The Difference between Japanese ESL Learners and L1 Speakers in Production and Perception:
A Study of Information Order in the English Dative Alternation

Kazuko KATSUFUJI
*The University of Tokushima*

This paper presents empirical data that demonstrate differences and relationships between two commonly used linguistic tasks performed by the L1 speaker (NS) and the Japanese ESL learner (NNS). Thirty-five NSs and thirty-five NNSs were tested in both their elicited production (production) and acceptability judgment (perception) of the English dative structure of which information order of two noun phrases (NPs) following the verb was controlled by preceding prompts. Results from both tests show that significant differences of scores between NSs and NNSs, i.e., higher sensitivity for the given-new information order for NSs. They indicate that production tasks principally evaluate syntactic-discourse competence for both NSs and NNSs. While perception tasks also measure the syntactic-discourse competence for NNSs as the production tasks do, they provide a less direct measure of such competence and seem to be influenced by extra-linguistic factors for the NSs. Furthermore, correlation between production and perception scores demonstrates an interesting pattern. While these two scores by NNSs are highly correlated, these scores by NSs are not as much correlated as by the NNSs'. Such findings are comparable to those found in early first language acquisition (Greenbaum and Quirk, 1970), and offers an evidence for Sorace's (1996) argument in L2 learners' internalized grammars.

1. Introduction

It is a broadly accepted practice of both linguists and second language (L2) acquisition researchers to depend on grammaticality judgments (perception) and orally produced linguistic utterances (production) to support their theoretical claims. The specific objective of these tasks are to measure linguistic competence: the stable state of knowledge of the NS, on the one hand, and the developing language competence of the NNS, on the other. Researchers have accumulated a large body of such data in the field of adult L2 acquisition (Rutherford, 1984). Generally, as
indicated in a study by Greenbaum and Quirk (1970), it is widely accepted that grammatically
judgments (perception) and linguistically produced utterances (production) are highly correlated
and it has been assumed that NSs have a tendency to use the same grammar for both the sentences
they accept and those they can produce. Subsequently, experimental design and approach to
measure L2 learners' internalized grammars have been borrowed from L1 theory without
questioning their validity and their applicability to the study of L2. Furthermore, in reading L2
acquisition literature, data obtained by these two different tests have been handled as if they are
compatible. Researchers such as Sorace (1996) question whether it is legitimate to assume that
L2 learners' linguistic intuitions in elicited production tasks and grammatical judgments are
indicators of L2 learners' internalized grammars. In fact, research in the past have not clarify
whether it is appropriate to treat native and nonnative competence as identical objects of
analysis.

Research in production and perception in L2 acquisition have been rigorously studied in the
area of phonology (Leather and James, 1996). Generally, previous investigations in the acquisi-
tion of new phonological contrasts by adult learners agree in that, unlike L1 acquisition, accurate
perception does not necessarily precede accurate production in L2. In the other language levels,
such as in syntactic and discourse levels, however, very little is known of the relationships between
production and perception. A systematic comparison between elicited imitation (production) and
act-out (comprehension) in a syntactic level is provided in a study by Flynn (1986). She shows
that production tests primarily evaluate a learner's knowledge of the second language while
comprehension tests principally evaluate a learner's sensitivity to pragmatic context. Her study
is significant in demonstrating the two different experimental tasks which access linguistic
knowledge differently, however, the comparison has been made only within L2 learners, but not
within L1 speakers.

The present study examines relationships between the production and perception of a
discourse factor by Japanese ESL learners in comparison with English NSs. The aim of the study
is not only to show how differently these two linguistic tasks measure linguistic knowledge, but
also to demonstrate how differently these tasks are performed by L2 learners and L1 speakers.

The study also attempts to utilize its findings to examine Sorace's argument. The findings
also offer a test for Flynn's concept in which production tasks primarily evaluate a learner's
knowledge of the language while comprehension tasks principally measure a learner's sensitivity
to pragmatic context. Before turning into the study, the structure to be tested in perception and
production is explained, and a brief summary of L1 crosslinguistic research with English and
Japanese on the relationship between discourse factors and the dative structure is provided. The
subsequent sections outline the research hypotheses, describe an experiment, and interpret the
results based on data collected from NSs and NNSs to examine the comparison of the two
linguistic tasks as reflections of L1 and L2 competence on the relationship between discourse
factors and the dative alternation. Finally, suggestions are made for further additional research.
2. Method

2.1 The Target Structure

In the present study, the English dative alternation was chosen as the area of investigation. The English dative alternation is well-known for its relationships with discourse, in addition to its sentence-level semantic and morphological constraints as pointed in Inagaki (1993). For example, even alternating verbs such as *give* is highly constrained in certain discourse contexts (Smyth, Prideaux and Hogan, 1979) as shown in the following sentences.

