When is the object relative clause easier to process than the subject relative clause?

Atsushi SATO†  Barış KAHRAMAN‡ and  Hiromu SAKAI††

† Faculty of Humanities, Nanzan University  18 Yamasato-cho, Showa-ku, Nagoya, 466-8673 Japan
‡ Faculty of Education, Canakkale Onsekiz Mart University  C1-217 17100 Canakkale, Turkey
†† Graduate School of Education, Hiroshima University  1-1 Kagamiyama, Higashi-Hiroshima, 739-8524 Japan

E-mail:  † sato.chiru.sakura@gmail.com,  ‡ heiwaeyu@gmail.com,  †† hsakai@hiroshima-u.ac.jp

Abstract  Subject relative clauses (SRCs) are processed more easily than object relative clauses (ORCs) in many languages. The present study aims to test the theories that have tried to explain why such an asymmetry exists in relative clause processing, focusing on the PDC approach (Gennari & MacDonald, 2008, 2009) that assumes that production, corpus distribution, and comprehension are closely related to each other. We conducted a gated sentence completion experiment and a self-paced reading experiment. Although the results do not show complete agreement between production and comprehension, the PDC approach captures many aspects of our results that is difficult to cover by other theories.

Keyword  Sentence processing, relative clause, self-paced reading task, gated sentence completion task

1. Introduction

Many researchers found that subject relative clauses (SRCs) are processed more easily than object relative clauses (ORCs) in many languages [1, 2]. There has been much effort made to explain this processing asymmetry [3, 4, 5, 6, 7, 8]. Some researchers regarded the filler-gap dependency [9] that is included in relative clauses as the main source of the processing asymmetry found in relative clauses. They proposed processing theories specialized to it, such as Structural Distance Hypothesis (SDH) [3] and Dependency Locality Theory (DLT) [4]. However, other researchers assumed relative clause processing as the ambiguity resolution processes and tried to explain the processing asymmetry in relative clauses by probabilistic approaches [5, 6, 7, 8].

In this paper, we will explore into this topic from the latter viewpoint, and particularly attempt to examine the Production Distribution Comprehension approach (PDC approach) [5, 6] in Japanese. We will report that the processing load between SRCs and ORCs is reversed when the animacy of subject and object NPs are controlled. These results are partly consistent with a production study and a corpus study conducted in Japanese [10], and thus argue for the PDC approach.

2. PDC approach

The PDC approach is the advanced version of a constraint satisfaction model and assumes that production principles, the distribution in corpora and the tendencies in comprehension are closely related. Gennari & MacDonald (2008, 2009) pointed out that production principles and distributional information help to disambiguate sentences during comprehension [5, 6]. For example, animacy is one of the important cues to resolve ambiguity during sentence comprehension. They compared ORCs like (1a) and (1b) with passive relative clauses like (1c) and (1d), and showed that passive relative clauses were read faster than ORCs [5].

(1a) The movie that the director watched had …
(1b) The director that the movie pleased had …
(1c) The movie that was watched by the director had …
(1d) The director that was pleased by the movie had …

Gennari & MacDonald also conducted the gated sentence completion experiment. Their results were compatible with the results of the comprehension experiment.

Although the PDC approach seems capable to explain a wide variety of experimental results, it has been tested in limited languages such as English and Chinese [11]. To test the validity of the PDC approach and take a step forward, research in wider range of languages is needed. For this purpose, Japanese is a language for a good benchmark because the structural properties in Japanese are different from those in English and Chinese. For example, Japanese relative clauses are prenominal and basically do not have overt relative clause markers. These
characteristics bring about ambiguity as to whether the current fragment like (2) belongs to the main clause, the subordinate clause, or the relative clause etc.

(2) Gaisyou-wo utagat-ta
foreign minister-ACC suspect-PAST

In this study, we employed relative clauses with inanimate object nouns and conducted a gated sentence completion experiment and self-paced reading experiment, and compared these results with the corpus study in Sato (2011) [10]. First, we explored how native speakers of Japanese complete SRC/ORC fragments, and then we compared the reading time of SRCs with animate head nouns and ORCs with inanimate head nouns.

