INSTRUMENTS FOR MEASURING GRAMMATICALITY AND PREFERENCE: THE STATUS OF MODIFIED NPs IN ENGLISH AND KOREAN

DEBORAH MASTERSON

University of Hawai‘i

Grammaticality judgments are a lot like 25-cent hot dogs — they contain meat, but they contain a lot of other ingredients, too.

—David Birdsong (1989)

We examine three data collection instruments in this study, the elicited imitation task (EI), the reaction time task (RT), and the written judgment, or “paper and pencil” task (P&P). We use these instruments to collect native language baseline data. Results of this experiment indicate that different tasks distinguish between judgments of what is “grammatical” and judgments of what is “preferred” by participants. Judgments of grammaticality appear to essentially binary, that is, a particular structure rates a very high score or a very low score on a scale which allows a range of values, whereas judgments of preference, a measure of ease of processing, take advantage of the full range of possible values, producing scores along a continuum of acceptability. Results also appear to be affected by other factors such as the type of participant and the experimental conditions.

IN ORDER TO UNDERTAKE MANY TYPES OF SLA RESEARCH, native language baseline data is needed. This information may be collected in a number of ways, using a number if different data collection instruments. Results of such data collection are even more valuable in making claims about native speakers’ performance if they can be cross-validated by data collected via more that one instrument or procedure. In this study, we gather NS baseline data for use in future SLA research. This study assumes a particular syntactic framework and compares data obtained from native speakers of English with data obtained from native speakers of Korean performing on the same structure types via three different tasks.
SYNTACTIC ANALYSIS

The structure under investigation is modified nominal phrases, that is, phrases in which nouns combine with adjectives, genitives, complements, and determiners. English and Korean differ with respect to the types of nominal phrases permitted and, of particular interest here, with respect to the permissible positions of specifiers and modifiers within the phrases. In this experiment we test participants on four specific nominal phrases types which reflect a number of contrasts between the two languages.

Contrasts between English and Korean noun phrases

Recent Government and Binding (GB) theory can account for these contrasts among nominal phrase types by making a distinction between languages such as English which have "functional" categories as well as lexical categories and languages such as Korean which have lexical categories only (see Fukui 1986, Fukui and Speas 1986). One of the characteristics of functional categories as opposed to lexical categories is that the former "close off" category projection; specifically, they prevent iteration of the specifier position. This is an important point for the discussion that follows.

Before we begin, we define the way in which we use the terms determiner, specifier, and modifier in this paper. Determiner and modifier are used in a general "pre-theoretical" sense, while specifier has meaning specifically within a particular theory of phrase structure.

Determiner is used to refer to the class of words such as a, the, and this in English and words such as ku, i, and ce in Korean. These words have the following characteristics in common: (1) they uniquely identify a noun, (2) they belong to a small closed set of lexical items, and (3) they are limited to one occurrence per nominal phrase. As discussed in a following section, the category determiner may also include genitive-marked nominals.

Specifier is defined relationally as the sister of X' and immediate daughter of a maximal projection. "Maximal projection" is used here to indicate the uppermost node in a constituent at which projection is closed off. Following Fukui, "functional categories have a unique specifier, but
lexical categories may iterate 'specifiers', as long as all 'specifiers' are fully licensed and can be interpreted at LF” (1986, p. 34). In English adjectives are specifiers of the noun; in Korean, all modifying elements are specifiers of the noun.

*Modifier* is used broadly as a cover term to refer to all elements in a nominal phrase other than the principal noun. For our purposes here these elements include determiners, genitive-marked nominals, adjectives, and complements of N. Within Fukui’s framework, determiners are of the functional category DET, which serves as the head of the English nominal phrase, thus the nominal phrase in English is actually a determiner phrase, or DP. As the head of DP, DET always occurs in the phrase-initial position in accordance with the more general English head-direction parameter. English determiners, as members of the functional category DET, prevent iteration of the specifier position. This has the effect of limiting the occurrence of determiners to at most one per DP. See Figure 1 for examples.

In contrast, Korean determiners do not function as heads. The nominal phrase in is an N’ with N as the phrase-final head, here in accordance with the Korean setting of the head-direction parameter. The lexical category N allows free iteration of the specifier position, i.e. it does not close off further category projection.

**Determiners and adjectives as modifiers:** ("MOD-NP")

Syntactically, Korean does not distinguish between specifiers with a determinative meaning such as *ku, i, and ce* and other modifiers of the head N. Because of this, the position of the determiner with respect to other modifiers in the nominal phrase is relatively free, accounting for the following contrasts between English and Korean:

(1)  

<table>
<thead>
<tr>
<th>English</th>
<th>Korean</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. the big blue book</td>
<td>a'. ku khun phwulun chayk the big blue book</td>
</tr>
<tr>
<td>b. * big the blue book</td>
<td>b'. khun ku phwulun chayk big the blue book</td>
</tr>
<tr>
<td>c. * big blue the book</td>
<td>c'. khun phwulun ku chayk big blue the book</td>
</tr>
</tbody>
</table>
Genitive-marked nouns as modifiers: possessives ("POS-NP")

In both languages, a genitive-marked possessive nominal occupies a specifier position, as defined above, in the nominal phrase. In English, only one genitive-marked nominal, the specifier of DP^2, is licensed per DP by the syntax; in Korean the syntax permits multiple specifier positions in N'—presumably it is semantics that limits the number of genitive-marked nominal with the meaning "possessor" as specifier of N'. These claims, together with previous claims of free ordering of modifiers (examples (2)b, b', c, c') and obligatory phrase-final N in Korean (example (2)d), account for the following contrasts among these nominal phrases containing genitive-marked nouns as modifiers:

(2)  
(a) David's house  
(b) * the David's house  
(c) * David's the house  
(d) the house of David's

Genitive-marked nouns as modifiers: Two subcategorized arguments ("2ARG-NP")

