Pronoun Advantage in L2 Relative Clause Processing

Jinsun Choe and Kamil Ud Deen

This study presents the results of an online self-paced reading task of Korean L2 learners of English reading subject and object relative clauses (SRCs, ORCs). We manipulate the kind of nominal inside RCs: lexical NP as in (1) and pronoun as in (2).

(1) a. The banker that _ praised the barber climbed the mountain.
   b. The banker that the barber praised _ climbed the mountain.

(2) a. The banker that _ praised you climbed the mountain.
   b. The banker that you praised _ climbed the mountain.

We find that the pronoun significantly improves reading times of ORCs, showing that L2 learners, like native English speakers, are sensitive to discourse accessibility of referents in processing long-distance dependencies.

The present paper is organized as follows. Section 1 introduces and reviews previous research on acquisition and processing of RCs in first and second language. Section 2 describes the method of the current experiment which tested Korean learners’ processing of English RCs. Section 3 presents the results of the comprehension question accuracy and of the online reading time and discusses the findings. Lastly, section 4 summarizes the results and concludes the paper.

1. Previous research on relative clauses
1.1. L1 acquisition and processing studies

Relative clauses (RCs) have attracted great interests in psycholinguistic research. There is a general consensus from numerous studies that ORCs (3b) are harder to process, read, and are acquired later by children than SRCs (3a), when both the head noun (the banker) and the embedded noun (the barber) are lexical NPs (e.g., Kidd & Bavin, 2002; King & Just, 1991; King & Kutas, 1995; Tanaka, 2016; Traxler, Morris, & Seely, 2002; Wanner & Maratsos, 1978).

*Jinsun Choe, Korea University of Technology and Education, jchoe@koreatech.ac.kr; Kamil Ud Deen, University of Hawaii at Manoa, kamil@hawaii.edu. This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2018S1A5A8029817).

This finding has been shown by a variety of measures, including reading time and eye movement. However, this well-known asymmetry is significantly reduced when the embedded noun is a pronoun, as in (4) (e.g., Gordon, Hendrick, & Johnson, 2001; Haendler, Kliegl, & Adani, 2015; Reali & Christiansen, 2007; Warren & Gibson, 2002).

Gordon, Hendrick, and Johnson (2001) found that native English speakers showed shorter reading times and higher comprehension accuracy when the embedded NP inside the relative clause was a pronoun, compared to a lexical noun phrase. Numerous accounts have been proposed to explain this finding. One prominent proposal (the Dependency Locality Theory, DLT) states that processing costs increase with the number of new discourse referents that intervene between two elements that make up a dependency (Gibson, 1998, 2000). Object relatives (3b) are thus, harder to process than subject relatives (3a) because two new discourse elements (the barber and praised in (3)) intervene between the relative head and the gap in object relatives, whereas the dependency in subject relatives is resolved immediately without any new discourse elements intervening. In terms of assessing the strength of intervention of the new discourse referents, this theory adopts the Givenness Hierarchy, proposed by Gundel, Hedberg & Zacharski (1993) in (5), which correlates NP types with the accessibility of their referents.

According to this hierarchy, referent types that are reduced in form, such as pronouns, are more accessible than less-focused and peripheral referential expressions, which are often expressed as a full NP. Therefore, as pronouns are not considered new discourse referents, the DLT predicts less processing cost for ORCs containing a pronoun, as in (4).

1.2. L2 acquisition and processing studies

In L2 processing of English RCs, a similar subject-object asymmetry exists when RCs contain lexical NPs, as in (3). (e.g., Doughty, 1991, Eckman et al., 1988; Gass, 1979; Hamilton, 1994; Schachter, 1973; Wolfe-Quintero, 1992). To investigate the effects of accessibility of different types of discourse referents in L2 relative clause processing, Kim (2016) conducted offline and online
comprehension tasks. In the offline matching task, L2 learners were asked to match a subject and its corresponding predicate within a doubly embedded sentence containing a pronominal or a nominal referent, as in (6).

(6) The student who the professor who [I/the scientist] collaborated with at the lab advised copied the article.

The results showed that the L2 learners were more accurate in identifying a subject of various predicates in the pronoun condition, compared to the full NP condition, confirming the advantage of pronouns. Similar results were obtained in the online task, which was a self-paced reading task. In particular, this task inspected L2ers’ sensitivity to an error arising from a morpho-syntactic mismatch between a subject and its predicate.

(7) The book that [I/the student] bought yesterday was [interesting/*interested].