3. Gated sentence completion experiment

To explore the tendency of relative clause production, we conducted a gated sentence completion experiment.

Method

Materials

First, we made 24 sets of sentences that contain relative clauses like (3a) and (3b) for the self-paced reading experiment. We selected transitive verbs inside relative clauses that tend to take an animate subject noun and inanimate object noun.

(3a) Kokkai-no honkaigi-de
Congress-GEN session-at
syougen-wo koban-da giin-wa ...
testimony-ACC refuse-PAST senator-TOP
"the senator that refused the testimony at the session ..."

(3b) Kokkai-no honkaigi-de
Congress-GEN session-at
giin-ga koban-da syougen-wa ...
senator-NOM refuse-PAST testimony-TOP
"the testimony that the senator refused at the session ..."

Frequency and the number of moras and characters were controlled to rule out their possible effect on reading times in Experiment 2 (all ts < 0.78). Then, the adverb phrases and the latter part of the sentences were deleted, and the relative clause part before its head noun was left.

(4a) Syougen-wo koban-da
testimony-ACC refuse-PAST

(4b) Giin-ga koban-da
senator-NOM refuse-PAST

Then, we added the animate object noun conditions like (4c) and (4d) to compare the results with the previous research [12, 13]. In these conditions, we selected transitive verbs that tend to take both animate subject and object noun.

(4c) Gaisyou-wo utagat-ta
foreign minister-ACC suspect-PAST

(4d) Giin-ga utagat-ta
senator-NOM suspect-PAST

Finally, each target item had 4 conditions. Experimental design was 2x2 (2 relative clause type and 2 animacy type). We made 4 lists along with Latin-square design and combined each of them with 48 filler sentences. As a result, each list contained 72 fragments.

Participants

30 university students participated in this experiment in exchange for a course credit.

Procedure

Participants were instructed to complete 72 fragments in natural and grammatical Japanese. They were also instructed that if they feel the fragment is natural and grammatical enough, they can finish the fragment by writing “.” (period). The experiment lasted for about 20 minutes.

Results

720 sentences were obtained and all sentences were grammatical. We classified 720 sentences as follows.

Table 1. Produced sentences in the gated sentence completion experiment

<table>
<thead>
<tr>
<th></th>
<th>Inanimate Object</th>
<th>Animate Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRC ORC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRC</td>
<td>79</td>
<td>104</td>
</tr>
<tr>
<td>ORC</td>
<td>145</td>
<td>129</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>relatives</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>Cleft</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Period</td>
<td>44</td>
<td>22</td>
</tr>
<tr>
<td>Sentence</td>
<td>47</td>
<td>8</td>
</tr>
</tbody>
</table>
If a sentence was completed by adding a subject relative clause in SRC condition like (4a) and (4c), or an object relative clauses in ORC condition like (4b) and (4d), we counted it as a “SRC/ORC”. Other kinds of relative clauses were counted as “Other relatives”. “Cleft” pertains to sentences that were completed by adding the particle “no-wa” which generally marks cleft sentences, and “Period” means sentences that were completed by adding “.” (period). “Sentence” includes sentences other than the above 4 kinds of sentences.

Generalized linear mixed model revealed a significant main effect of relative clause type (Wald’s $Z = 8.27$, $p < .01$) and a significant interaction between relative clause type and animacy type (Wald’s $Z = 3.11$, $p < .01$). No significant main effect of animacy type, however, was observed (Wald’s $Z = -0.946$, $p = .344$).

