Grammaticality Judgments as Measurements of Linguistic Competence: A Case Study of Japanese EFL Learners

Seishin MIURA
Akita University

Abstract

The present study investigates whether or not grammaticality judgment tests reflect a learner’s linguistic competence.

The results of this experiment on grammaticality judgment showed the following:

1) On our grammaticality test, learners with high scores on a TOEFL-type pre-test did better than those with low scores on the TOEFL-type pre-test, notably showing that the former are stronger than the latter with Universal Grammar (UG) items.

2) Those with high scores on the TOEFL-type pre-test apparently make grammatical judgments more frequently based on their theoretical knowledge of grammar than those with low scores.

3) In making judgments, those with low scores more frequently utilized the strategy of guessing, as opposed to those with high scores.

Finally, several other issues related to this study are mentioned for continued research which can help prove grammatical judgment tests to be a valid method of measuring a learner’s competence and creativity of the learner language, i.e., Interlanguage.

1. Background

Grammaticality judgment tests have been considered as one of the most useful methods to measure a learner’s competence in the target language. (See Schachter, Jacquelyn et al.(1976), Chaudron (1983).) However, recently, some scholars are cautious to use these tests for eliciting data, claiming that these tests are not necessarily valid unless one pays much attention to factors such as guessing and time pressure so that the results will not be affected by them. (See Davies et al.(1998), Ellis (1991), Mazurkewich, I. & White, L.(1984).)

The debate on the validity of grammaticality judgment tests concerns three major
issues: (1) The distinction between competence and performance, (2) Universal Grammar vs. Variationist positions, and (3) Explicit vs. Implicit knowledge. Behind these issues are implied (a) the different degrees of grammaticality, i.e., grammaticalness, acceptability, and appropriateness, (b) the different nature of Interlanguage itself, i.e., gradability, uniqueness, and creativity, and finally, (c) learner factors. These ideas may be paraphrased as: Can these tests be used to measure competence or performance of a learner? What is competence? Is it an ability to be able to achieve even after a certain age; i.e., a critical period? If a person knows a language well, does it mean that a person is able to explain whether or not that sentence is well-formed in an outright manner? Bialystok’s (1979) distinction between explicit and implicit knowledge is suggestive of the changes of the learners’ language. If a learner uses explicit knowledge of the target language, he or she may be able to improve his or her language competence. For a learner to be fully competent, he must be able to judge both grammatical and ungrammatical sentences accurately. (See Arthur (1980: 178-183).) Therefore, it is important to consider both accurate and inaccurate responses.

The patterns of grammaticality judgments in a test to elicit data for these questions may be classified into four cases below:

(a) A case where a grammatical sentence is judged as grammatical.
(b) A case where a grammatical sentence is judged as ungrammatical.
(c) A case where an ungrammatical sentence is judged as grammatical.
(d) A case where an ungrammatical sentence is judged as ungrammatical.

Nobody will surely doubt that those who make their judgments as in cases (b) and (c) are grammatically incompetent because these people have produced an inaccurate response. However, it is not as easy a task as it may seem to determine whether a learner is grammatically competent when responding as in cases (a) and (d). Although most of the respondents in patterns (a) and (d) can be said to be competent because they made an accurate judgment, there is no denying a slight possibility that they responded accurately by chance. In recent studies of grammaticality judgments, this methodological difficulty, i.e., exclusion of chance from judgments of grammaticality, has been a big issue to make the use of grammaticality judgments valid as a means of measuring a linguistic competence. To solve this problem, in our study two more exploratory items were added in an attempt to make grammaticality judgment tests more reliable. (See Appendix I and Appendix II.)

Research questions for this study are as follows:

1. Do learners with high scores on our TOEFL-type pre-test score better than those with low scores?
2. How do those with high scores make grammatical judgments?
3. How do those with low scores make their judgments?
2. Method

2.1. Subjects

Ninety-four university students participated in this experiment. Some are studying English as their major. Most students are studying general English to develop their reading comprehension.