There are classes of nouns in both languages that subcategorize for arguments. In either language the external argument, that is the subject of the noun, appears as a genitive-marked nominal in a specifier position. In English, the external argument is a specifier under DP; in Korean it is a specifier under N'. In both languages the internal argument is a complement of N (see Figure 1). In English, the subcategorized internal argument often occurs as a PP; in Korean, it is optionally a genitive-marked nominal complement of the head N. The effect of this in Korean is the appearance of two genitive-marked nominals in the surface string. Note the following contrasts among these nominal phrases containing two arguments (external and internal):
Figure 1
Examples of Nominal Phrase Structures

a. noun modified by determiner + adjectives

```
        D'  N'
       /   |  /
      DET  the  A'
         big  blue
textbook
```

b. noun modified by (determiner+) genitive

```
           N'
          /  | 
         N'  'the'
        /   |
       N'  N
     'blue'
   'book'
```

```
           N'
          /  | 
         N'  'the's'
        /   |
       N'  N
     'friend's'
   'office'
```

```
           N'
          /  | 
         N'  'the'
        /   |
       N'  N
     'pohyu's'
   'criticism'
```

```
           N'
          /  | 
         N'  'the'
        /   |
       N'  N
     'changho's'
   'criticism'
```

```
           N'
          /  | 
         N'  'the'
        /   |
       N'  N
     'JJJJJ'
   'JJJ'
Genitive-marked nouns as modifiers: Three subcategorized arguments ("3ARG-NP")

There are cases of nouns that subcategorize for two internal arguments. Below are examples of nominal phrases containing three arguments (one external and two internal). Korean arguments all carry a final genitive case-markers; in addition, the recipient argument is dative-marked:

(4) a. * Steven’s to Jenny of flowers gift
   a’. Steven-uy Jenny-ey-key-uy kkoch-uy senmwul
   Steven-GEN Jenny-DAT-GEN flowers-GEN gift
b. * Steven’s to Jenny of flowers the gift
   b’. Steven-uy Jenny-ey-key-uy kkoch-uy ku senmwul
   Steven-GEN Jenny-DAT-GEN flowers-GEN the gift
c. Steven’s gift of flowers to Jenny
   c’. * Steven-uy senmwul kkoch-uy Jenny-ey-key-uy
   Steven-GEN gift flowers-GEN Jenny-DAT-GEN
d. Jenny’s gift of flowers from Steven
   d’. * Jenny-ey-key-uy senmwul kkoch-uy Steven-uy
       Jenny-DAT-GEN gift flowers-GEN Steven-GEN

e. the gift of flowers from Steven to Jenny
   e’. * ku senmwul kkoch-uy Steven-uy Jenny-ey-key-uy
       the gift flowers-GEN Steven-GEN Jenny-DAT-GEN

Phrase-final determiners: Ungrammatical in either language

Because the head in English is DET and that in Korean is N and because the English head, DET, is phrase-initial while the Korean head, N, is phrase-final, phrase-final determiners are ungrammatical in both languages:

(5) a. * big blue book the
    a’. * khun phwulun chayk ku
        big blue book the

b. * house David’s
   b’. * cip David-uy
        house David-GEN

c. * Japan’s Seoul’s destruction the
   c’. * ilpon-uy sewul-uy phakoy ku
        Japan-GEN Seoul-GEN destruction the

d. * destruction of Seoul by Japan the
   d’. * phakoy sewul-uy ilpon-uy ku
        destruction Seoul-GEN Japan-GEN the

e. * Steven’s to Jenny of flowers gift the
   e’. * Steven-uy Jenny-ey-key-uy kkoch-uy senmwul ku
        Steven-GEN Jenny-DAT-GEN flowers-GEN gift the

f. * gift of flowers from Steven to Jenny the
   f’. * senmwul kkoch-uy Steven-uy Jenny-ey-key-uy ku
        gift flowers-GEN Steven-GEN Jenny-GEN the
Summary: Contrasts between English and Korean noun phrases

Fukui's version of GB theory accounts for contrasts between English and Korean in four types of nominal phrases. These types, exemplified above, are:

- **MOD-NP**: a noun modified by a determiner and adjectives,
- **POS-NP**: a noun modified by a genitive-marked possessive nominal,
- **2ARG-NP**: a noun modified by a genitive-marked nominal as external argument plus one subcategorized internal argument,
- **3ARG-NP**: a noun modified by a genitive-marked nominal as external argument plus two subcategorized internal arguments.

This study examines how native-speaker data for both English and Korean speakers performing on three separate “grammaticality judgment” tasks either supports or challenges the predictions of Fukui's characterization of nominal phrase structure.

**THE THREE TASKS: OVERVIEW**

The decision to use three data collection instruments with native speakers was motivated, in part, in response to two issues in SLA research, namely the importance of native-speaker baseline data and the need for cross-validation.

Not only is native-speaker data necessary when making claims about how people use language (first or otherwise) it is important, in order to try to eliminate unnecessary bias, that the data be “naive” as far as possible. The use of native-speaker controls in SLA research has increased somewhat since Gass (1983) observed that in SLA research “by in large, grammaticality judgments are not asked of native speakers”, but by no means has the practice of supplying native-speaker baseline data become standard. In place of controls, a researcher often includes personal intuitions about the structure under investigation. Such intuitions are assumed to represent the native speaker’s underlying competence, but the impropriety of this biased practice should be obvious. On this subject, Labov (1973, p. 199) offers the following comment: “linguists cannot continue to produce theory and data at the same time”.
"Intuition", in this case, as in all cases of grammaticality judgments, should not be mistaken for observation (Itkonen 1981). A case in point, Spencer (1972) compares the grammaticality judgments collected from both linguists and non-linguists to find that while the non-linguists agreed among themselves 80% of the time, they agreed with the linguists’ judgments on only half the sentences.

What Shimura (1990) has to say about the nature of data collected and the need for cross-validation applies equally well to the collection of L1 and L2 data. As pointed out by Shimura, grammaticality judgment tasks test not linguistic competence but rather performance which can be seen as an interaction between competence, problem solving devices, and other cognitive modules (following a model by White (1989)); he states:

Drawing a conclusion about learners’ L2 competence from the performance data obtained through the use of GJ [grammaticality judgments] alone can be very dangerous, because...performance reflects not only linguistic knowledge, but other interacting factors. Different tasks tap different combinations of cognitive modules (including linguistic competence) which make up performance of L2 learners and, therefore, have different degrees of access to competence. Ideally, conclusions about linguistic competence...need to be made based on data from several different types of experimental tasks. (p. 68)