It was observed that in the regions 5 (bought) and 6 (yesterday), the L2ers’ RT was shorter in the pronoun condition than in the full NP condition, although the pronoun advantage did not lead to a significant difference in region 8 (interesting/interested). Overall, Kim (2016) demonstrated that the pronoun facilitated L2 sentence comprehension and processing at least for a subset of Korean L2ers of English. However, only non-reversible ORCs with an inanimate head noun were tested in Kim’s (2016) online task. In the present study, therefore, we used both types of reversible RCs to explore the effect of the pronoun across RC type.

2. Method
2.1. Participants

Eighty-three Korean-speaking L2 learners of English participated in this study (along with fifteen native English speaker controls). The L2 learners were recruited from a university in Seoul, Korea, and the native English speakers were recruited from a university in the United States of America. Both the native speakers and L2 learners were paid $5 for their participation.

2.2. Design and materials

The present study employed a 2 x 2 factorial design, crossing RC type (SRC vs. ORC) and the Embedded NP type (Lexical NP vs. Pronoun). There were six sentences for each condition, and thus, a total of twenty-four experimental sentences were created. These sentences were distributed across four lists in a Latin Square design. A sample set of experimental items is shown in Table 1 below. Finally, there were 76 filler sentences which were aimed at distracting the participants from guessing the research question of the study.
Table 1. A sample set of experimental items

<table>
<thead>
<tr>
<th></th>
<th>RC type</th>
<th>Embedded NP type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SRC</td>
<td>Lexical NP</td>
<td>The nurse that <strong>ignored the doctor</strong> last evening answered the phone quickly.</td>
</tr>
<tr>
<td>2</td>
<td>SRC</td>
<td>Pronoun</td>
<td>The nurse that <strong>ignored you</strong> last evening answered the phone quickly.</td>
</tr>
<tr>
<td>3</td>
<td>ORC</td>
<td>Lexical NP</td>
<td>The nurse that <strong>the doctor ignored</strong> last evening answered the phone quickly.</td>
</tr>
<tr>
<td>4</td>
<td>ORC</td>
<td>Pronoun</td>
<td>The nurse that <strong>you ignored</strong> last evening answered the phone quickly.</td>
</tr>
</tbody>
</table>

2.3. Procedure

Participants performed self-paced reading in a word-by-word moving window display. At the start of each trial, a sentence appeared on the screen with all characters replaced by dashes. Participants pressed the space bar to change a string of dashes into a word. Each time the space bar was pressed, the next word appeared and the previous word reverted back into dashes. The time between space bar presses was recorded as their reading time. Participants were instructed to read the sentences at their natural speed. After each sentence, a true/false comprehension question was presented that asked about the content of the preceding sentence. Half of the comprehension questions were true and half were false. For the experimental items, correct answers required an understanding of the syntactic-semantic relations between the two NPs and the matrix verb or the embedded verb in the RC.

3. Results and discussion

Before the data analysis, participants with lower than 80% accuracy on the filler items were excluded. From this criterion, none of the native English speakers’ data were eliminated but seventeen L2 learners’ data were removed. The analysis below are thus from the remaining sixty-six L2 learners and the fifteen native English speakers.

3.1. Comprehension question accuracy

First, we looked at the accuracy of true/false comprehension questions. The results are presented in the form of mean accuracy for the native L1 English speakers in Figure 1.
Accuracy was significantly higher for SRCs than for ORCs ($F_1(1,14) = 35.76, p < .001; F_2(1,23) = 18.04, p < .001$). It was also higher for the Pronoun condition than for the Lexical NP condition ($F_1(1,14) = 9.89, p = .007; F_2(1,23) = 9.14, p = .006$). Finally, there was a marginally significant interaction effect between RC type and the embedded NP type in the participants analysis ($F_1(1,14) = 3.65, p = .08$) but not in the items analysis ($F_2(1,23) = 1.05, p = .32$), indicating that the effect of RC type was greater for the Lexical NP condition than for the Pronoun condition. Figure 2 below presents the comprehension question accuracy results for the L2ers.
The accuracy results of L2 learners showed a similar pattern as that of the native speakers. The accuracy was higher for SRCS than for ORCs ($F_1(1,65) = 43.97, p < .001; F_2(1,23) = 12.61, p = .002$), and for the Pronoun condition than for the Lexical NP condition ($F_1(1,65) = 26.69, p < .001; F_2(1,23) = 34.03, p < .001$). However, there was no interaction effect ($F_1(1,65) = 1.89, p > .1; F_2(1,23) = 1.78, p > .1$).