2.2. Procedures

The subjects were divided into three groups according to the results of a test which included several grammatical categories that can be found in a TOEFL test. Our test was composed of Universal Grammar items and basic grammar items as can be seen in Appendix I and II. In order to find which participants scored better, and how they reached the correct answer with confidence, accurate and erroneous responses were counted for each sentence in these three groups to compare the ratio of accuracy vs. error data. To account for learners’ competence, the responses were calculated as correct or incorrect.

2.3. Categories of Test Items

Our test consisted of two major parts: Universal Grammar items such as Complex Noun Phrases, Relative Clauses, Extraction of Subjects or Objects in Finite and Infinite Clauses, Adjuncts, Subjects, and basic grammar items such as Preposition, Number, Pied Piping, Tense, and Preposition Stranding. Each sentence is categorized as follows. (The numbers stand for each sentence in Appendix I and Appendix II.)

Note 1)  
- RC: 5, 25, 27  
- CN: 1, 3, 11  
- S: 4, 14, 24  
  [RC= Relative Clauses  
  CN= Complex Noun Phrases  
  ESF= Extraction of Subjects from Finite or Infinite Clauses  
  EOF= Extraction of Objects from Finite or Infinite Clauses  
  A= Adjuncts  
  S= Subjects  

Note 2)  
- P: 3, 7, 11, 14, 16, 19, 21, 24, 32  
- PP: 1, 6, 12, 18, 23, 29  
- PS: 2, 8, 13, 22, 27, 30  
  [P= Preposition  
  N= Number  
  PP= Pied Piping  
  T= Tense  
  PS= Preposition Stranding]
3. Results

To begin with, let us note that the subjects in this study are divided unequally into three groups as shown in Figure 1 (High achiever: 21 people, middle achiever: 58 people, and low achiever: 15 people).

![Figure 1: Number of Subjects in Each Group](image)

In Table 1, the overall data are provided for accurate responses, confident responses with their percentage against the total of the subjects, and the average of judgment grounds within each category of both tests.

<table>
<thead>
<tr>
<th>Test 1 Category</th>
<th>Accurate Responses</th>
<th>%</th>
<th>Confident Responses</th>
<th>%</th>
<th>Theory</th>
<th>Analogy</th>
<th>Guessing</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN</td>
<td>63.0</td>
<td>67.0</td>
<td>9</td>
<td>9.9</td>
<td>6.7</td>
<td>1.7</td>
<td>1.0</td>
</tr>
<tr>
<td>A</td>
<td>66.0</td>
<td>70.2</td>
<td>11</td>
<td>12.1</td>
<td>7.0</td>
<td>3.3</td>
<td>1.0</td>
</tr>
<tr>
<td>S</td>
<td>83.3</td>
<td>88.7</td>
<td>17</td>
<td>17.7</td>
<td>9.3</td>
<td>5.0</td>
<td>2.3</td>
</tr>
<tr>
<td>RC</td>
<td>51.7</td>
<td>55.0</td>
<td>5</td>
<td>5.3</td>
<td>4.3</td>
<td>0.7</td>
<td>0.0</td>
</tr>
<tr>
<td>ESF(a)</td>
<td>32.3</td>
<td>34.4</td>
<td>7</td>
<td>7.8</td>
<td>3.0</td>
<td>3.0</td>
<td>1.3</td>
</tr>
<tr>
<td>ESF(b)</td>
<td>37.3</td>
<td>39.7</td>
<td>6</td>
<td>6.4</td>
<td>2.7</td>
<td>2.3</td>
<td>1.0</td>
</tr>
<tr>
<td>EOF(a)</td>
<td>48.2</td>
<td>51.2</td>
<td>10</td>
<td>10.1</td>
<td>4.8</td>
<td>3.2</td>
<td>1.5</td>
</tr>
<tr>
<td>ESF(b)</td>
<td>37.3</td>
<td>39.7</td>
<td>6</td>
<td>6.4</td>
<td>2.7</td>
<td>2.3</td>
<td>1.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test 2 Category</th>
<th>Accurate Responses</th>
<th>%</th>
<th>Confident Responses</th>
<th>%</th>
<th>Theory</th>
<th>Analogy</th>
<th>Guessing</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>44.7</td>
<td>47.5</td>
<td>7.6</td>
<td>8.0</td>
<td>3.4</td>
<td>3.0</td>
<td>1.1</td>
</tr>
<tr>
<td>PP</td>
<td>39.0</td>
<td>41.5</td>
<td>10.2</td>
<td>10.8</td>
<td>5.5</td>
<td>3.7</td>
<td>1.0</td>
</tr>
<tr>
<td>PS</td>
<td>59.3</td>
<td>63.1</td>
<td>17.5</td>
<td>18.6</td>
<td>10.3</td>
<td>5.3</td>
<td>1.8</td>
</tr>
<tr>
<td>T</td>
<td>62.0</td>
<td>66.0</td>
<td>29.3</td>
<td>31.2</td>
<td>14.3</td>
<td>11.0</td>
<td>4.0</td>
</tr>
<tr>
<td>N</td>
<td>42.0</td>
<td>44.7</td>
<td>17.5</td>
<td>18.6</td>
<td>8.0</td>
<td>5.5</td>
<td>4.0</td>
</tr>
</tbody>
</table>

