Understanding Pronouns in EFL Reading:
Processing Differences between
Skilled and Less-skilled Learners*

Tatsuhiro YOSHIDA
Hyogo University of Teacher Education

The process of pronoun understanding in skilled and less-skilled EFL learners was investigated based on approaches used in L1 studies. Japanese university students at two different levels of English language proficiencies as determined by a standard test were asked to read the sentences which included the pronouns he or she, and their reading times and comprehension accuracy were measured. The result showed that the less-skilled learners read the sentences focusing on the morphosyntactic information of the pronoun, whereas the skilled learners encoded the semantic information as well as the morphosyntactic cues. The reasons why the skilled learners performed better are discussed.

1. Introduction

Pronouns are classified by Halliday & Hassan (1976) into ‘reference’, one of the cohesive devices. A pronoun almost always presupposes an item which has already been mentioned (pronouns are occasionally used to refer cataphorically to items appearing in the following text). Halliday & Hassan emphasize that the link between a pronoun and its antecedent, which is a semantic relation, establishes cohesion and, thus, contributes to the reader’s (or listener’s) comprehension of the text. Since establishing cohesive links is a significant task for learners of English as a foreign language (EFL), investigating how some learners successfully understand pronouns in a text while others fail to do so would appear to be worthwhile. Surprisingly, however, few such studies have been conducted so far in the EFL context. In the present paper, I will first review the approaches and findings of research conducted in the first language (L1) context and raise some central issues on pronoun understanding. I will then conduct an experiment which was designed based on this L1 research to determine how EFL learners
understand pronouns and what causes processing differences between skilled and less—skilled learners.

2. Research findings from L1 studies

How people understand pronouns in a text has been actively investigated in psycholinguistics. A very simple picture of the process is that when a pronoun is encountered, the reader first searches for its antecedent and then incorporates it into the current interpretation of the text. Oakhill, Garnham & Vonk (1989) surveyed recent studies on anaphoric resolution, from which they list several linguistic and cognitive factors which affect the process:

(1) the distance between a pronoun and its antecedent (Daneman & Carpenter, 1980; Ehrlich & Rayner, 1983)
(2) the nature of the material between a pronoun (or anaphor) and its antecedent (i.e. whether the antecedent is kept in focus in the intervening material) (Clifton & Ferreira, 1987)
(3) the grammatical position of the antecedent (i.e. whether the antecedent is in the subject or object position)
(4) the availability of the pronoun’s morphosyntactic information (e.g. gender and number) (Caramazza, et al. 1977; Ehrlich, 1980; Garnham & Oakhill, 1985; Vonk, 1984)
(5) the bias of verb semantics (i.e. implicit causality) (Caramazza, et al. 1977; Ehrlich, 1980; Garnham & Oakhill, 1985; Vonk, 1984)

Much still remains to be investigated in this area, and here I am unable to consider all the above factors due to lack of space (for further discussion, see Oakhill, Garnham & Vonk, 1989). I will, therefore, concentrate on the two factors which are most actively discussed in psycholinguistic studies, i.e. gender cue and implicit causality; the reason why I believe these factors to be especially significant is explained in the following section. First, however, I will briefly summarize previous studies on how these two factors influence the process of pronoun understanding by native speakers of English.

(1) Gender Cue

[1a] Ken lent a pen to Mary because she wanted to write a letter.
[1b] Susan lent a pen to Mary because she wanted to write a letter.

In [1a] it is obvious that she in the subordinate clause refers to Mary, but in [1b] the referent of she cannot be identified thoroughly by its gender cue. It is only by the meaning of the sentence and our pragmatic knowledge (i.e. a person who wants to write a letter usually needs a pen) that the antecedent of the pronoun is identified as Mary. I will refer to the condition in which the referent of a pronoun is identified by the gender of the pronoun (i.e. [1a]) as the Gender Cue condition, and that in which there is no gender cue to identify the referent (i.e.[1b]) as the No Gender Cue condition. Intuitively, [1b] is cognitively more demanding than [1a]. Caramazza,
et al. (1977), Ehrlich (1980), Garnham & Oakhill (1985) and Vonk (1984) investigated which type of sentence is understood faster. Subjects were presented with sentences on a computer screen and the time taken to read them was measured. The results provided evidence supporting our intuitive hypothesis; [1b] was read significantly more slowly than [1a].