The three data collection instruments that we used with native English and Korean speakers, RT, EI, and P&P tasks, are described below.

Reaction time (RT)

Reaction time tasks have been used to determine grammaticality according to the participant’s internal representation of strings, both for native speakers and for second and foreign language learners. Freedman and Forester (1985) claim native speakers require more time to process ungrammatical strings than grammatical ones in a presentation/timed response task for the following reason: when strings are grammatical, higher level representations can be formed which facilitates processing. The assumption in analysing data obtained from such a task is that a shorter response time indicates that a high level representation has been formed and thus distinguishes the grammatical
from the ungrammatical without the participant’s conscious realization of having made a judgment.

Using a reaction time task, Lynn Eubank, as reported in Eubank 1991, has done perhaps the most exhaustive study of native speakers as baseline data for L2 studies to date. In five separate experiments of native speakers of German and English, he was able to obtain a significant effect for grammaticality only in the final experiment in which he attributes to lengthening the input sentences.

Elicited imitation (EI)

The Elicited Imitation task is similar to the RT task, but unlike a paper-and-pencil task, in that information about grammaticality can be collected without having to question the participant directly. EI tasks are used as tests of grammatical competence, originally with native speakers, but more and more in recent years with second and foreign language learners as well. Chaudron & Russell (1991) claim results from previous L1 studies employing elicited imitation have found general correlation with results obtained from other measures, but that such studies have been less successful in eliciting performance reliability on specific grammar points. The present study is designed to compare results from the three tasks on an extremely constrained data set. Previous studies (Montgomery 1978, Perkins 1986, Perfetti 1969) have found length to be more critical than either syntactic complexity or word familiarity. These are all factors we have tried to control for in our study. Additionally, we consistently “buried” the structure of interest in the same sentence internal position to thwart primacy/recency advantage, that is to avoid the sentence-initial and sentence-final positions which have been shown to be more easily retrieved (Lewandowsky & Murdock, 1989).

Bley-Vroman & Chaudron (1992) hypothesize that elicited imitation works this way: First, a participant hears the input and processes it, forming a representation on a number of levels, they suggest the phrasal level may be the “controlling” level. For this reason we strictly controlled our input sentences for number of phrases. Next, the representation(s) are stored in short term memory (STM), then the participant formulates an output sentence based on
what can be retrieved from STM. The smaller the number of chunks that must be stored in STM, the easier the sentence will be to repeat, so grammatical sentences composed of a few coherent chunks should be easier than ungrammatical sentences composed of many disjointed phrasal fragments. The number “seven” (Miller, 1956) has been hypothesized as the approximate upper limit for the number of chunks that may be successfully stored and retrieved. Figure 2 outlines the number of chunks in the output sentences depending upon what unit is considered to be a “chunk”:

**Figure 3**

<table>
<thead>
<tr>
<th>English/Korean chunks</th>
<th>syllables</th>
<th>words*</th>
<th>phrasal chunks**</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOD-NP</td>
<td>11–12/16–20</td>
<td>9/7–8</td>
<td>7/8</td>
</tr>
<tr>
<td>POS-NP</td>
<td>12–16/19–22</td>
<td>8–10/7–8</td>
<td>7/8</td>
</tr>
<tr>
<td>2ARG-NP</td>
<td>12–18/20–21</td>
<td>7–11/6–7</td>
<td>7/8</td>
</tr>
<tr>
<td>3ARG-NP</td>
<td>14–22/24–27</td>
<td>11–13/7–8</td>
<td>7/8</td>
</tr>
</tbody>
</table>

* includes bound morphemes
** based on a formula by Abney (1991)

Note that while the English input sentences contain fewer syllables than do the Korean sentences, they contain more words. The number of phrasal chunks is comparable across languages.

**Pencil and paper (P&P)**

P&P-type tasks in their various forms are the most common means of collecting data about grammaticality in first language and second language studies. As both the field and the instruments are refined, we have become more aware of the fallacy of regarding the data collected as reflecting a participant’s underlying grammar; these tasks types are not the “window on competence” that they were previously considered to be. Nevertheless, this type of instrument is widely used for its simplicity and ease of administration. The relatively homogeneous data collected via P&P tasks tends to lend itself well to complex statistical analysis and opens up the possibility for replication and comparison in subsequent studies. As with other types of instruments,
several researchers have recommended and utilized cross-validating tasks with varying amounts of convergence.

Our version of the P&P task differs from the other tasks in two important aspects. First, while participants may have suspected that the point of these tasks was to gather information about the grammaticality or acceptability of certain sentences in their native language (although this was not particularly apparent from unsolicited comments made by a few of the participants after taking one or both of the first two tasks and before taking the P&P task), the instructions for the P&P task make this goal clear. And, second, the P&P task, unlike the previous two, was not timed — participants were allowed to take as much time as necessary to complete the task. The type of processing involved is “off-” rather than “on-line”.

Predicted results of this study