3.1. Data Analysis

In what follows, the ratio of accurate responses, the degree of confidence in making judgments, and the ground on which they made those judgments, are compared...
and analyzed among these groups in terms of the research questions given at the end of Section 1.

3.1.1. Ratio of Accurate Responses

As a whole, the ratios of accurate responses in Test 1 and Test 2 are almost identical. (See Figure 2.)

![Figure 2](image1)

However, if we take a look at each item, it is evident that a few students attached more accurate responses than other students in both tests. This simply means that although the two tests were of the same difficulty, both tests as a whole included both easy and difficult sentences for them to judge. (See Figure 3 and Figure 4.)

![Figure 3](image2)

![Figure 4](image3)

By analyzing each group separately, it seems that middle achievers give more
accurate answers. (See Figure 5 and Figure 6.)

![Figure 5](image1)

However, if analyzing the group as a whole, high achievers produce a higher percentage of accuracy for almost all items in both tests as indicated in Figure 7 and Figure 8. Only a few exceptions are found in ESF(a) and EOF(a) in Test 1 and P and PP in Test 2. At the present the reason for this is unclear.

![Figure 6](image2)

3.1.2. Degree of Confidence

For the degree of confidence in grammatical judgments, Test 2 shows a slightly higher percentage of respondents than in Test 1. (See Figure 2.) It can be speculated that this is due to the fact that Test 2 includes a lot of basic grammatical items. Although the degree of confidence also varies from item to item in both tests in general, high achievers demonstrate a higher percentage for all items in both tests. (See Figure 9 and Figure 10.) For some items the percentage of middle and low achievers is reversed. (See the categories A, ESF(b), EOF(a), and EOF(b) in Figure 9, and the category PP in Figure 10.)
3.1.3. Ground for Making Judgments

For the ground on which they made judgments, only those who were confident in their judgments were targeted for this study. More than 40% of the subjects utilize some kind of theoretical basis in grammar. (See Figure 11 and Figure 12).

This allows us to conclude that from what we have observed, greater knowledge of grammar gives confidence when making grammatical judgments and helps to produce accurate and correct responses.

