Binding and Dependency Length in Gapless Relative Clauses

Hajime ONO  and  Yu IKEMOTO

† Faculty of Science & Engineering, Kinki University  3-4-1 Kowakae, Higashi-Osaka-shi, Osaka, 658-8502 Japan
E-mail: onohajime@kindai.ac.jp, ikemotoyu@gmail.com

Abstract  Previous studies on Japanese relative clauses suggested that prediction makes an effect on the processing cost for the relative clause head noun. Our study used a pronoun *soko* "there" in the relative clause in order to control the influence of the relative clause prediction. The pronoun in the relative clause was bound by a universal quantifier, which appeared as a relative clause head. The relative clause head that binds the pronoun attached to the subject was read slower than that binds the pronoun attached to the object, suggesting that the previously observed preference for subject relative clause in Japanese could be largely due to the influence of prediction.

Keyword  Japanese, sentence processing, relative clause, pronoun, dependency formation

1. Introduction

Theories of sentence processing have benefited quite a lot from the cross-linguistic investigation, and the issue of dependency formation could be one of the areas that enjoyed the diversity found in human language. Probably even within a language, there are a large number of variety with respect to the dependency, and examining each one of them is highly valuable for combining the results from those obtained elsewhere. In this paper, we focus on the diversity found in relative clauses, aiming at examining to what extent seemingly different processing steps can be handled by a common procedure in the dependency formation.

2. Relative clauses

There have been a lot of studies showing that subject relative clause (SRC) is easier to process than object relative clause (ORC) in a number of different languages (King & Just, 1991; Mak, et al., 2006; Traxler, et al., 2002; Holmes & O’Regan, 1981; Kwon, et al., 2006, among others). For instance in (1), the relative clause structure in (1a) took longer to read than that in (1b).

(1a) ORC: the student [s who, the professor [VP saw _i]]
(1b) SRC: the student [s who _i [VP saw the professor]]

One of the major hypotheses that account for the different processing cost observed between (1a) and (1b) uses the linear distance between the filler and the gap as an index (Dependency Locality Theory, Gibson, 1998, 2000, often referred to in the literature as Linear Distance Hypothesis (LDH)). According to Gibson’s DLT, ORC is more costly to process than SRC because there are more words intervening between the filler and the gap. Longer dependency length is more taxing because the parser has to hold the filler in its working memory in a longer period of time.

Another prominent hypothesis that accounts for the slower reading time for ORC uses the structural complexity as an index for calculating the processing cost (referred to as Structural Distance Hypothesis, Hawkins, 1999; O’Grady, 1997). According to Structural Distance Hypothesis, ORC is harder to process than SRC because the dependency between the filler and the gap in ORC crosses more phrasal boundaries such as S and VP nodes than that in SRC does. The more phrasal boundaries the dependency crosses, the more costly the processing of relative clause will be.

Some of the processing studies on Japanese relative clauses pointed out that investigating Japanese relative clauses was particularly relevant in deciding which of the two hypotheses is better to account for the processing contrast between SRC and ORC (Ishizuka, 2005; Miyamoto & Nakamura, 2003; Ueno & Garnsey, 2008; Sato, Kahraman, Ono, & Sakai, 2010). Japanese relative clauses are different from English relative clauses in that the head noun follows the relative clause part that contains a gap as shown in (2).

(2a) ORC: [s kyouzyu-ga [VP _i mita] gakusei, ... professor-NOM saw student]

‘the student who the professor saw’

(2b) SRC: [s _i [VP kyouzyu-o mita] gakusei, ... professor-ACC saw student]

‘the student who saw the professor’

This syntactic property of Japanese relative clauses is important because the two competing hypotheses reviewed above make different predictions with respect to the processing cost of the relative clauses. According to LDH, the processing cost of ORC in Japanese should be smaller than that of SRC because there is only one word, a verb,
intervening between the gap and the filler. In SRC, there are two words, an object and a verb; hence, it is more costly to process SRC than ORC. On the other hand, SDH predicts that the processing cost of SRC is smaller than that of ORC. Assuming that the relative structural positions of the subject and the object are the same in English and Japanese, SDH makes the same prediction in Japanese as it does in English.

Using a self-paced reading method, Miyamoto & Nakamura (2003) examined the reading times for SRC and ORC in Japanese, and found that the reading time of the head noun in SRC was faster than that in ORC. Their results supported that the structural complexity that is involved in the dependency between the relative clause gap and the head noun worked well as a measure computing the processing cost of the relative clauses.