(2) Implicit causality

In addition to morphosyntactic cues, such as gender, the semantic factor also affects the process. Caramazza and his colleagues (Caramazza, et al. 1977; Garvey & Caramazza, 1974) presented subjects with sentence fragments like [2] and [3] and asked them to complete the sentences. In all the sentences the participants in the main clause, such as John and Bill, were of the same gender.

[2] John sold his car to Bill because he...
[3] John punished Bill because he ...

The result was very interesting; the subjects' perception of the referent of the subject pronoun he was biased toward either the subject or the object according to the verb in the main clause. For [2], for example, significantly more subjects identified the referent of he as John and continued the sentence fragment as in [2a] rather than as in [2b];

[2a] John sold his car to Bill because he needed some money.
[2b] John sold his car to Bill because he needed a means of transport.

On the other hand, for [3], Bill was more likely to be perceived as the referent of he and, thus, a sentence such as [3b] was produced more than [3a].

[3a] John punished Bill because he was very annoyed.
[3b] John punished Bill because he was very annoying.

These results indicate that verb meanings, causality in this case, affect the subjects' assignment of the referent of the pronoun. For instance, when A sells something to B, the causal source of the action is more likely to be found in A (the seller) than B (the buyer), whereas, when A punishes B, the causal source is more likely to be found in B (the punishee) than A (the punisher).

The same researchers further examined whether verb meaning affects on-line reading time. Subjects were asked to read sentences like [2a]-[3b] shown on a computer screen. The time taken to read each sentence was measured and analyzed. A basic finding was that the subjects read sentences like [2a] and [3b] significantly faster than those like [2b] and [3a], indicating that the subjects predicted the referent of the pronoun based on the verb's semantic information. Thus, when the structure of the sentence they read was congruent with the verb's implicit causality, the sentence was read faster than when it was not. I will refer to the former condition as the Congruent condition and the latter as the Incongruent condition.
3. Pronoun understanding in EFL reading

Having described the basic findings on pronoun resolution in L1 studies, I will now turn my attention to the pronoun understanding of EFL learners. In the EFL context, very few studies of on-line pronoun understanding have been conducted. One of the reasons for this may be that understanding pronouns such as he and she has not been seen as posing a serious problem to EFL learners; for example, Japanese EFL learners learn the system of the personal pronoun at a very early stage of the junior high school course. However, since various factors have been shown to affect pronoun resolution by native speakers of English, can EFL learners really be expected to understand pronouns so easily? Assuming that some or all of the factors mentioned in Section 2 do affect EFL learners' understanding of pronouns, my concern further goes to the question whether there are processing differences between skilled and less-skilled EFL learners. Given that skilled learners' perception and use of linguistic information is doubtlessly different from that of less-skilled learners, it should be possible to identify the processing difference which may partly cause the overall difference in reading proficiency.

In the next section, an experiment on the pronoun understanding of EFL learners is described, focusing on how gender cue and implicit causality affect learners' pronoun understanding.

There are two reasons why I choose to examine the effects of these two factors out of the five mentioned earlier. First, as I have already described, gender cue provides surface, morphosyntactic information, whereas implicit causality is the verb's semantic information. By exploring the effects of these two factors, it should be possible to clarify which aspects of pronoun understanding cause the overall difference between skilled and less-skilled learners. As Jenkin, et al. (1993) found, EFL learners are constrained to encode linguistic information 'propositionally' (i.e. in the form of surface linguistic information) because of their lack of native speaker-like proficiency. However, Jenkin, et al. did not investigate whether there is a difference in the way of encoding linguistic information between skilled and less-skilled learners.

Secondly, the effect of implicit causality is closely related to the learners' prediction ability. Implicit causality biases readers in favor of a particular source of causation. Thus, implicit causality has a facilitating effect only if a learner can successfully utilize the verb's semantic information and predict the following part of the sentence. Consequently, if there is correlation between learners' language proficiency and their ability to predict a up-coming material in a text, skilled learners will be more likely to benefit from implicit causality.