As noted above, this study is being undertaken, for two main purposes: first, to collect native-speaker baseline data for future second language acquisition research, and second, as an experiment in cross-validation of complementary data collection instruments. With respect to the native-speaker data, what we expect to observe is that all three instruments will detect differences in grammaticality values within each language, that is, there will be a significant difference between grammatical and ungrammatical sentences on each task. “Grammaticality”, in this instance, is defined by the selected syntactic theory. As for the issue of cross-validation, we do not expect results to be identical across tasks. We expect to see differences in the data collected by the three methods, however, we do not expect these differences to falsely indicate fundamental differences in the participants’ internal grammars, but rather to reflect differences in what particular aspects of the participants’ performance is being measured. To be specific, while we expect all three tasks to detect differences in grammatical versus ungrammatical sentences, we further predict that the EI task and the RT task, but not the P&P task will be especially sensitive to differences in ease of processing, that is, tasks which involve on-line processing will show a significant difference among
modification types, which vary in length and complexity. These predictions are stated below:

**Prediction 1.** The data collected by each of the three instruments (RT, EI, and P&P) will show significant effects for grammatical versus ungrammatical sentences on each task.

**Prediction 2.** The data collected by the RT and EI tasks, in contrast to the data from the P&P task, will show significant effects for modification types.

### DATA COLLECTION

**Generation of input sentences**

Input sentences were generated according to a strict formula. Based on Eubanks' findings (above) and on pilots of both the RT and EI tasks, we decided to make our input sentences fairly long. Each sentence consisted of a frame of the following form: SUBJECT+ VERB + NP + PP. Into the NP slot of this frame was inserted an example type made up of a vocabulary set. Here is an example of an input sentence generated by the formula:

**Figure 3**

**Input Sentence Generation Formula**

<table>
<thead>
<tr>
<th>FRAME</th>
<th>Mary found _____ under the desk.</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ VOCABULARY SET</td>
<td>the, new, black, briefcase</td>
</tr>
<tr>
<td>+ EXAMPLE TYPE</td>
<td>ADJ ADJ NOUN DET</td>
</tr>
<tr>
<td>INPUT SENTENCE</td>
<td>Mary found new black briefcase the under the desk.</td>
</tr>
</tbody>
</table>

There were four nominal phrase types, MOD-NP, POS-NP, 2ARG-NP, and 3ARG-NP as detailed above. For each of these nominal phrase types there were three frames and three vocabulary sets. The number of example types varied according to nominal phrase type. All possible combinations were
generated producing a total of 180 input sentences for each language. See Appendices C and D for the entire sets of English and Korean input sentences.

Participants: English speakers

The 49 English-speaking participants were recruited from Psychology 100, 220, and 371 classes at the University of Hawai‘i. Each earned $10.00 and 2% credit towards his or her final grade in return for participating. They ranged in age from 18 to 42, with a mean age of 23. All were native English speakers. However, beyond this common trait there was a wide variety of second language experience. Two participants claimed no knowledge whatsoever of a second language. While many had had minimal experience with a second language (e.g. high school Spanish), others were exposed to other languages in the home from childhood — some professed good passive knowledge of these languages.

Participants: Korean speakers

The 44 Korean-speaking participants were all high school seniors from Haman Boys’ High School in Haman, Gyeongsang Namdo, South Korea. The students were not remunerated in any way for participating; in fact, they were required to do so. All 44 participants were in the same homeroom together and attended all the same classes together every day. This particular group was a “vocational track” group; that is, none of them had plans to apply to a university following graduation. Virtually all had grown up in Haman and had attended the same schools together. Their language learning history was equally similar: as is standard practice in Korea, all had begun instruction in English at the age of thirteen or fourteen and continued to the present time taking three to five hours a week. All had also studied German for two hours a week during the preceding year. They ranged in age from 17 to 19.

Procedure

91 participants performed the three tasks in the same sequence: RT, EI, and P&P; the remaining two, both Koreans, followed the sequence: EI, RT, P&P due to scheduling constraints.
The RT experiment was run on a Macintosh Powerbook 140 using a HyperCard program designed specifically for the task. A native-speaking researcher sat to one side, first to guide the participant through the instruction and practice session and answer questions, and then, to observe the process. Both during the 5-pair practice session and during the actual task the priming sentence was displayed for 4.16 seconds. The priming sentence then disappeared and a pointing finger appeared for .36 seconds to indicate in which of four random locations the pair sentence would appear. As soon as the pair sentence appeared, the participant was to indicate, as quickly as possible, if it matched exactly the priming sentence by pressing the red key ("f", covered with a red tag), or if the sentence differed in some way by pressing the blue key ("j", covered with a blue tag). The participant initiated the next set of sentences by pressing the space bar. The program presented 60 randomly ordered sentences and recorded time between the appearance of a possible matching sentence and the participant's decision (key press). The resolution of the system clock is 1/60 of a second (.016 seconds), or one Macintosh "tick". 45 sentences were equally distributed among grammaticality value, vocabulary sets, and three of the four modification types (POS-NP not used). The remaining 15 sentences were unrelated distractors.

During the EI task, the experimenter used two Sony TCM-5000EV tape recorders. One was used to play taped instructions for the participant which included four examples of elicited imitation. In the instructions it was stressed that the participant should do his or her best to recall as much of the sentence as possible, even if the imitation were incomplete. The experimenter controlled the presentation of each elicitation by pausing and resuming the elicitation tape. The second tape recorder was placed in front of the participant to record the elicitation and the following imitations. The experimenter did not speak except to ask if the participant were ready to begin, to thank the participant at the end and, during the experiment, to offer a little encouragement when it seemed necessary. Each of the three pre-recorded versions of the EI task contained 60 randomly ordered sentences equally distributed among grammaticality value, vocabulary sets, and modification type.