### 3.2. Online reading time

Next, we looked at reading time (RT) results for sentences for which participants correctly answered the true/false comprehension question. To eliminate RT outliers, raw reading times were trimmed in the following ways. First, any RTs longer than 2500ms were replaced with 2500ms. Next, RTs more than 2.5 standard deviations above or below the mean for each region in each condition were discarded. Then, the residual reading time transformation was calculated from the raw RTs to control for differences in word length and reading speed. Residual RTs for each participant were obtained by calculating the expected reading time of each region (using a linear regression equation according to word length) and subtracting it from the raw RTs (Ferreira & Clifton, 1986).

There were two regions of interests for the present study. The first critical region, the dark-shaded part in Figures 3 and 4, labeled as R1, was the embedded noun for SRC and the embedded verb for ORC. The second critical region, the light-shaded part in Figures 3 and 4, labeled as R2, was the matrix verb of the sentence. Figure 3 presents the mean residual reading times of the native English speakers.

![Figure 3. Natives’ mean residual RTs](image)

For the first critical region, there was a marginal main effect of NP type by participants ($F_1(1,14) = 3.85, p = .07; F_2(1,23) = .71, p > .1$), indicating that the
Pronoun condition was read faster than the Lexical NP condition. And there was a marginally significant interaction effect, but only by items ($F_1(1,14) = 5.89, p = .029; F_2(1,23) = 3.51, p = .076$). Critically, when the two pronoun conditions were compared against one another, there was no difference between SRC and ORC ($ps > .1$). For the second critical region, there was also a main effect of NP type ($F_1(1,14) = 11.69, p = .004; F_2(1,23) = 20.46, p < .001$), but there was neither a main effect of RC type nor an interaction effect between the two. Finally, there was no effect of RC type between the two pronoun conditions ($ps > .1$), that is between the SRC-Pronoun condition and the ORC-Pronoun condition. Let us now turn to look at the mean residual reading times of the L2 learners, illustrated in Figure 4 below.

![Figure 4. L2ers’ mean residual RTs](image)

For the first critical region, there was a main effect of RC type ($F_1(1,65) = 3.73, p = .058; F_2(1,23) = 5.04, p = .035$), a main effect of NP type ($F_1(1,65) = 29.18, p < .001; F_2(1,23) = 34.51, p < .001$), and a significant interaction effect ($F_1(1,65) = 6.81, p = .011; F_2(1,23) = 8.60, p = .007$). As with L1 native speakers, there was no difference between SRC and ORC for the two pronoun conditions ($ps > .1$). For the second critical region, the only significant effect was that there was a main effect of NP type by participants ($F_1(1,65) = 4.56, p = .037; F_2(1,23) = 1.58, p > .1$): that pronoun conditions were read faster than the lexical NP conditions.

4. Concluding remarks

The present study investigated whether L2ers read ORCs faster when the embedded NP is a pronoun versus when it is a full lexical noun by conducting a self-paced reading task with Korean-speaking L2ers of English. There were three dependent variables: the accuracy of the true/false comprehension question that tested the meaning of the preceding experimental sentence, and residual reading
times for the two critical regions: For SRCs (8), the first critical region was the embedded noun phrase (*the doctor* in (8a) and *you* in (8b)), and for ORCs (9), the first critical region was the embedded verb inside the RC (*ignored* in (9a) and (9b)). For both SRCs and ORCs, the second critical region was the matrix verb of the sentence (*answered* in (8) and (9)).

(8) a. The nurse that ignored **the doctor** last evening **answered** the phone quickly.
   b. The nurse that ignored **you** last evening **answered** the phone quickly.

(9) a. The nurse that the **doctor** ignored last evening **answered** the phone quickly.
   b. The nurse that you **ignored** last evening **answered** the phone quickly.

The results can be summarized as follows. Both comprehension question accuracy and reading time data show that having a pronoun significantly reduces the processing cost for both the natives and L2 learners, resulting in better comprehension (as shown by higher accuracy) and faster processing (as shown by shorter RTs). In particular, the reading time data reveal no effect of the RC type when the pronoun condition was examined alone, indicating that an embedded pronoun eliminates the well-known subject-object asymmetry observed in numerous studies on L1 and L2 populations. These results demonstrate that not only L1 speakers’ but also L2 speakers’ processing cost is influenced by the discourse accessibility of the intervening constituent (Gibson, 1998, 2000). Further work is required in order to investigate the effect of accessibility of other types of discourse referents in L2ers’ sentence processing and how it interacts with different types of structurally complex sentences.

**References**