(1) a. Who did Paul give the book to?
   b. Paul gave the book to Jane.
   c. ?Paul gave Jane the book.
(2) a. What did Paul give Jane/to Jane?
   b. Paul gave Jane the book.
   c. ?Paul gave the book to Jane.

If sentences (1b), (1c), (2b), and (2c) are read with normal intonation without placing contrastive stress on any particular words, (1c) and (2c) sound less natural than (1b) and (2b) as answers to (1a) and (2a) respectively. The naturalness of the sentences is due to the distinction between the first noun phrase (NP) which carries given information and the second NP which bears new information. Thus, in English, there is a discourse constraint operating to discourage an NP carrying new information from filling in the NP position immediately after the verb (Note also that sentences (1c) and (2c) would sound natural if a stress were placed on a word which carries new information.).

Results from L1 crosslinguistic research with English (Smyth et al., 1978) and Japanese (Katsufuji, 1998) have shown that the dative alternation of the two languages exhibits quite different relationships to its discourse context. There is a significant effect for information ordering on the English dative alternation: English L1 speakers prefer dative structure with two object NPs in given-new information order as shown in (1b) and (2b). On the other hand, Japanese L1 speakers prefer the new-given information order to the given-new information order on the whole. For a detailed overview, see Katsufuji (1998).

2.2 Research Questions

The focus of the study is to compare production with perception within and between NSs and NNSs. The present study addresses the following research questions (RQ):

• RQ1. What are the differences between NSs' and NNSs' performance on production and perception? Will NSs' production and perception scores be higher than those of NNSs?
• RQ2. Are the NS's production and perception of the dative alternation in correct information order highly correlated as was in studies by Greenbaum and Quirk (1970)?
• RQ3. Are the NNS's production and perception of the dative alternation in correct information order highly correlated as much as those of the NS's?
2.3 Design

Two tasks were designed using the English dative alternation in discourse to examine the above-mentioned RQs. In Task 1, the participants are asked to judge dative constructions that alternate in relation to their preceding wh-questions. For Task 2, contextualized production tasks used in a study by Wolfe-Quintero (1992) were adopted and modified for the present study.

2.4 Participants and procedure

Both tasks contrasted groups of native English-speaking participants (mean age twenty-eight years) with Japanese learners of English (mean age twenty-seven years). The same participants volunteered for both tasks. There were 35 native speakers (NSs) and 35 Japanese learners of English (NNSs) for a total of 70 participants in both tasks. The participants, the NSs and NNSs, were either undergraduates, graduates, or faculty at the University of Hawai‘i at Manoa.

For both Tasks 1 and 2, the pre-recorded materials were listened to through a headset. In Task 1, the participants were asked to listen to 36 sets of questions and answers and to indicate their judgment for each answer by circling a number on a sheet. In Task 2, which was an elicited production task, they were asked to utter the answers according to a cue in a set of materials.

Each participant’s performance was recorded individually through a microphone attached to the head set. It took 40 minutes in total to administer both tasks.

2.5 Materials

Task 1 (Acceptability judgment task) – perception

The pre-recorded material consisted of 36 pairs of questions and answers between two native English speakers. To control the intonation of the sentences, the questions were recorded separately from the answers and edited later. In all 36 sets, half of the responses had two object nouns in given-new information order which was “contextually motivated” by the previous interrogative sentence as in (3a) and (3b); the rest had two object nouns in new-given information order as in (4a) and (4b), which would be inappropriate unless new information was stressed when being read.

(3) a. What did Eddie offer to Patti?
    Eddie offered Patti some tea.
b. Who did Kate give a watch to?
    Kate gave a watch to David.

(4) a. What did Chris give Irene?
    Chris gave flowers to Irene.
b. Who did Charlie offer a coke to?
    Charlie offered Sheila a coke.

Then the participants were asked to indicate the naturalness of the answer by circling a number from 1 (totally unnatural) through 7 (totally natural) on the Likert scale as they heard each
pair. The scoring for each response in the judgment was based on (a) correctness, i.e., whether or not the participants were able to produce dative structures with given–new ordered NPs as in (3a) and (3b), and (b) the acceptability judgment rating. Correct responses were assigned the positive value of the acceptability judgment rating and incorrect responses received the negative value. Scores thus ranged from +1 to +7 for correct responses and -1 to -7 for incorrect answers. The total of acceptability judgment scores on each dative construction converted into percentages serve as the first variable of this study.