Discussion

The significant main effect of relative clause type means that more object relative clauses were produced across conditions, and the significant interaction indicates the differences in the number between SRCs and ORCs are larger in the inanimate object condition than the animate object condition. PDC approach predicts that both in the animate object condition and inanimate object condition, reading time of ORCs should be faster than SRCs. In Japanese, however, SRCs with animate object nouns are read faster than ORCs with animate object nouns like those in English [12, 13]. Though the results obtained here are not consistent with the comprehension results, these results can be understood by taking the characteristics of Japanese into consideration. In SRC conditions, participants saw an object noun and a transitive verb while in ORC conditions, a subject noun and a transitive verb. In Japanese, subject nouns tend to be omitted without surrounding context, and the omitted subject tends to be understood as the first person. However, object omissions in no surrounding context condition are a little unnatural though they are still acceptable. These differences might have an impact on the number and the type of produced sentences. Ueno & Garnsey also reported similar results

<table>
<thead>
<tr>
<th>Region</th>
<th>Inanimate Object</th>
<th>Animate Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRC</td>
<td>916</td>
<td>150</td>
</tr>
<tr>
<td>ORC</td>
<td>1478</td>
<td>120</td>
</tr>
</tbody>
</table>

Chi-square test revealed the difference between SRCs and ORCs in the animate object condition is marginally significant ($\chi^2 = 3.33$, $p = .068$), and the difference between SRCs and ORCs in the inanimate object condition is significant ($\chi^2 = 131.93$, $p < .001$). Taking these results into consideration, PDC approach predicts that ORCs should be read faster than SRCs in the inanimate object condition. Next, we tested this prediction using the self-paced reading method.

4. Self-paced reading experiment

To test the above prediction, we conducted a self-paced reading experiment.

Method

Materials

We employed materials like (3a) and (3b) used in the gated sentence completion experiment. Regions were
divided according to “bunsetsu”, which is generally the complex of a word and one or two particles, as in Table 3. We made 2 lists along with Latin-square design and combined each of them with 60 filler sentences. As a result, each list contained 84 sentences.

**Participants**

25 university students participated in this experiment. Participants received 500 yen in return for their participations. The experiment lasted about 20 minutes.

**Procedure**

The sentences were presented on a computer monitor using non-cumulative, word-by-word, self-paced reading method [14]. The experiment was run on a laptop computer using Linger 2.94 (developed by Doug Rohde). Each trial started with an image of the sentence in which dashes replaced all the printed characters. Participants pressed the spacebar to reveal each new “bunsetsu”, causing the preceding “bunsetsu” to revert to dashes. At the end of each sentence, participants answered a comprehension question by pushing ‘‘Y’’ for Yes and ‘‘N’’ for No. Participants were encouraged to read as naturally as possible and to answer the questions according to their first impulse.

**Results**

Both the comprehension accuracy rates of SRCs and ORCs with inanimate object nouns are 94.3%. The trials that participants could not answer correctly were removed from further analyses and 4528 data points remained. The data points whose reading time is less than 200ms or longer than 5000ms were also removed. Finally, we calculated 2.5SD in each region by condition, and the data points that exceeding 2.5SD were removed. The number of removed data points was 158 and it was 3.49% of 4528 data points.

From region 1 to 4, there was no significant difference between SRCs and ORCs ($F$s < 2.917). In region 5, the difference reached significance in the analyses by subjects and items ($F(1,24) = 8.228$, $p < .001$, $F(1,23) = 6.332$, $p < .05$). In region 6, though the analysis by items showed marginal significance, we could not confirm the significant difference in the analysis by participants ($F(1,24) = 1.052$, $p = .315$, $F(1,23) = 4.276$, $p = .05$). In region 7, there is no difference in reading time between SRCs and ORCs ($F(1,24) = 1.693$, $p = .206$, $F(1,23) = 0.795$, $p = .382$). In region 8, the difference was marginally significant only in the analysis by participants ($F(1,24) = 3.012$, $p = .096$, $F(1,23) = 1.323$, $p = .262$).