Assumptions about how the two factors affect EFL learners' processing can thus be summarized as follows;

(1) EFL learners understand pronouns based more on gender cue (morphosyntactic factor) than on implicit causality (semantic factor).
If skilled EFL learners are proficient enough to predict the subsequent part of the sentence, they will be affected by the implicit causality of a verb in the main clause. Less-skilled learners will be affected only by gender cue; when gender cue is not available, whether the condition is Congruent or Incongruent, less-skilled learners will perform poorly.

(3) There will be an overall processing difference (i.e. in reading time and comprehension) between the two groups.

4. Experiment

(1) Method

Subjects Subjects were first-year students at Hyogo University Teacher Education. They have all learned EFL for 6–7 years and at the time of the experiment were taking an EFL class twice a week as a university requirement. None of the subjects had experience of studying or living overseas. A pretest based on the Standard Test of English Proficiency, Second Grade (Max score=60) was given to sixty-five students, from whom 13 ‘skilled’ and 12 ‘less-skilled learners’ were selected based on their scores ($t=7.133, p<.001$).

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean (max=60)</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skilled learners</td>
<td>13</td>
<td>38.8</td>
<td>4.26</td>
</tr>
<tr>
<td>Less-skilled learners</td>
<td>12</td>
<td>25.8</td>
<td>4.43</td>
</tr>
</tbody>
</table>

Materials All the experimental materials used were formulated based on previous L1 studies. There were 28 experimental sentences, each consisting of two clauses. The first was a main clause, which included one of the following verbs; *give, send, lend, ask ... a question, telephone, help, and kill*, describing the interaction between two participants. The second was a subordinate clause which began with the conjunction *because*; the subject of the subordinate clause was always either *he* or *she*, which referred to either of the participants mentioned in the main clause (see [1a]–[3b] above).

The two participants in the main clause were sometimes of different genders (Gender cue condition) and sometimes of the same gender (No gender cue condition). Each main clause was followed by a subordinate clause in such a way that the subject of the subordinate clause was either congruent (Congruity condition) with the implicit causality of the verb in the main clause, or incongruent with it (Incongruity condition). The bias of the verbs was determined based on Caramazza, et al. (1977). Thus, four versions of each sentence were constructed, as in [4a]–[4d]:

135
[4a] Jim sold the computer to Mary, because he needed some money.  
[4b] Tom sold the computer to Mary, because she needed it.  
[4c] Tom sold the computer to Jim, because he needed some money.  
[4d] Mary sold the computer to Susan, because she needed it.  

(GC = Gender Cue, NG = No Gender Cue, CON = Congruity, INC = Incongruity)

In addition to the experimental sentences, a further twenty-eight filler sentences were presented in order to distract subjects' attention from pronoun understanding. Examples of these are;

[5a] Jim finished eating three apples, before Mary finished two.  
[5b] Mary came to the table, after Jim finished his dinner.

Procedure    Each sentence was divided into two parts and successively presented on a computer screen. The first part included the main clause and the conjunction because, while the second part comprised the rest of the sentence. The subject first saw the symbol **** on the screen, which signaled the beginning of the trial. When the space bar was pressed, the first part appeared on the screen. The subject could proceed to the second display when he felt he understood the first part (self-paced reading). To encourage the subjects to fully understand each sentence, a simple comprehension check followed the second display (e.g. [6a], [6b]). This was designed so that subjects had to determine which participant in the main clause the pronoun referred to. It was answered by pressing keys labeled “True” or “False”:

[6a] Jim sold the computer to Mary, because he needed some money.
       Comprehension check: Mary needed some money.

[6b] Mary sold the computer to Susan, because she needed it.
       Comprehension check: Susan needed the computer.

The experiment was conducted on an individual basis in a quiet room. The subjects saw twenty-eight experimental sentences and twenty-eight filler sentences. To reduce the effect of the presentation order, four versions with different presentation orders were constructed and one of them was randomly assigned to each subject. For each sentence the subject’s response time (RT) for the second part of the sentence and for the comprehension check, as well as the comprehension accuracy (true or false), were recorded by the computer and submitted to statistical analyses. In each case before starting the experiment, 10 practice sentences were given to ensure that subjects were familiar with the procedure. If subjects felt tired during the experiment, they were permitted to rest as long as they wished between presentations.