3.2. Implications

The present study has been concerned with analysis of the accurate data obtained from our test. Its results are limited in that it attempted to describe exclusively how the participants responded correctly. Analysis of error data is also a contributing factor when examining a learner’s competence, because they complement the complete and truly accurate knowledge by a learner of the target language.
The present study could not provide a deeper analysis of the learner’s thought process, either. It could have covered different tasks focusing equally on production as well as comprehension. Other issues may be considered for future studies through the same set of data, such as consistency in the participant’s judgment, instruction effects, and correction of erroneous sentences.

4. Conclusion

Findings of this study confirmed that unlike Ellis’s (1991) claim, if the notion of competence can be interpreted to imply that it may develop given an appropriate learning context, grammaticality judgment tests can serve as a measurement of linguistic competence. They also suggest that, although second language learners may not achieve the same level of competence as that of native speakers, they should be able to develop a higher level of ultimate attainment and language creativity through learning grammar. This is, however, a question open for future research.

References

273-291.

Appendix I (Sentences for Test 1)
（指示文）以下の各文について、次の三つの点に答えてください。
① その文が文法的に正しいと思ったらT,文法的に正しくないと思ったらFを例のように問題文の番号の前の（）に書いてください。
② その判断について、(a)「確信がある」ならば数字の1, (b)「まあまあ確信がある」ならば数字の2,(c)「まったくない」ならば数字の3を例のように書いてください。
③ その判断をしたとき、その判断の根拠が、(a)「全く当てずっぽう」ならば数字の1, (b)「類似の構文からの推測」ならば数字の2, (c)「文法理論等からの知識」ならば数字の3を例のように書いてください。

（例） ① ② ③
<table>
<thead>
<tr>
<th>の</th>
<th>の</th>
<th>の</th>
</tr>
</thead>
</table>

（T）3（1）1. Who does Sam deny the story that he kissed?
( ) ( ) ( ) 1. Who does Sam deny the story that he kissed?
( ) ( ) ( ) 2. Who did Anne say likes her friend?
( ) ( ) ( ) 3. What do you believe the claim that Anne stole.
( ) ( ) ( ) 4. Who did a story by please the children?
( ) ( ) ( ) 5. What did Sam see the man who stole?
( ) ( ) ( ) 6. What does the man think crashed into the car?
( ) ( ) ( ) 7. Who did Jane say her friend likes?
( ) ( ) ( ) 8. Who does Anne want to win the election?
( ) ( ) ( ) 9. Who do you suppose Bill wants to marry?
( ) ( ) ( ) 10. Who did you meet Tom after you saw?
Appendix II (Sentences for Test 2)

1. The girls are sitting on the park bench this afternoon.
2. The man is knocking on the kitchen door right now.
3. The young girl waited the school bus yesterday morning.
4. The new teacher arrived the next day.
5. His parents is buying a new car today.
6. On which bench are the girls sitting this afternoon?
7. The little boy played the teddy bear yesterday.
8. Which door is the man knocking on right now?
9. Which day did the new teacher arrive?
10. Did his parents bought a new car last year?
11. Which bus did the young girl wait yesterday morning?
12. This is the bench on which the girls are sitting this afternoon.
13. This is the door that the man is knocking on right now.
14. Which teddy bear did the little boy play yesterday?
15. That is the day the new teacher arrived.
16. This is the bus that the young girl waited yesterday morning.
17. Which car my parents buy last year?
18. The children are laughing at the funny pictures this morning.
19. This is the teddy bear that the young boy played yesterday,
20. Her best friend graduated last year.
21. The tall nurse worked the doctor last year.
22. The children are looking at the tall trees right now.
23. At which pictures are the children laughing this morning?
24. Which doctor did the tall nurse work last year?
25. Your brother are eating a sandwich in the lunchroom.
26. What year did her best friend graduate?
27. Which trees are the children looking at right now?
28. Did your brother ate a sandwich in the lunchroom.
29. These are the pictures at which the children are laughing this morning.
30. These are the trees that the children are looking at right now.
31. What is your brother eating it in the lunchroom.
32. This is the doctor that the tall nurse worked last year.
33. That is the year her best friend graduated