Task 2 (Elicited production task) – production

The experiment for Task 2 took place at a language lab of the University of Hawai‘i as Task 1 was conducted. Each participant sat in an individual booth, listened to pre-recorded audio material to elicit production through an individual head set, and their answers were recorded. In this task, it was necessary to elicit either the double–object dative (DOD) structure as in (3a) or the prepositional dative (PD) structure as in (3b) motivated in a natural context. Participants listened to a short passage which presented descriptions of the agent and either the recipient or the theme. Then, they were asked a question about either (a) an unknown recipient who was transferred a theme object by the agent, or (b) an unknown theme object that was being transferred to the recipient. An example of one of the pre–recorded situations and questions to elicit DOD structures is given below:

(5) (A pre–recorded situation and a question to elicit a DOD structure)

Ann found Jim wearing a cast on his leg. Ann offered something so that Jim could sit on it. What did Ann offer to Jim?

In this example, the characters (Ann and Jim) are assumed as given information, and the theme object (what) is assumed to be focused attention. The response section included a picture of a chair, and participants were expected to answer as follows:

(6) (The expected answer)

Ann offered Jim a chair.

The participants had to identify the unknown recipient or object from the picture, and then utter a response to the question asked about the recipient or object. This format was adapted and developed from the task used by Wolfe–Quintero (1992). The scoring for each response was based on whether or not the participant could utter a dative construction in given–new information order. Dative responses in which information was given–new ordered were assigned one point. Since there were 36 responses in total, scores thus ranged from 0 to 36. For comparison with Task 1, the scores were converted into percentages.

The situation and question were heard only once in order to avoid encouraging participants to analyze the sentence. In the booklet given to each participant, the final question of the situation was repeated (e.g., What did Ann offer to Jim? ), and the verb they were to use was provided in parenthesis (e.g., offered).
3. Results and Discussion

First, two-way repeated measures ANOVA is performed for mean comparisons. The first independent variable of interest was the difference between NSs and NNSs. This variable is labeled Group in the analyses reported below, and it has two levels: NS and NNS. The other independent variable was the difference between elicited production scores and acceptability judgment scores. This variable is labeled Task Type, and it also has two levels: production and perception. The dependent variable is the total scores of the two tasks, and it is labeled Given-New Information Sensitivity. Descriptive statistics were computed, and a two-way ANOVA was calculated with Groups as the single between-groups factor and Task Type as a within-groups factor. The null hypotheses of no difference between groups was adopted. The overall α was set at <.05. Second, scatter plots are drawn, and correlation coefficients are calculated in order to show the relationships between two variables, i.e., production and perception within each group.

The means (M) and standard deviations (SD) for Given-New Information Sensitivity for two tasks as scored by NSs and NNSs are shown in Table 1. The table shows that NSs’ mean scores of Given-New Information Sensitivity are higher for the two tasks than NNSs’.

<table>
<thead>
<tr>
<th></th>
<th>Production</th>
<th>Perception</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M(SD)</td>
<td>M(SD)</td>
</tr>
<tr>
<td>NSs</td>
<td>87.38 (11.38)</td>
<td>60.03 (9.89)</td>
</tr>
<tr>
<td>NNSs</td>
<td>48.02 (19.50)</td>
<td>51.54 (5.46)</td>
</tr>
</tbody>
</table>

The ANOVA results are presented in Table 2. There were significant effects for Group, \( F(1, 68) = 125.78^* \), and Task Type, \( F(1, 68) = 31.16^* \). This indicates that there are consistent mean differences between the total scores by NSs and those by NNSs; as well as between the production scores and perception scores.

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>( F )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>20039.93</td>
<td>1</td>
<td>20039.93</td>
<td>125.78</td>
<td>.0001*</td>
</tr>
<tr>
<td>Task Type</td>
<td>4965.28</td>
<td>1</td>
<td>4965.28</td>
<td>31.16</td>
<td>.0001*</td>
</tr>
<tr>
<td>Group x Task Type</td>
<td>8338.99</td>
<td>1</td>
<td>8338.99</td>
<td>52.34</td>
<td>.0001*</td>
</tr>
<tr>
<td>Residual</td>
<td>21668.20</td>
<td>136</td>
<td>159.33</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(* p < .05\)
The results of the study also show a significant interaction between Group x Task, $F(1, 68) = 52.34^*$ (Figure 1). This interaction indicates that NSs are more sensitive to given-new information order of dative sentences when they produce than when they perceive. On the other hand, although the NNSs' scores are lower than the NSs' in both perceiving and producing given-new information ordered dative structures, the NNSs' scores do not interact with the type of tasks as much as the NSs' scores do.

