwise model selection procedure which was outlined in Zuur et al. (2009, Chapter 5). This procedure can be summarized as follows: First, we began with a model which contained all of the fixed and random effects. Second, we tried to find the optimal random effects structure (varying intercepts for one or more factors and/or varying slopes for one or more factors). In the last step, because the optimal random effects structure has been found, we tried to find the optimal fixed effects structure. In the last two steps, an optimal model could be chosen according to some criterion such as significance testing/*p*-values or information criteria. The *p*-values could be used when a model m_1 contains (i) only random effects that made m_1 significantly better than the model m_2 where these effects were not included and (ii) only fixed effects (including their interactions) that make the model m_1 significantly better than the model m_2 where these effects and their interactions were not included.

3.2 Initial model

Our statistical analysis started from the description of variables, which were used in the analysis. The following table enumerates the variables used in this paper.

Table 3. Variables and their description

Variables	Description
DiagType	3 types of diagnostics (Hong's classification)
Diagnostics	6 diagnostics
AgreedNP	Whether the predicate shows an agreement with NP1 or NP2
ConstType	SSCs vs. MSCs
Score	Acceptability score (line drawing)

Among the variables, *Score* was the dependent variable, and the others became independent variables.

The following step was to decide which variables belonged to the fixed effects and which ones were the random effects. These were decided as follows. First of all, since *ConstType* was encoded in all of the data (i.e., all the data in our data sets must have the values for this factor), it became a fixed effect. The other three variables (*DiagnType*, *Diagnostics*, and *AgreedNP*)[3] were applied only some parts of the data set (i.e., *Honorific Agreement* was applied only to the sentences with honorific markers), these three variable belongs to the random effects.

Before the model for our data sets was made, there was one more thing to consider. As you can find, Figure 1, each *Diagnostics* has 2 values (*AgreedNP*; agreement with NP1 and agreement with NP2) and 6 *Diagnostics* were organized into *DiagType*, the hierarchical structures of these three variables must be reflected in the statistical analysis as *DiagType/Diagnostics/AgreedNP*. Consequently, all the variables in Table 3 were classified as follows.

Table 4. Classification of variables

Fixed Effect	Random Effect	Dependent Variable
ConstType	DiagType/Diagnostics/AgreedNP	Score

Note that a hierarchical structure exists in the random effect.

Based on these classifications, our initial model was constructed as follows. Here, note that the fixed effect was represented without a parenthesis and that the random effects were represented with a parenthesis.

Table 5. Initial model

Model	Description
model00	Score~ConstType+(1|DiagType/Diagnostics/AgreedNP)
model01	Score~ConstType+(1|Hierarchy1)+(1|Hierarchy2)+(1|Hierarchy3) where Hierarchy1=DiagType, Hierarchy2=DiagType:Diagnostics, and Hierarchy3=DiagType:Diagnostics:AgreedNP

3 In fact, whether the predicate shows an agreement with NP1 or NP2 (*AgreedNP*) was encoded for all the data. However, which diagnostics was applied was different. That's why *AgreedNP* was organized below the *Diagnostics*.

The intended model was *model00*. This model could be re-written as in *model01*, where each hierarchical factor was scrutinized separately. That is, *model00*=*model01*. The model *model01* was used in the actual analysis so that the statistical values for each hierarchical factor could be obtained.

3.3 Finding an optimal model

Since the initial model was constructed, the next step was taken. Following the guideline in Zuur et al. (2009), the model selection procedure was performed. If the *p*-value of the ANOVA tests of two models was smaller than 0.05 ($p<0.05$), this means that two models are significantly different. This implies that the deleted factor plays an important role in the model selection and that the factor must NOT be deleted in the model. If the *p*-value was greater than 0.05 ($0.05<p$), this means that two models are not significantly different. This implies that the deleted element plays little role in the model selection and that the element CAN be deleted in the model. Through the model selection process, the following final model was obtained.

Table 6. Final model

Model	Description
model02	Score~ConstType+(1|Hierarchy2)+(1|Hierarchy3)

As you can observe in Table 6, the top hierarchy (i.e., Hierarchy1) was eliminated from the model. It means that the grouping factor of subjecthood diagnostics in Figure 1 was statistically NOT significant. However, Table 6 also demonstrated that each diagnostics (*Diagnostics*) and the agreement with which NPs (*Diagnostics*: *AgreedNP*) are statistically significant.