After completing the first two tasks, each participant was directed to read a full page of instructions describing the P&P task and was given the
opportunity to ask questions. Also, the participant filled out a questionnaire describing any exposure to other languages before starting (see Appendices A and B). Three versions of the task contained 60 randomly ordered sentences, each version equally distributed among grammaticality value, vocabulary sets, and modification type. The participants were asked to judge the sentences and rate them on a Likert scale containing the values 1 through 5, 5 being "good", 1 being "bad" (see Appendices A and B).

RESULTS

Tables 1 through 6 show the effect of modification type and grammaticality on the various task measurements. A two-way repeated measures ANOVA was done for each task and for each language group. The independent variables were grammaticality and modification type and the dependent variable depended upon the task, i.e. time used to determine match/no match for RT, correctness of imitation for EI, and rating on the Likert scale for P&P.

RT results (Tables 1 and 2)

Results from both languages show a significant effect for modification type as stated in Prediction 1. The English participants' data is consistent with the expectation of a significant effect for grammaticality, but the effect for grammaticality in Korean, though conforming to the pattern we predicted, did not reach statistical significance. Additionally, English shows an interaction between modification type and grammaticality. Not only did ungrammatical sentences take more time to process than their grammatical counterparts of identical length and vocabulary indicating the participants' judgments about grammaticality, but increasing complexity as measured both by greater number of syllables and words and by greater structural complexity required more processing time, indicating the participants' preferences for certain types of sentences.
EI results (Tables 3 and 4)

Both the English and Korean results show significant effects for both grammaticality (Prediction 1) and for modification type (Prediction 2) as well as an interaction between the two. These findings are consistent with the suggestion that EI measures both grammaticality and preference. However, the Korean results are less clear. In the case of the Korean participants, an extremely low rate of success in achieving complete repetitions of the elicitation sentences is evidenced by the “floor effect” (see Table 4) — virtually none of the sentences were repeated identically word-for-word. Therefore, in order to see any kind of pattern at all, we considered only the nominal phrase portion of the input sentence. For this first pass at the data, very strict scoring criteria were used. Only 100% correct repetitions of nominal phrases were counted. No allowance was made for lexical replacement, for example, “Richard” for “Robert”, “discussed” for “described”, “after the party” for “before class”. Such replacements did not alter the basic structure of the input and usually only involved vocabulary that appeared elsewhere in the input data. We expect that in future analyses of the EI data a scoring system which allows for lexical replacement will yield more meaningful results.

There are a number of possible explanation for the Korean EI results, for example, there is the difference in length in terms of chunks as discussed above. Another explanation may have to do with differences between the two groups of participants. We observed a striking difference in the general attitudes of the two groups, especially in terms of the EI task. While the English speaking participants seemed quite willing, and at times eager to take part in the experiment, the Korean participants often appeared nervous and confused. This is not surprising; the EI task was administered by an native English-speaking researcher with poor Korean language skills (in both the English and Korean versions all instructions were pre-recorded), therefore making it difficult for the participants to ask additional questions, although the researcher did make an effort to provide encouragement in Korean. Also, the Korean participants, unlike the English-speakers who were paid for their participation, were required by their teacher to participate. It was stressed to
Reaction Time Task Data: The effect of modification type and grammaticality on reaction time (1 unit = 16 milliseconds)

Table 1: English RT

Table 2: Korean RT

Elicited Imitation Task Data: The effect of modification type and grammaticality on the percentage correctly repeated

Table 3: English EI

Table 4: Korean EI

Pencil & Paper Task Data: The effect of modification type and grammaticality on sentence rating

Table 5: English P&P

Table 6: Korean P&P
these participants that individual performances on the tasks would not be revealed to the school administration. In addition, though it is somewhat difficult to compare the two groups, the general level of education of the Koreans was lower than that of the native English participants. The mean age of the Koreans was also lower — 18 as compared to 23. Gleitman & Gleitman (1970) have shown that selection of college students, as opposed to other types of participants, can strongly affect metalinguistic performance.

Additional factors complicating the tasks for the Koreans may have been the unfamiliarity of the task and anxiety about the ability of the instrument to fairly represent an individual’s abilities. Such effects have been documented in a study (Oh 1992) on Korean students being tested on their knowledge of English. Oh found that performance suffered significantly as the students’ unfamiliarity with a task and their anxiety about the results increased. The English-speaking participants in our experiment were relatively sophisticated with respect to familiarity with psychological testing in general — they had discussed it in class and, for many of the participants, this was the fourth experiment in which they had voluntarily participated during the current semester. It is also less likely that their anxiety level was as high as that of the Koreans, as they were being rewarded for their participation rather than being required to participate regardless of their wishes.

P&P results (Tables 5 and 6)