**Table 4. F values in the relative clause part and the split-over region**

<table>
<thead>
<tr>
<th>Region</th>
<th>By participants</th>
<th>By items</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>df</td>
<td>$F_1$</td>
</tr>
<tr>
<td>3 (noun)</td>
<td>1,24</td>
<td>2.777</td>
</tr>
<tr>
<td>4 (verb)</td>
<td>1,24</td>
<td>1.692</td>
</tr>
<tr>
<td>5 (relative head)</td>
<td>1,24</td>
<td>8.228</td>
</tr>
<tr>
<td>6 (adverb)</td>
<td>1,24</td>
<td>1.052</td>
</tr>
</tbody>
</table>

**Discussion**

Region 5, which is the critical region, showed clear main effect of relative clause type. In this region, ORCs were processed faster than SRCs. This result was along with the PDC approach because it predicts ORCs with inanimate head noun should be processed more easily than SRCs with animate head. In the next section, we will argue the validity of the PDC approach against Structural Distance Hypothesis (SDH) [3] and Dependency Locality Theory [4].

**5. General discussion**

Experiment 1 showed that participants produce more ORCs in fragments like (5b) while they produce less SRCs in fragments like (5a).

(5a) Syougen-wo koban-da testimony-ACC refuse-PAST

(5b) Giin-ga koban-da senator-NOM refuse-PAST

Experiment 2 revealed that ORCs with inanimate head
nouns are processed faster than SRCs with animate head nouns. Through our experiments, the results showed the close relation between production and comprehension. And they also matched with the distribution revealed in Sato [10]. From these results, it can be said that PDC approach seems to be supported in Japanese as well.

On the other hand, SDH [3], which is one of the major hypotheses for relative clause processing, predicts that SRCs are processed faster than ORCs because the gapped position in SRCs is structurally higher than that in ORCs. Another major hypothesis, DLT predicts that ORCs should generally be read faster than SRCs because the number of unresolved dependency is smaller in ORCs than in SRCs. These 2 hypotheses can explain one side of experimental results but cannot explain the other. However, Traxler, Morris, & Secly (2002) [15] and Mak, Vonk, & Schriefers (2002) [16] have showed that SRCs and ORCs with inanimate object nouns show no difference in processing difficulty. They claimed that animacy was one of the important semantic cues and it helped the parser to process ORCs and the processing asymmetry disappeared. The major difference between structure- or dependency-based approaches and PDC approach is how animacy affects comprehension. Researchers that emphasize the importance of structural processing assume that animacy is a semantic cue for the parser to construct a syntactic structure or dependency. However, if the structural complexity is the main source of processing cost and other semantic cues only reduce it, reversed processing asymmetry like our results should not be expected. On the other hand, the PDC approach considers animacy as one of probabilistic cues to assess how probable the expected structure is. In this case, reversed processing costs will be possible because the comprehension tendency depends on the distributional information and production principles. Considering these results, the PDC approach seems to be supported also in Japanese. However, there is a persisting problem in the use of corpus distribution because how frequency is defined varies among researchers. For example, Reali & Christiansen [15] counted the structural frequency and Gennari & MacDonald [5] calculated the transitional probability. The corpus study in Sato [10] did not calculate the probability, merely counting the number of structures. It is possible that a structure might have less structural frequency while it might have high transitional probability against another structure. Kahraman et al. [16] investigated the Japanese cleft sentences through a corpus study and self-paced reading experiment and indicated that the transitional probability is better in predicting reading time than the structural frequency. This might also be valid for relative clauses in Japanese. Therefore, we should investigate the transitional probability of relative clauses and take a step forward to define what “frequency” really means in human cognitive system.

6. Conclusion

The aim of the present study was to explore the PDC approach in Japanese. A series of experiments in Japanese indicated that the tendency in production seems to be mirrored in the tendency in comprehension. In this respect, PDC approach is the hypothesis that has a good explanatory power and there are numerous reasons why it could be more convincing than the other hypotheses. However, there is a problem in handling distributional information (structural frequency or transitional probability, or both). We should examine how to define distributional information and ascertain PDC approach through various ways such as corpus study, behavioral experiment, and computational modeling in wider range of languages.

Reference