Since the final model was obtained, it was possible to calculate the detailed statistical (numerical) values for the final model. Table 7 shows us the detailed statistics.

Table 7. Statistics for the final model

Fixed	Coefficient	*se*	*t*	*p*	Random	*sd*
(intercept)	33.081	2.702	12.24	<.001	Hierarchy2	5.444
ConstType	17.231	0.517	33.36	<.001	Hierarchy3	9.275

As you can witness, the fixed factors (intercept) and *ConstType* were statistically significant. Also note that there is some variations in two random factors (0<*sd*).

4. Analysis result

4.1. Fixed effect: ConstType

Now that the final model was obtained, let us examine each fixed and random effects closely. The first predictor/variable that we have to consider is the fixed factor *ConstType*. The following graph demonstrates the effect plot for this factor. Here, the I-shaped lines represent 95% confidence intervals (CIs).

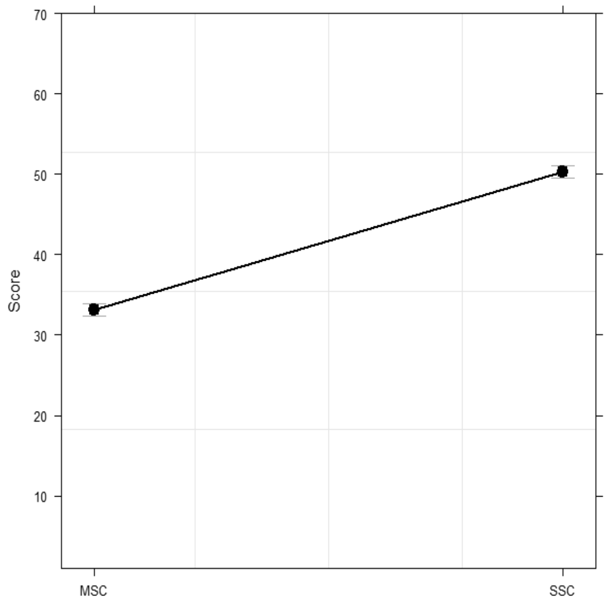

Figure 2. Effect plot for the factor *ConstType*

As this plot represents, the acceptability scores were highly affected depending on whether the sentence contained in a SSC environment or a MSC one. Note that the 95% CIs did NOT overlap. The statistical difference was also observed in the *p*-value of this factor in Table 7 (p<.001).

4.2. Random effects: Diagnostics and their groupings

Now, let us move on to the random factors. As Table 6 and Table 7 indicated, the grouping factor of subjecthood tests in Figure 1 was statistically NOT significant (.05<p), but each diagnostics was statistically significant (p<.05).

However, it is not the whole of the story. Even though the grouping factor of subjecthood diagnostics was not statistically significant, the factor might play a role. In order to examine how each grouping factor affected the subjecthood diagnostics, the following graph was generated based on the coefficient values of each member of *DiagType* and each *Diagnostics*.[4]