The results from the English participants’ data conform to our predictions. The Korean data shows a significant effect for grammaticality, however this data also shows an unexpected effect for the modification type, challenging our prediction. Note, for example, that a short ungrammatical sentence is considered more grammatical than a long grammatical one. This suggests that preference data is captured by the P&P task in the case of the Koreans, due in part perhaps to the fact that the Koreans made wider use of the intermediate ratings on the 1–5 Likert scale. There is no immediate explanation for this finding; one possible explanation that comes to mind is that, although both sets of instructions were based on instructions developed in Bley-Vroman, Felix, and Ioup (1988), the Korean varied from the English in some way that
encouraged the Korean participants to behave differently. While the Korean instructions were not a direct translation of the English ones, great care was taken in constructing the P&P instructions so that not only would the instructions be clear to each group of participants, but also that, as far as possible, instructions would be comparable in terms of content and emphasis (see Appendices A and B). However, the presentation of the instructions differed between the two language groups; each English-speaking participant completed the P&P task by him- or herself in a room attended by a research assistant. The participant was encouraged to ask questions as to how to make use of the scale, though of course, no specific responses were suggested. The Korean participants all completed the P&P task together in a classroom attended by the researcher, a native Korean-speaking assistant, and the participants’ homeroom teacher. Upon the researcher’s recommendation, the Korean assistant went through the instructions with the class, stressing that they should make wide use of all ratings if it seemed appropriate. It is possible that this point was overly stressed.

DISCUSSION AND CONCLUSION

The results of the experiment are summarized below:

**Figure 4**
Data Collection Results with respect to the Research Predictions:
Results which do not support the research predictions are shown in CAPS.

<table>
<thead>
<tr>
<th></th>
<th>P1: effect for grammaticality</th>
<th>P2: effect for modification type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>English</td>
<td>Korean</td>
</tr>
<tr>
<td>Reaction Time</td>
<td>yes</td>
<td>NO</td>
</tr>
<tr>
<td>Elicited Imitation</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Pencil &amp; Paper</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

The English data confirmed our expectations on all three tasks. However, as shown above, the Korean results failed to yield the expected result with
respect to the P&P task, showing an unexpected effect for modification type. Additionally, although the expected pattern was present, the Korean results on the RT task failed to attain significance. As mentioned above, performance by the Korean participants on the EI task was especially disappointing. We suspect that this is due in part to the inordinate load on short term memory imposed by the extremely long (in terms of syllables) input and that when this data is reanalyzed utilizing a more relaxed scoring system which ignores non-syntactic errors such as lexical replacement, the analysis will yield more meaningful results. However, length is not a sufficient explanation for the difficulties that the Korean participants evidenced on both the RT and EI tasks, and possibly even on the P&P task. Other explanations include those of age, education level, and attitude as discussed above. The difference between the Korean and English-speaking participants is not one of knowledge — for we assume that native-speaker knowledge is comparable cross-linguistically, but one of control. This distinction is discussed at length in Birdsong (1989, Chapter 2 in particular).

Though the results of the present experiment are mixed, they do tend to indicate that different tasks distinguish between judgments of what is “grammatical” and judgments of what is “preferred” by the participants. Specifically, at least for the native English speakers, the written judgment task elicited judgments of grammaticality that were essentially binary, that is, a particular structure rates a very high score or a very low score on a Likert scale, whereas the other two tasks, the reaction time task and the elicited imitation task generated judgments of preference, i.e., they are measures of ease of processing, taking advantage of the full range of possible ratings and producing scores along a continuum of acceptability. As discussed above, results are by no means “noiseless”; they appear to be affected by other factors such as the type of participant and the experimental conditions.

Thus, the news is both good and bad. Future researchers should regard the results of this experiment both as encouragement and as caveat when comparing and analysing data collected by the three tasks employed or any others. We must be not only sensitive to the population we are testing, especially if we plan to compare it with a distinct population, but we must also be aware of exactly what it is that we claim to measure with the instrument we use.
NOTES

1. I gratefully acknowledge the valuable assistance of Robert Bley-Vroman, who has contributed immeasurably to this project from its inception. Special thanks go to Choi Cheol-kyu for arranging for the Korean participants. Many thanks for help with data collection and analysis to Kate Wolfe-Quintero, Craig Chaudron, Soo ah Yuen, and Bill Johnston. I am particularly grateful to Jay Jeong-ryeol Kim for the many hours he has contributed, not only for discussions of theoretical issues and help in collecting data in Korea, but also for the design of the computer software used for the generation of the input sentences as well as the HyperCard adaptation of the reaction time instrument by Robert Bley-Vroman. Dr. Bley-Vroman’s program was based on a similar program by Derek Dunn-Rankin. This study was supported in part by a grant from the University of Hawai‘i College of Languages, Linguistics and Literature; Craig Chaudron and Robert Bley-Vroman, Principal Investigators. Any errors in this study are, of course, my own.

2. Fukui proposes two types of elements that may appear under DET. These are case-assigners such as genitive “’s” and non-case-assigners such as determiners a, the, and that.

The example shows the case-assigning element ‘s in DET which licenses projection to the higher level, DP and assigns genitive case to the specifier position. It is here that David appears. Fukui reserves the term "specifier" for this position and defines specifiers as elements that "close off" a category projection. In contrast, the, unlike ‘s, is a non-case-assigning element and licenses projection only to the D' level, thereby not allowing a specifier position and also closing off further projection.