However, as Miyamoto & Nakamura (2003) as well as some others pointed out, there are a number of issues that might have affected their findings. One of those issues that are also of interest in the current study is a concern that there is a difference between the two conditions regarding the point where the parser may notice the existence of a gap. In SRC, the parser encounters an accusative NP at the beginning of the sentence, which tells the parser that the nominative NP is missing for some reason (it could be due to a topic drop or scrambling, for instance). This does not provide the parser any definite information that the parser is now dealing with a relative clause structure, but the accusative NP appearing at the beginning of the sentence may be sufficient for increasing the probability of the upcoming relative clause structure to a certain level. In ORC, on the other hand, the parser may not be able to recognize the presence of a gap until it encounters the transitive verb in the relative clause. Although it is not straightforwardly clear how this difference about the gap recognition (or maybe more properly “the recognition of the dependency”) would affect the reading times of the head noun of the relative clause, the difference nonetheless seems to be a confounding factor for the reading time results.

Furthermore, it has been suggested that English relative clause processing involves Active Gap Filling (Crain & Fodor, 1985; Stowe, 1986; Klunder & Kutas, 1993) where the parser actively tries to identify the gap position after it sees the filler (i.e., the head noun in the relative clause). However, the relative ordering between the filler and the gap is different in Japanese: The relative clause that involves a gap precedes the head of the relative clause. It is then quite unlikely that Active Gap Filling is at work in the processing of Japanese relative clauses, though it is assumed elsewhere that Active Gap Filling is quite relevant for the processing of scrambling in Japanese elsewhere (Miyamoto & Takahashi, 2002; Aoshima, Phillips, & Weinberg, 2004).

Summing up the discussion so far, we have seen that SRC was easier to process than ORC in English, and two competing hypotheses, Linear Distance Hypothesis and Structural Distance Hypothesis, have been shown to account for the processing contrast. Processing studies on Japanese relative clauses are potentially useful in order to test which hypothesis has a wider empirical coverage. It has been observed that, in Japanese, SRC was easier to process than ORC as well, suggesting that SDH is empirically superior to LDH. However, there is some concern about the results in Japanese with respect to the timing when the parser might be able to notice the presence of the relative clause structure.

3. The current study

It is quite disappointing if the results on Japanese relative clauses cannot provide firm empirical findings toward the source of the processing difficulty of relative clause, especially about the competition between LDH and SDH, due to some confounding. However, even though Japanese relative clause structure involves a dependency that is different from English, namely the one in which the gap precedes the filler, it seems quite valuable to investigate how the parser deals with such a dependency. For instance, to what extent is the processing of a gap-filler dependency similar to the processing of a filler-gap dependency? This is particularly an interesting issue if looking at languages like Japanese. Note that a lot of studies on the processing of scrambled sentences in Japanese argue that it involves processing of a filler-gap dependency. That means Japanese involves both a filler-gap dependency and a gap-filler dependency within one language. Then, whether or not those two dependencies are similar to each other in terms of processing is crucial because it has a direct influence on how many different types of processing algorithms a language has to be equipped with.

Furthermore, although it is quite unlikely that the processing of the dependency formation in English relative clauses is exactly the same as that in Japanese relative clauses, it is still conceivable that there are some processing steps that are shared between the two. In other
words, the term ‘dependency formation’ obviously involves multiple lower-level steps. For example, we can imagine that, as a step in the dependency formation in Japanese relative clauses, there is a process of linking between the head noun of the relative clause and the gap once the parser finally encounters the head noun and the relative clause gap is posited in the parsed structure. Assuming that this process of linking occurs only after both the filler and the gap are identified, it seems plausible that it occurs as a late stage even in English. As we have seen above, previous studies may have failed to fully control when the parser notices the existence of the dependency, then they may not have been able to appropriately measure the processing cost of this linking step.

In the current study, we used a Japanese pronoun soko ‘there’ in the relative clause instead of having a gap in an argument position, i.e., subject or object, and the pronoun soko forms a dependency with the head noun of the relative clause. Therefore, they are called gapless relative clauses. This type of relative clause is arguably quite useful in order to examine the processing cost of linking, which is one of the steps in the dependency formation, while controlling the timing of the recognition of the dependency by the parser.

Pronoun soko allows two types of usage: a referential reading or a bound variable reading, as shown in (3). In (3a), the pronoun is used to refer to the entity, a place Saijo, in the previous sentence. In (3b), the pronoun is bound by a universal quantifier dono-ryokan-mo ‘every inn’. To obtain a bound variable reading, the pronoun must be bound by a quantifier denoting a place or organization.

(3a) referential reading
Taro-ga Saizyo-eitta. Soko-no sake- wa oisii.
T-NOM S-to went there-GEN sake-TOP good
‘Taro went to Saijo. Sake which is found there is good.’
(3b) bound variable reading
Dono-ryokan-mo soko-no okami-ga yoku syabureru.
every-inn-also there-GEN landlady-NOM much talk
‘For every inn, its landlady talks a lot.’