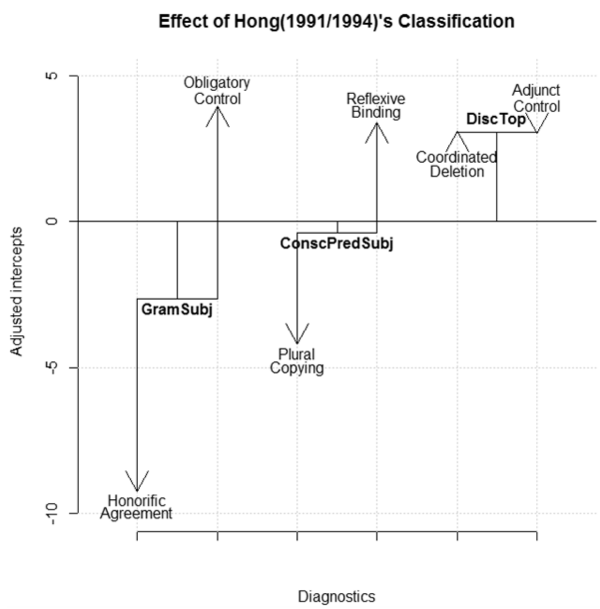

Figure 3. Effect of hierarchical structure

4 The values for *Hierarchy3* were not included here, since they make the plot too complicated.

The graph in Figure 3 can be interpreted theoretically as follows. Figure 3 visualizes the effect *DiagType/Diagnostics*. Two important observations should be made. First, compared to the overall baseline, the estimates of *DiagType* for *DiscTop* have to be adjusted upwards (by 3.05) whereas the estimates for *GramSubj* have to be adjusted downwards (by −2.64). This reflects the fact that the diagnostics in *DiscTop* had a higher probability to make a correct choices than the diagnostics in *GramSubj*. Second, in addition to *DiagType*, there are also adjustments for each *Diagnostics* that have to be added to those for *DiagType*. Crucially, these adjustments are quite large compared to those for *DiagType*, which shows that the *Diagnostics* distinction is more influential than the *DiagType* distinction or, from the opposite perspective, that the between-*Diagnostics* differences are larger than the within-*DiagType* differences. Those findings confirm that, qualitatively, *Diagnostics* played more crucial role than *DiagType* in the subjecthood tests.

More specifically, the plot in Figure 3 can be interpreted as follows. As Pinheiro and Bates (2000) mentioned, the *y*-values in this graph refers to the deviance of coefficients from the average. The horizontal line 0 indicates the point where the average value of all the coefficients for the random factors is located. This line can be the base line of the comparison. If a value is located above this base line, it indicates that the coefficient for the given random factor is bigger than average. In turn, this implies that the given random factor demonstrates a higher performance than other (random) factors. In our case, this means that the given random factor is good at identifying the subject in the sentences. On the other hand, If a value is located below this base line, it indicates that the coefficient for the given random factor is smaller than average. In turn, this implies that the given random factor demonstrates a lower performance than other (random) factors. In our case, this means that the given random factor is NOT good at identifying the subject in the sentences.

In the plot, the values for each diagnostics are indicated by the arrows, and those for each group are represented by horizontal lines which covers the members of the group. That is, the plain characters and arrows indicate the influence of each diagnostics, and the bold-faced words and horizontal lines indicate the influence of each group. Each value indicates how much each coefficient of the random factors fluctuated from the average (base line 0).

Let us first examine the influence of each diagnostics group. Figure 3 contains three group names: *GramSubj*, *ConscPredSubj*, and *DiscTop*. In Figure 3, only *DiscTop* is located above the base line 0, and *GramSubj* is located below the base line. This implies that the former group of diagnostics had a higher coefficient than average, which means that the subjecthood diagnostics in this group were good at identifying the subject in the sentences. On the other hand, the latter group of diagnostics had a lower coefficient than average, which means that the subjecthood tests in this group were not good at identifying the subject in the sentences. The value for *ConscPredSubj* had the value very close to 0, which implies that the subjecthood diagnostics in this group showed the average performance.[5]

Now, let us go to each group and examine each diagnostics method more closely.

The first group is *GramSubj*, which includes *Honorific Agreement* and *Obligatory Control*. Among these two diagnostics, *Obligatory Control* is located above the base line for *GramSubj*, but and *Honorific Agreement* is located below the line. This implies that the former had a higher coefficient than average, which means that this diagnostics had a higher performance in identifying the subject of the sentence. On the other hand, the latter diagnostics have a lower coefficient than average, which means that the method had a lower performance in identifying the subject of the sentence.

The second group is *ConscPredSubj*, that includes *Plural Copying* and *Reflexive Binding*. Among these two diagnostics, *Reflexive Binding* is located above the base line for *ConscPredSubj*, but and *Plural Copying* is located below the line. This implies that the former had a higher coefficient value than the average values of this group, which means that this method was good at identifying the subject of the sentence. On the other hand, the latter diagnostics have a lower coefficient value than the average of this group, which means that it is NOT good at identifying the subject of the sentence.

The third group is *DiscTop*, which includes *Coordinated Deletion* and *Adjunct Control*. Among these two diagnostics, *Coordinated Deletion* is located above the base line for *DiscTop*, but and *Adjunct Control* is located below the line. This implies that the former had a higher coefficient than average, which means that this

---

5 The reason why the analysis results in Table 1 and those in Figure 3 is a little different is that both SSCs and MSCs were separately analyzed in Table 1 but that they were analyzed together in Figure 3.

test had a higher performance in identifying the subject of the sentence. On the other hand, the latter diagnostics have a lower coefficient than average, which means the diagnostics had a lower performance in identifying the subject of the sentence.