REFERENCES


Deborah Masterson
c/o Linguistics – Moore 569
1890 East West Road
University of Hawai‘i
Honolulu, HI 98622

e-mail: mastersn@uhunix.uhcc.hawaii.edu
ENGLISH NATIVE SPEAKER INSTRUCTIONS

name __________________________________ __ age

Speakers of a language seem to develop a 'feel' for what is a possible sentence, even in many cases where they have never been taught a particular rule. For example, you might feel sentence number 1 below sounds like a good English sentence, while sentence number 2 is bad, even though you can't state a particular rule. Sentence number 3 probably sounds bad, but not as bad as number 2.

1. John is likely to win the race.
2. John the race likely is to win.
3. John is probable to win the race.

We are going to present you with list of sentences. We would like you to rate these sentences on a scale of bad to good, 1 being "bad" and 5 being "good". Use the numbers 2 through 4 for those sentences you are not sure about, like sentence number 3 above.

After you read the sentence, rate it by circling a number to the right of it, for example:

1. John is likely to win the race.  
   ![Circle 5]
2. John the race likely is to win.  
   ![Circle 1]
3. John is probable to win the race.
   ![Circle 1]

You would probably rate sentence number 3 as a "2", "3", or "4". There are no right or wrong answers, we just want to know how you, as a native speaker of English, feel about these sentences.
Before you begin, please tell us about any other language you speak or have studied. How did you learn the language? Did you study it formally or pick it up naturally? How old were you? Did you live in a place where the language was spoken? Are you studying a language currently? Please include all languages you have knowledge of.

Are you principally left-handed?  
☐ yes  ☐ no

Do you use your left hand for some things?  
☐ yes  ☐ no  
(such as using tools, utensils, writing, throwing, playing tennis)

Is anyone in your immediate family left-handed?  
☐ yes  ☐ no  
(mother, father, sister, brother)

Please take you time and read each sentence on the following pages carefully before you answer. Concentrate on the structure of the sentence. Ignore any problems with spelling, punctuation, etc. Mark only one answer for each sentence. Make sure you have marked all 57 sentences.

Thank you for your cooperation.
APPENDIX B: Korean P&P task

한국어 모국어 화자용 교본

이름 ___________________________ 나이 __________

사람들은 따로 문법 교육을 받지 않고도 모국어에서 가능한 문장과 그렇지 않은 것을 구분하는 능력을 갖고 있습니다. 예를 들어, 구체적인 구칙을 말할 수는 없어도, 아래 주어진 문장 1번은 바로 한국말 이지만 2번은 아니라고 느낄 것입니다. 그린가하면, 3번은 이상할 지 모르지만 2번만큼 나쁘지는 않은 것입니다.

1. 창수의 머리가 좋다.
2. 머리가 창수의 좋다.
3. 머리가 창수가 좋다.

여러분들께 주어진 문장 리스트를 보고 나쁜 문장에서 좋은 문장 스케일 1번에서 5번 중 여러분의 적극과 일치하는 번호를 하나 선택 하시오. 스케일의 1번은 나쁜 문장이고 5번은 좋은 문장이며 2번, 3번, 4번은 위의 3번째에와 같이 확실치 않은 경우입니다. 주어진 문장을 읽고, 오른쪽의 번호 중 여러분의 적극에 맞는 문법 스케일 번호를 하나 골라 동그라미를 치시오. 아래 예를 참고 하시오.

1. 창수의 머리가 좋다.  
2. 머리가 창수의 좋다.  
3. 머리가 창수가 좋다.

3번 문장은 아마 문법 스케일 2번, 3번 아니면 4번을 선택 했을 것입니다. 여기서 맞고 틀린 답은 없으며, 우리가 알고자 하는 것은 단지 여러분들이 한국어 모국어 사용자로서 주어진 문장들을 문법 스케일 상에서 어떻게 판단 하는지 하는 것 입니다.
시작하기 전에, 여러분들이 아는 외국어에 대해서 아래에 주어진 여백에 묻는 대로 대답하시오.

외국어를 배운 적이 있습니까? 아래에 해당하는 곳에 가위표를 하시오.

☐에 ☐아니요

외국어를 배운적이 있으면, 그 외국어에 관한 정보가 필요합니다. 그 언어를 어떻게 배우였으며 얼마나 배웠습니까? 그리고, 배우기 시작했을 때 몇살이었으며 몇 년 동안이나 배웠습니까? 지금도 외국어를 배우고 있습니까? 아는 외국어를 잘 부르시나요.

사람 집이 있습니까? ☐에 ☐아니요

부분적으로, 사람집이 있습니까? ☐에 ☐아니요
(예를 들어, 도구를 다루거나 공을 던질때)

가족중에 사람들과 있습니까? ☐에 ☐아니요

시간을 충분히 갖고, 여러분이 답하기 전에 다음페이지에 있는 문제들을 잘읽어보십시오. 문장의 맞춤법이나 문장구조가 틀린 것은 무시합니다. 대신에 여러분들이 이 문장을 들었을 때 얼마나 자연스러운 한국말로 느낄 것인지 지적해서 답합합니다. 각 문장에 하나의 문법 스케일 번호에 만 동그라미 하십시오. 주어진 57 개 문장 모두 빠짐없이 답하시오.

여러분의 협조에 감사합니다.
APPENDIX C: English input sentences

David lost the big blue textbook in the classroom
David lost the small red notebook in the classroom
David lost the new black briefcase in the classroom
David lost big blue the textbook in the classroom
David lost small red the notebook in the classroom
David lost new black the briefcase in the classroom
David lost big blue textbook the in the classroom
David lost small red notebook the in the classroom
David lost new black briefcase the in the classroom
Linda saw the big blue textbook on the chair
Linda saw the small red notebook on the chair
Linda saw the new black briefcase on the chair
Linda saw big blue the textbook on the chair
Linda saw small red the notebook on the chair
Linda saw new black the briefcase on the chair
Linda saw big blue textbook the on the chair
Linda saw small red notebook the on the chair
Linda saw new black briefcase the on the chair
Mary found the big blue textbook under the desk
Mary found the small red notebook under the desk
Mary found the new black briefcase under the desk
Mary found big blue the textbook under the desk
Mary found small red the notebook under the desk
Mary found new black the briefcase under the desk
Mary found big blue the textbook under the desk
Mary found small red the notebook under the desk
Mary found new black the briefcase under the desk
We went to the house of a teacher of David's after class
We went to the shop of a neighbor of Jenny's after class
We went to the office of a friend of Father's after class
We went by David's teacher's the house after class
We went by Jenny's neighbor's the shop after class
We went by Father's friend's the office after class
They arrived at David's teacher's house before noon
They arrived at Jenny's neighbor's shop before noon
They arrived at Father's friend's office before noon
They arrived at the house of David's teacher before noon
They arrived at the shop of Jenny's neighbor before noon
They arrived at the office of Father's friend before noon
They arrived at the house of a teacher of David's before noon
They arrived at the shop of a neighbor of Jenny's before noon
They arrived at the office of a friend of Father's before noon
They arrived at David's teacher's house before noon
They arrived at Jenny's neighbor's shop before noon
They arrived at Father's friend's office before noon
They arrived at David's teacher's house before noon
They arrived at Jenny's neighbor's shop before noon
They arrived at Father's friend's office before noon
They arrived at David's teacher's house before noon
They arrived at Jenny's neighbor's shop before noon
They arrived at Father's friend's office before noon
They arrived at David's teacher's house before noon
They arrived at Jenny's neighbor's shop before noon
They arrived at Father's friend's office before noon
They arrived at David's teacher's house before noon
They arrived at Jenny's neighbor's shop before noon