Apparatus    The experiment was controlled on-line by an NEC-9801VM personal computer running an originally developed program. Each part of the sentence was presented in the center of a CRT (NEC N5913L). The subjects responded to the presentation by pressing the space bar and the keys labelled “True” or “False” on the keyboard.

136
Prediction Gender cue will have an overall effect on the performance of both groups. Also, since the information which less-skilled learners are able to utilize will be restricted to gender cue, their performance will be better when only gender cue is available (i.e. [GC+ CON] and [GC+ INC]) than when it is not (i.e. [NG+ CON] and [NG+ INC]). In the case of the skilled learners, the effects of gender cue and implicit causality are expected. Although the effect of gender cue will still be dominant, skilled learners will perform better under NG conditions than the less-skilled when the referent of a pronoun is congruent with the verb bias (i.e. [NG+ CON]). In short, the performance of each group will be:

Skilled Learners

<table>
<thead>
<tr>
<th>Gender Cue</th>
<th>No Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Congruent</td>
<td>Incongruent</td>
</tr>
<tr>
<td>[GC+ CON]</td>
<td>[GC+ INC]</td>
</tr>
<tr>
<td>[NG+ CON]</td>
<td>[NG+ INC]</td>
</tr>
</tbody>
</table>

Less-skilled Learners

<table>
<thead>
<tr>
<th>Gender Cue</th>
<th>No Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Congruent</td>
<td>Incongruent</td>
</tr>
<tr>
<td>[GC+ CON]</td>
<td>[GC+ INC]</td>
</tr>
<tr>
<td>[NG+ CON]</td>
<td>[NG+ INC]</td>
</tr>
</tbody>
</table>

(2) Results

RTs for the subordinate clauses and for the comprehension check and the accuracy of comprehension were subjected to a three-factor ANOVA (2X2X2) with repeated measures, with gender cue (Gender Cue-No Gender Cue), implicit causality (Congruent-Incongruent) and proficiency (skilled-less-skilled learners) as fixed factors and subjects [F1] or items [F2] as a random factor.

RT for subordinate clauses There was a main effect of gender cue [F1(1,23)=31.604, \( p<.001 \); F2(1,12)=32.639, \( p<.001 \)]. For both skilled and less-skilled learners, reading took longer when gender cue was present than when it was not. [F1(1,23)=33.408, \( p<.001 \); F2(1,12)=30.163, \( p<.001 \) for the skilled; F1(1,23)=4.711, \( p<.05 \); F2(1,12)=6.694, \( p<.05 \) for the less-skilled]. No significant difference in RT between the skilled and the less-skilled groups was observed. The interaction between proficiency and gender cue was significant [F1(1,23)=6.515, \( p<.05 \); F2(1,12)=4.219, \( .05<p<.10 \)]

<table>
<thead>
<tr>
<th></th>
<th>Congruent</th>
<th>Incongruent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender Cue</td>
<td>No Gender</td>
<td></td>
</tr>
<tr>
<td>Skilled Learners</td>
<td>3566</td>
<td>4311</td>
</tr>
<tr>
<td>Less-skilled Learners</td>
<td>3471</td>
<td>4045</td>
</tr>
</tbody>
</table>

Table 2. Mean RT for subordinate clause (msec)

RT for comprehension checks The only main effect observed was gender cue [F1(1,23)=42.917, \( p<.001 \); F2(1,12)=30.843, \( p<.001 \)]. No other effects were observed.
Table 3. Mean RT for comprehension checks (msec)

<table>
<thead>
<tr>
<th></th>
<th>Congruent</th>
<th>No Gender</th>
<th>Incongruent</th>
<th>No Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skilled Learners</td>
<td>2460</td>
<td>3078</td>
<td>2558</td>
<td>3591</td>
</tr>
<tr>
<td>Less-skilled Learners</td>
<td>3098</td>
<td>3345</td>
<td>2669</td>
<td>3627</td>
</tr>
</tbody>
</table>

Accuracy of comprehension checks  Table 3 shows the mean proportions of the correct answers for each group. There was a significant difference between the skilled and the less-skilled learners \( F(1,23)=5.021, p<.05; F(1,12)=7.280, p<.05 \). Gender cue also had a main effect \( F(1,23)=42.869, p<.001; F(2,12)=45.096, p<.05 \). An interaction was observed between proficiency and gender cue \( F(1,23)=3.602, 05<p<.10; F(1,12)=4.837, p<.05 \).