Here, note that the variation of two diagnostics (*Coordinated Deletion* and *Adjunct Control*) from the line for *DiscTop* was smaller than the deviation of *DiscTop* from the base line 0. It implies that the two tests in *DiscTop* demonstrated similar performances, which means that they can be classified into a single group. On the other hand, the variations of two diagnostics in the groups *GramSubj* and *ConscPredSubj* were much bigger than the deviation of each group from the base line 0. It implies that the two tests in *GramSubj* and *ConscPredSubj* did not show similar performance, which means that they CANNOT be classified into a single group.

5. Discussions

In this paper, the following was examined through the statistical analysis: if and how the grouping factor of subjecthood diagnostics affected each test and the acceptability scores of native speakers. Our statistical analysis revealed the following facts.

First, as Figure 2 showed us, it was found that the acceptability scores were highly affected depending on whether the sentence contained a SSC or a MSC environment. The acceptability scores in the SSC environments were much higher than those in the MSC sentences. It is not a surprising result, since previous experimental studies also showed this fact.

Second, Table 6 and Table 7 illustrated that the grouping factor of subjecthood diagnostics was statistically NOT significant. Figure 3 provided the reason why the grouping factor was statistically in significant. The subjecthood diagnostics in *DiscTop* demonstrated smaller deviation from the base line for this group, whereas the tests in *GramSubj* and *ConscPredSubj* showed much larger deviation from the base line for each group. These larger deviations from the base line resulted in the insignificant influences of the grouping factor.

One more interesting fact is that Figure 3 demonstrated different tendencies that Hong (ibid.) predicted. If we follow the reasoning of Hong (1994:99-100), two diagnostics in *GramSubj* (*Honorific Agreement* and *Obligatory Control*) had to demonstrated the best performance, since Hong (1994:131) said that "*Honorification* and *equi control* pick out all and only subjects: so they may be used as necessary and sufficient subjecthood tests." Then, two diagnostics in *ConscPredSubj* (*Plural Copying* and *Reflexive Binding*) followed the tests in *GramSubj*, and two diagnostics in *DiscTop* (*Coordinated Deletion* and *Adjunct Control*) had to demonstrate the worst performance. However, the results in Figure 3 illustrated the opposite tendency.[6] It is a little surprising.

Then, why did these phenomena happen? Let us consider the following examples.[7]

(4) c. *Kim Kyouswunim-i cengwen-i alumtawu-si-ta.*
Kim professor.NOM garden.NOM beautiful.HON.DECL
'Professor Kim's garden is beautiful.'

d. *Cheli-ka emeni-ka musewu-si-ta.*
Cheli.NOM mother.NOM scary.HON.DECL
'Cheli's mother is scary.'

(5) a. *Ku haksayng-tul-i muncey-ka acwu-tul manh-ta.*
the student.PL.NOM problem.NOM very.PL many.DECL
'The students have a problem.'

b. *Ku yein-i ai-tul-i acwu-tul manh-ta.*
the woman.NOM child.PL.NOM very.PL many.DECL
'The woman has many children.'

In these sentences, the predicates showed an agreement with two different NPs. The predicate *alumtawu-si-ta* agreed with NP1 *Kim Kyouswunim-i* in (4c), but the predicate *musewu-si-ta* showed an agreement with NP2 *emeni-ka* in (4d). Likewise, the predicate *acwu-tul manh-ta* agreed with NP1 *Ku haksayng-tul-i* in (5a), but the same predicate showed an agreement with NP2 *ai-tul-i* in (5b). These different choices of NPs seemed to make these two diagnostics unreliable as subjecthood tests.

6 The statistical analysis was conducted again (step by step) since the analysis results in Figure 3 was just opposite to the predictions of Hong (ibid.), but the same results were obtained.

7 Two sentences in (4) are the same sentences which were presented in Section 2.1.

However, more experimental studies are necessary to examine why these two subjecthood diagnostics behaved differently from the others.

6. Conclusion

In this paper, the grouping effects of subjecthood diagnostics were closely examined with a statistical analysis. Even though there have been a lot of theoretical investigations on the subject diagnostics in Korean, there have been only a few studies which examined the subjecthood tests with experimental designs. Moreover, there may be no or few studies which examined the grouping effects of subjecthood diagnostics based on the experimental data.