It is possible to have this pronoun inside the relative clause and bound by a universal quantifier that appears as a head noun of the relative clause as in (4) (Hoji 1991, 1995; Hoji, Kinsui, Takubo & Ueyama, 2000, Ueyama 1998).

(4a) Gapless Relative Clause: Pronoun in Subject-RC
[sono-no okami-ga kyaku-o kantaisiteiru]
there-GEN landlady-NOM guest-ACC welcome
dono-ryokan-mo yoku syabu
‘Every inn where its landlady welcomes its guests cleans the bathroom thoroughly.’

(4b) Gapless Relative Clause: Pronoun in Object-RC
[okami-ga sono-no kyaku-o kantaisiteiru]
landlady-NOM there-GEN guest-ACC welcome
dono-ryokan-mo yoku syabu
‘Every inn where the landlady welcomes its guests cleans the bathroom thoroughly.’

Note that examples in (4) do not have a gap in the relative clauses; while the pronoun soko is attached to the subject in the relative clause in (4a) “Pronoun in Subject-RC”, the same pronoun is attached to the object in the relative clause in (4b) “Pronoun in Object-RC.”

Using this type of pronoun as a bound variable inside the relative clause has some advantages for our current concerns. First, although this kind of relative clauses do not have a real “gap” in a strict sense, the pronoun forms a similar dependency with the head noun of the relative clause. Then, it is conceivable to hypothesize that the way the parser handles dependency formation of this type share some features with the dependency formation between a gap and the filler. In addition, in examples (4), all arguments (subject and object) are present; unlike examples in (2), both of the relative clauses in (4) start with a nominative NP followed by an accusative NP. This suggests that it is quite unlikely that there is a difference between the conditions about the timing of the parser’s recognition of the relative clause. In other words, the parser’s reanalysis for the relative clause structure should occur at the same point in a sentence: The parser has to reanalyze the main clause structure to the relative clause structure just when it encounters the head noun. Furthermore, there has been suggested that the processing costs of the relative clause reflect the frequency of the structure (see Sato, 2011). It seems plausible to assume that the paradigm in (4) is independent from the frequency issue.

Using the two conditions, we would like to examine what determines the processing cost of establishing the dependency with the head noun of the relative clause. More specifically, in the case of dependency formation in Japanese relative clauses, where the dependent element (i.e., usually a gap) precedes the licensor (i.e., a filler), which hypothesis, SDH or LDH, makes a correct prediction about the processing cost associated with the
dependency formation. Recall that this is exactly one of the questions the previous research on Japanese relative clauses aimed to investigate.

4. Experiment

A self-paced reading experiment was conducted to examine which type of relative clause is more costly to process (Just, Carpenter & Woolley, 1982). Twenty-six sets of target sentences were prepared with 2 conditions (pronoun in subject-RC and in object-RC, similar to sentences in (4)), distributed into 2 lists. Each list was combined with 50 filler sentences. Thirty university students participated the study and read sentences presented in a moving window fashion, where the reading times for each phrase (‘bunsetu’ in Japanese) were recorded. Participants were instructed to read the sentences at the speed as they do normally. Each sentence was followed by a comprehension question, which encourages the participants to pay attention to the content of the sentence.

Data from 4 participants whose comprehension accuracy rates were below 66% were eliminated. Data from 4 target sentences were also eliminated because their comprehension accuracy rates were below 66%. Data from the target sentences where the participant answered incorrectly to their comprehension question were also eliminated from further analyses. Reading time data were trimmed so that data points beyond 2 SD from the relevant condition × region cell mean were discarded. Figure 1 illustrates the reading time data. Regions 2 and 3 showed a large difference between the two conditions, but this was due to the length of the phrase; in the pronoun in subject-RC condition, the pronoun soko showed up with the nominative NP in region 2 (soko-no wakaokami-ga ‘there-GEN young-landlady-NOM’), while in the pronoun in object-RC condition, the pronoun showed up with the accusative NP in region 3 (soko-no syukuhakukyaku-o ‘there-GEN guest-ACC’). Region 4 contains a transitive verb in the relative clause, and region 5 is the critical region that contains a head noun of the relative clause (dono ryokan-mo ‘every inn-also’). At region 5, the pronoun in object-RC condition was read reliably faster than the pronoun in subject-RC condition ($F_{1(1.25)}=6.45, p<0.02; F_{2(1.21)}=7.76, p<0.01$). The reading time data in region 4 and 6 did not show any statistically reliable difference. The overall results indicated that the processing of the gapless relative clause where the head noun forms a dependency with the pronoun attached to the object was less costly than that involves a dependency with the pronoun to the subject.