Table 4. Mean proportion of comprehension accuracy (%)

<table>
<thead>
<tr>
<th></th>
<th>Congruent</th>
<th>No Gender</th>
<th>Incongruent</th>
<th>No Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skilled Learners</td>
<td>96.7</td>
<td>91.2</td>
<td>97.8</td>
<td>82.4</td>
</tr>
<tr>
<td>Less-skilled Learners</td>
<td>95.2</td>
<td>79.7</td>
<td>95.2</td>
<td>73.8</td>
</tr>
</tbody>
</table>

5. Discussion

Analyses of RTs for the second displays and comprehension checks show greater overall effects of gender cue, as was expected; that is, learners of both groups were able to read the second part of the sentence faster when gender cue was available. This suggests that the subjects, unlike native speakers of English, read the sentences based on the surface, morphosyntactic information of the sentences. However, no significant differences were observed between the two groups for RTs. It may be reasonable to suppose that this was because the sentences presented were not long enough to create RT differences. Alternatively, it could be that the proficiency difference was not large enough to show RT differences. This issue must be further investigated in a longer passage with different subjects. The reason why RT differences did not appear could be due to the fact that the Congruity condition did not facilitate the skilled learners' RTs. This seems contradict the prediction above.

However, there was an effect of Congruity not on the skilled learners' reading speed, but on their comprehension. This is clearly shown by an analysis of the comprehension accuracy. When gender cue was available, there was almost no difference in the mean accuracy between the two groups (97.3% for the skilled vs. 95.2% for the less-skilled) but when gender cue was not available, a significant difference was observed (86.8% vs. 76.8%, respectively). Since there
was no two-way interaction between gender cue and implicit causality, caution is required in further discussion of the interaction, but it is safe to say that the difference was the result of differences in processing between the No Gender Cue+Congruent and No Gender Cue+Incongruent conditions. The skilled learners were able to answer fairly accurately under the No Gender Cue condition when semantic information was congruent with the structure; i.e. when their prediction about the referent of a pronoun was correct (91.2%). However, the less-skilled learners answered poorly under the conditions in which gender cue was not available (79.7% and 73.8%). This clearly indicates that less-skilled learners are likely to focus on morphosyntactic, surface information, whereas skilled learners are able to encode not only the morphosyntactic information but also the verb's semantic information. Accuracy of the answers by the skilled learners was the lowest under No Gender Cue+Incongruent condition (82.4%), but it was still higher than that of the less-skilled. This also shows that the skilled learners were better at making inferences about the referent of a pronoun when there was no morphosyntactic clue to identify it.

As Garnham & Oakhill (1982) suggested, two participants of different genders in the main clause are distinctively represented in memory and recalled better. In addition to this, the ability to make use of the verb's semantic information and to predict up-coming material may contribute to establishing a stable mental representation of the sentence. In short, the skilled learners performed better in answering the questions because they could utilize multiple sources of information, while the less-skilled learners could access only surface linguistic information, resulting in the poorer performance in comprehension.

6. Conclusion

The process of pronoun understanding by skilled and less-skilled EFL learners was examined. The results showed that gender cue was a factor which greatly affected the processing of both groups. Although congruity of the implicit causality of the verb with the sentence structures did not facilitate the learners' RTs, it tended to affect the accuracy of the skilled learners' answers. This difference in the way of drawing upon information may be related to the overall difference in the learners' language proficiency; however, experimental studies need to be conducted to verify this. Here, the experimental materials were restricted to single sentences and a particular sentence structure (i.e. Noun+Verb+Noun because pronoun...). It will be more informative to investigate how learners process anaphoric expressions in a longer text and establish coherent comprehension of it. Also, to account for the processing difference more reliably and precisely, the number of subjects in each group should be increased and their language proficiencies reliably determined.
I would like to express my gratitude to Professors Masamichi Tanaka, Toshihiko Yamaoka and John Chick, Hyogo University of Teacher Education for their valuable and critical comments on the early versions of the paper. Any remaining inadequacies are, of course, my own. My special thanks are also due to Yuya Koga, Hiroshima University graduate school, for developing a computer program for the experiment.

References