In this paper, 6 subject diagnostics (*Honorific Agreement*, *Plural Copying*, *Obligatory Control*, *Reflexive Binding*, *Coordinated Deletion*, and *Adjunct Control*) were classified into 3 groups (*Grammatical Subject*, *Consciousness and Predication Subject*, and *Discourse Topichood*). Then, a multi-level/hierarchical GLMM was applied in order to examine how the grouping factor of subjecthood diagnostics affected each test and the acceptability scores of native speakers. The analysis results demonstrated that the grouping factor of subjecthood diagnostics was statistically insignificant. It was also observed that *Honorific Agreement* and *Plural Copying* demonstrated different behaviors from the others.

Though the analysis results in this paper did not show all the aspects of subjecthood tests, they were enough to demonstrate how subjecthood diagnostics and their grouping factor affected the acceptability scores of native speakers. We hope that this study can contribute to the description and explanation of Korean.

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.

Cowart, Wayne. 1997. *Experimental syntax: Applying objective methods to sentence*

judgments. Thousands Oaks, CA: Sage Publications.
Gelman, Andrew and Jeniffer Hill. 2006. *Data analysis using regression and multi-level/hierarchical models*. Cambridge: Cambridge University Press.
Gries, Stefan St. 2013. *Statistics for linguistics with R: A practical introduction*. Second Edition. Berlin: Gruyter.
Hong, Ki-Ssun. 1991. *Argument selection and case-marking in Korean*. PhD Dissertation, Stanford University.
Hong, Ki-sun, 1994. Subjecthood tests in Korean. *Language Research* 30(1): 99-136.
Johnson, Keith. 2008. *Quantitative methods in linguistics*. Oxford: Blackwell.
Keller, Frank. 2000. *Gradient in grammar: Experimental and computational aspects of degrees of grammaticality*. PhD Dissertation, University of Edinburgh.
Kim, Eunah, Ji-Hye Kim, and Yong-hun Lee. 2015. Honorific agreement and plural copying as subject diagnostics in Korean: An experimental approach. Manuscript. Seoul National University.
Kim, Ji-Hye, Yong-hun Lee, and Eunah Kim. 2015. Obligatory control and coordinated subject deletion as Korean subject diagnostics: An experimental approach. *Language and Information* 19(1): 75-101.
Lee, Yong-hun, Eunah Kim, and Ji-Hye Kim. 2015. Reflexive binding and adjunct control as subject diagnostics in Korean: An experimental approach. *Studies in Language* 31(2): 427-449.
Lodge, Milton. 1981. *Magnitude scaling: Quantitative measurement of opinions*. Beverley Hills, CA: Sage Publications.
Pinheiro, José and Douglas Bates. 2000. *Mixed-effects models in S and S-Plus*. Berlin; New York: Springer.
Schütze, Carson. 1996. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago, IL: University of Chicago Press.
Sprouse, Jon and Norbert Hornstein. 2013. Experimental syntax and island effects: Toward a comprehensive theory of islands. In Jon Sprouse and Norbert Hornstein (eds.), *Experimental syntax and island effects*, 1-17. Cambridge: Cambridge University Press.
Stevenson, Stanley. 1975. *Psycholinguistics: Introduction to its perceptual, neural, and social prospects*. New York: John Wiley.
Yoon, James. 1986. Some queries concerning the syntax of multiple subject constructions in Korean. *Studies in the Linguistic Sciences* 16: 215-236.
Youn, Cheong. 1990. *A relational analysis of Korean multiple nominative constructions*. PhD Dissertation, State University of New York at Buffalo.
Zuur, Alain, Elena Ieno, Neil Walker, and Anatoly Saveliev, and Graham Smith. 2009. *Mixed effects models and extensions in Ecology with R*. New York: Springer.

Yong-hun Lee
Department of English Language and Literature
Chungnam National University
99 Daehak-ro, Yuseng-gu, Daejeon 34134, Korea
E-mail: yleeuiuc@hanmail.net

Yeonkyung Park
Department of English Language and Literature
Hannam University
70 Hannam-ro, Daedeok-gu, Daejeon 34430, Korea
E-mail: withbyk@hanmail.net

Eunah Kim
Faculty of Liberal Education
Seoul National University
1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea
E-mail: eakim2@gmail.com

Received: 2015. 10. 12.
Revised: 2015. 11. 05.
Accepted: 2015. 11. 05.