5. General discussion and conclusion

Our results showed that the gapless relative clause was easy to process where the dependency was linearly short, supporting the idea that processing cost for the dependency formation between a pronoun and the relative clause head is accounted for by LDH. This result contrasts with the previous results that SRC was easier to process than ORC, suggesting that the previous results do not directly reflect the processing cost associated with establishing the dependency between the filler and the gap that seems to occur after the parser encounters the head noun of the relative clause and identification of the gap. As was discussed in section 2, the processing cost found in previous studies may have resulted from the difference based on the predictability of the relative clause structure. In addition, there are some more recent proposals for the processing cost connected to the ORC based on interference (Gennari, et al., 2012; Gennari & MacDonald, 2008), frequency (Reali & Christiansen, 2007), discourse (Roland, et al., 2012), and so on. Any one of those may have affected the previous results. One possible reason for getting a result that contrasts with those in the previous studies could be that the use of pronoun in the relative clause makes those factors ineffective in processing the structure.

Based on the idea that the parser does not recognize the existence of the dependency until it encounters the head noun of the relative clause, we can hypothesize that the parser conducts a backward search triggered by the head noun. It has been argued that a certain dependency formation involves a backward search (Sprouse, Fukuda,
Ono & Kluender, 2011), then the current case with a bound variable use of a pronoun inside the relative clause might be another situation where a backward search is in order.

Related to the backward search in Japanese, Yasunaga (2010:88, experiment 6) conducted an experiment to measure the processing cost observed in the dependency formation between an NP and a numeral quantifier in the paradigm in (5). According to Yasunaga (2010), the examples in (5a) and (5b) involve a dependency between an NP 3-satu ‘3-classifier’ and a numeral quantifier. However, no effect showed up at 3-satu when comparing (5a) and (5b). On the other hand, a P600 effect was observed by comparing 3-satu in (5b) and kinnou in (5c).

(5a) gakusee-ga 3-satu katta.
student-NOM bought
‘The student bought three magazines at a convenient store in his neighborhood.’

(5b) gakusee-ga 3-satu katta.
student-NOM bought
‘The student bought a magazine at a convenient store in his neighborhood.’

(5c) gakusee-ga kinnou katta.
student-NOM bought
‘The student bought a magazine at a convenient store in his neighborhood yesterday.’

It is slightly disappointing for not observing the distance effect for the backward dependency in the above cases, but the results may indicate that the numeral quantifier actually forms a dependency not with an NP but with a verb (Nakanishi, 2007). It is desirable to investigate further with the paradigm used in our experiment.

There are some further concerns for our current findings. First, we assumed that there was no processing difference between the two conditions up to the point when the parser reads the verb in the relative clause in our experimental paradigm, but it should be made clear what exactly the parser does when it sees an NP with the pronoun soko with no context related to that pronoun. There are multiple possibilities; it could be that the parser takes the pronoun to be a referential pronoun, considering some appropriate context not mentioned in the experiment. Or, the parser might actively construe the pronoun to be a bound pronoun, somehow come up with a structure in which a pronoun is bound by a quantificational expression, such as a quantifier or a wh-phrase. Although it seems to us quite unlikely that the parser actively set up a structure and search for the antecedent (i.e., the filler), it is necessary to figure out whether this assumption is well defended. Related to the construal of the pronoun, we must make sure that the parser establishes the appropriate dependency when it hits the head noun of the relative clause. As mentioned above, the pronoun allows the referential use, and it is possible for the parser to construal the pronoun as referential, not bound, even if there is a quantifier that is a potential licensor for the bound variable usage for the pronoun. A follow-up experiment, possibly an offline experiment, is needed to explore the interpretation of the pronoun: a similar concern was investigated in Aoshima, et al. (2009).

Finally, there is an ambiguity (a left-edge ambiguity) regarding the relative clause structure the parser can build. We have assumed so far that, when the parser read the head noun of the relative clause, it reanalyzes the structure so that both nominative and accusative NPs are inside the relative clause. However, it might be possible for the parser to build a structure where only the accusative NP is in the relative clause but the nominative NP is in the main clause. It could be problematic if the parser has some bias for this ambiguity in either condition, it might affect the reading time.

To sum up, the previous studies on Japanese relative clause have some concerns regarding the predictability of the gap and the relative clause structure itself due to the case-marking pattern. Such factor may have masked some effects observed at the head noun of the relative clause, which is a critical region. We tried to overcome the problem by using a pronoun inside the relative clause, and examine the processing cost of dependency formation between the gap and the filler, which has been the main goal as in the previous studies. We found that the condition where a pronoun was attached to the object was read faster, somewhat in contrast to the results in the literature. We suggested that the repeatedly found reading time pattern that SRC is easier to process than ORC arises due to the influence such as the predictability of the gap before the parser sees the head noun of the relative clause.

References

References