They arrived at Father's friend's office before noon
We discussed Japan's destruction of Seoul at home
We discussed Jenny's criticism of Richard at home
We discussed Steven's photo of Linda at home
We discussed Seoul's destruction by Japan at home
We discussed Richard's criticism by Jenny at home
We discussed Linda's photo by Steven at home
We discussed the destruction of Seoul by Japan at home
We discussed the criticism of Richard by Jenny at home
We discussed the photo of Linda by Steven at home
We discussed Japan's Seoul's destruction at home
We discussed Jenny's Richard's criticism at home
We discussed Steven's Linda's photo at home
We discussed Japan's Seoul's the destruction at home
We discussed Jenny's Richard's the criticism at home
We discussed Steven's Linda's the photo at home
We discussed Japan's Seoul's destruction the at home
We discussed Jenny's Richard's criticism the at home
We discussed Steven's Linda's photo the at home
I described Japan's destruction of Seoul to the group
I described Jenny's Richard's criticism to the group
I described Steven's photo of Linda to the group
I described Seoul's destruction by Japan to the group
I described Richard's criticism by Jenny to the group
I described Linda's photo by Steven to the group
I described the destruction of Seoul by Japan to the group
I described the criticism of Richard by Jenny to the group
I described the photo of Linda by Steven to the group
I described Japan's Seoul's destruction to the group
I described Jenny's Richard's criticism to the group
I described Steven's Linda's photo to the group
I described Japan's Seoul's the destruction to the group
I described Jenny's Richard's the criticism to the group
I described Steven's Linda's the photo to the group
I described Japan's Seoul's destruction the to the group
I described Jenny's Richard's criticism the to the group
I described Steven's Linda's photo the to the group
They discussed Mary's donation of money to Robert at home
They discussed Linda's offer of a job to Richard at home
They discussed the donation of money from Mary to Robert in the letter
They discussed the offer of a job by Linda to Richard in the letter
They discussed Robert's donation of money from Mary to Richard in the letter
They discussed the gift of flowers from Steven to Jenny in the letter
They discussed Steven's gift of flowers to Jenny at home
They discussed Linda's offer of a job to Richard at home
They discussed the offer of a job by Linda to Richard at home
They mentioned Steven's gift of flowers to Jenny in the letter
They mentioned Linda's offer of a job to Richard in the letter
They mentioned Mary's donation of money to Robert in the letter
They mentioned Robert's donation of money from Mary to Richard in the letter
They mentioned the gift of flowers from Steven to Jenny in the letter
They mentioned Steven's gift of flowers to Jenny at home
They mentioned Mary's donation of money to Robert in the letter
They mentioned Linda's offer of a job to Richard in the letter
They mentioned Japan's destruction of Seoul in the letter
They mentioned Jenny's criticism of Richard in the letter
They mentioned Steven's photo of Linda in the letter
They mentioned the destruction of Seoul by Japan in the letter
They mentioned the criticism of Richard by Jenny in the letter
They mentioned Steven's photo of Linda in the letter
They mentioned Japan's Seoul's destruction in the letter
They mentioned Jenny's Richard's criticism in the letter
They mentioned Steven's Linda's photo in the letter
They mentioned Japan's Seoul's the destruction in the letter
They mentioned Jenny's Richard's the criticism in the letter
They mentioned Steven's Linda's the photo in the letter
They mentioned Japan's Seoul's destruction the in the letter
They mentioned Jenny's Richard's criticism the in the letter
They mentioned Steven's Linda's photo the in the letter
They discussed Steven's gift of flowers to Jenny at home
They discussed Linda's offer of a job to Richard at home
They discussed Mary's donation of money to Robert at home
They discussed Jenny's gift of flowers from Steven at home
They discussed Richard's offer of a job from Linda at home
They discussed Robert's donation of money from Mary at home
They discussed the gift of flowers from Steven to Jenny at home
They discussed the offer of a job by Linda to Richard at home
They discussed the donation of money from Mary to Robert at home
They discussed Steven's to Jenny of flowers gift at home
They discussed Linda's to Richard of a job offer at home
They discussed Mary's to Robert of money donation at home
They discussed Steven's to Jenny of flowers the gift at home
They discussed Linda's to Richard of a job the offer at home
They discussed Mary's to Robert of money the donation at home
They discussed Steven's to Jenny of flowers gift the at home
They discussed Linda's to Richard of a job offer the at home
They discussed Mary's to Robert of money donation the at home
We described Steven's gift of flowers to Jenny to the group
We described Linda's offer of a job to Richard to the group
We described Mary's donation of money to Robert to the group
We described Jenny's gift of flowers from Steven to the group
We described Richard's offer of a job from Linda to the group
We described Robert's donation of money from Mary to the group
We described the gift of flowers from Steven to Jenny to the group
We described the offer of a job by Linda to Richard to the group
We described the donation of money from Mary to Robert to the group
We described Steven's to Jenny of flowers gift to the group
We described Linda's to Richard of a job offer to the group
We described Mary's to Robert of money donation to the group
We described Steven's to Jenny of flowers the gift to the group
We described Linda's to Richard of a job the offer to the group
We described Mary's to Robert of money the donation to the group
We described Steven's to Jenny of flowers gift the to the group
We described Linda's to Richard of a job offer the to the group
APPENDIX D: Korean input sentences

우리가 파티 테이블에 아바라필의 카드의 사무실에 갔다.
우리가 파티 테이블에 엄수의 선생님의 집에 갔다.
우리가 파티 테이블에 보의의 이웃의 가게에 갔다.
우리가 파티 테이블에 아바라필의 카드의 사무실에 갔다.
내가 수업 테이블에 엄수의 선생님의 집에 갔다.
내가 수업 테이블에 보의의 이웃의 가게에 갔다.
내가 수업 테이블에 아바라필의 카드의 사무실에 갔다.
내가 수업 테이블에 엄수의 선생님의 집에 갔다.
내가 수업 테이블에 보의의 이웃의 가게에 갔다.
내가 수업 테이블에 아바라필의 카드의 사무실에 갔다.
내가 수업 테이블에 엄수의 선생님의 집에 갔다.
내가 수업 테이블에 보의의 이웃의 가게에 갔다.
내가 수업 테이블에 아바라필의 카드의 사무실에 갔다.
내가 수업 테이블에 엄수의 선생님의 집에 갔다.
내가 수업 테이블에 보의의 이웃의 가게에 갔다.
내가 수업 테이블에 아바라필의 카드의 사무실에 갔다.
내가 수업 테이블에 엄수의 선생님의 집에 갔다.
내가 수업 테이블에 보의의 이웃의 가게에 갔다.
내가 수업 테이블에 아바라필의 카드의 사무실에 갔다.
내가 수업 테이블에 엄수의 선생님의 집에 갔다.
내가 수업 테이블에 보의의 이웃의 가게에 갔다.
내가 수업 테이블에 아바라필의 카드의 사무실에 갔다.
내가 수업 테이블에 엄수의 선생님의 집에 갔다.
내가 수업 테이블에 보의의 이웃의 가게에 갔다.