Figure 1. Interaction between Group and Task in NSs' and NNSs' of Given-New Information Sensitivity

The scattergram and regression for the production and perception within NSs is shown in Figure 2, and its correlation coefficient is .26. $r^2$ is $(.26)^2 = .068$. This means that only 6.8% of production scores is accounted for by perception scores or vice versa. The scattergram indicates that most of production scores of the NSs range from 70 to 98% accuracy, while most of perception scores range from 45 to 70%. This means that the NSs' scores are relatively lower in grammatical judgements than in elicited production tasks.

Figure 2. Scattergram and regression for production and perception within the NS
The scattergram and regression for the production and perception within NNSs is shown in Figure 3, and its correlation coefficient is .63. \( r^2 \) is \( (.63)^2 = .400 \) This means that 40% of production scores is accounted for by perception scores or vice versa. The scores for both grammatical judgments and elicited production tasks by the NNSs are much more highly correlated with each other than the scores by the NSs are.

Figure 3. Scattergram and regression for production and perception within the NNS

![Scattergram and regression](image)

Research question 1 (RQ1) was concerned with the differences between the NS's and the NNS's performance on production and perception of the relationship between information order and the dative alternation. Based on the results from the previously conducted research in which Japanese dative alternation does not have regularity in the relationship with information as English does, it was hypothesized that the NNS would have difficulty in learning the relationship in the English dative structure. The results show that the NS produce and perceive dative sentences in correct information order, i.e., given-new information order, more frequently than the NNS do. There was an interesting interaction between two main effects: Group and Task Type. This indicates that the NS are more sensitive to the correct information order when they produce it than when they perceive it. On the other hand, the NNS exhibit much less sensitivity to the correct information order when they perceive it as well as when they produce it. This interaction takes clearer view later when correlation between production and perception is discussed.

Research Question 2 (RQ2) was concerned with whether or not the NS's performance on production and perception of the dative alternation in correct information order is correlated. Then Research Question 3 (RQ3) was concerned with whether or not the NNS's performance on the two tasks are correlated as much as the NSs'. As shown in Figure 2, the correlation coefficient for the NS turned out to be inconsistent with a study by Greenbaum and Quirk (1970). In the study, Greenbaum and Quirk (1970), using grammatical judgments and production tasks,
indicate that the NS have a tendency to rely on the same grammar for both the sentences they perceive and those they produce. On the contrary, the results from the present study show that the NS's scores of elicited production and acceptability judgments are much less correlated than the NNS's scores are. The NS produced correct information ordering at 87.4% of whole elicited linguistic utterances, however, they perceived given–new information ordering much less correctly in their acceptability judgments (60.0%). The reason behind these results could be explained with an extra-linguistic factor: probably it is due to the NS's inclination to base their ratings on their willingness to tolerate usage in others that does not correspond to their own habitual forms. These findings of the current study in correlation between production and perception in the NS's performance basically support the argument of Sorace (1996) in which she argues that "linguistic intuitions reflect grammatical competence, but they are also open to the influence of other cognitive or contextual factors" (Sorace, 1996, p.378).

4. Conclusion

It is noted that the results are partially inconsistent with the arguments by Flynn (1986). She demonstrates that production tests primarily evaluate a learner's knowledge of the second language while comprehension tests principally evaluate a learner's sensitivity to pragmatic context. The results from L1 data follow Flynn's argument, however, L2 data do not agree with the argument. This means that consideration must be given to the requirements of each task in terms of understanding what aspect of linguistic knowledge that task most directly measures. The present study shows that production tasks measure an L1 speaker's competence on a discourse factor, while grammatical judgments may not be able to evaluate the competence of the L1 speaker as accurately as the production tasks. As was pointed by Sorace (1996), the study indicates that the L1 speaker's grammatical judgments are open to the effect of extra-linguistic factors. Researchers should not always assume that results from L1 grammatical judgments indicate the stable state of knowledge of the native speaker. This is not to say that only production tasks should be used in comparing L2 grammars with the L1 competence. Rather, we need focused evidence from a wide range of experimental tasks in which methodology is accurately controlled, and we must be informed about how different tasks relate to each other.

References

Inagaki, Shunji. "The acquisition of constraints on the dative alternation in English by native


