Crossover Effects with Set Indices: Evidence from Japanese Scrambling[*]

YUSUKE YAGI
University of Connecticut

YUTA TATSUMI
Meikai University

In Minimalist Program after Chomsky (1995), researchers have attempted to abandon *indices* from the syntax system. However, this paper points out that indices are still helpful (and even necessary) notions in the theory by examining crossover effects in Japanese scrambling. Furthermore, it will be argued that two different definitions of *disjointness* are needed for Binding Theory.

1 Introduction: Set Indices

Sportiche (1985) observes ill-formedness of the question in (1), which is intended to ask *for which x, Bill thinks Bill and x saw x*.

(1) Strong Crossover with Split Antecedents: Sportiche (1985: 466)
Which man$_{\{1\}}$ does Bill$_{\{2\}}$ think they$_{\{1,2\}}$ saw t$_{\{1\}}$?

The example is similar to the Strong Crossover (SCO) configuration in (2), ruled out by Binding Condition C. A trace left by wh-movement, which is argued to be an R-expression, cannot be *bound* by a coindexed element.

(2) *Who_1 *does* he_1 *love* t_1?

However, attributing (1) to SCO calls for revision of the definition of *binding*. Under the standard definition of *binding* in (3), the trace in (1) is only bound by *which man*, and crucially not bound by *they*. The trace and *they* do not have an *identical* index.

(3)   *x binds* by *y* iff
      a. *x* c-commands *y*, and
      b. *x* and *y* have the identical referential index.

Sportiche (1985) thus suggests that a syntactic item should have a *set* of indices and that we should revise the definitions of *binding* and *disjoint* as (4).

(4)   a. *x binds y* iff *x* c-commands *y*, and *x* and *y* have identical *index sets*.
      b. *x* is *disjoint* from *y* iff the intersection of the index sets of *x* and of *y* is the empty set.

An R-expressions must be *disjoint* from any other elements that c-command it from an A-position. The trace in (1) is not *disjoint* from *they* under the new definition because the intersection of the index set of *they* and the one of the trace is non-empty. In this way, ill-formedness of (1) is attributed to SCO, as expected.

In this paper, we point out that the set indices are also called for to account for a crossover effect in A-movement, namely in Japanese scrambling. It suggests the necessity of set indices in the syntax theory. The rest of this paper is organized as follows. Section 2 lays out the relevant data, and Section 3 suggests an analysis. Further empirical consequences are discussed in Section 4. Section 5 discusses a precise definition of *disjointness*. It is pointed out that we need two definitions of disjointness, *weak disjointness* and *strong disjointness*, to cover the data. Section 6 concludes.

2   Crossover and Set Indices in A-movement

It has been argued that short Scrambling (SS; scrambling within a local clause) is A-movement in Japanese (Saito 1992). As one of the A-movement characteristics, SS creates a new binding configuration by movement. This is shown in (5). In (5a), coreference of *Taro and Hanako* and reciprocal *otagai* 'each other' is degraded since the latter is not c-commanded by the former. SS makes it perfect.

(5) a. ??[*otagai*$_{\{1,2\}}$*-no*   *sensei*]*-ga*   [*Taro*$_{\{1\}}$ *to*   *Hanako*$_{\{2\}}$]*-o*
       each.other-GEN teacher-NOM   Taro   and Hanako-ACC

   hihansita.
   criticized

   Lit. 'Each other's teacher criticized [Taro and Hanako].'

   b. [*Taro*$_{\{1\}}$ *to*   *Hanako*$_{\{2\}}$]*-o* [*otagai*$_{\{1,2\}}$*-no*   *sensei*]*-ga*
       Taro   and Hanako-ACC   each.other-GEN teacher-NOM

   t$_{Obj}$ *hihansita.*
        criticized

   Lit. '[Taro and Hanako], each other's teacher criticized t$_{Obj}$.'

Now consider (6), our main concern in this paper. In (6), the embedded object *Ziro* undergoes SS. Since *Ziro* and the matrix subject *Taro* c-command the plural pronoun *karera* 'they,' in principle, they can be antecedents of the pronoun. However, the pronoun cannot take *Taro* and *Ziro* as its antecedents.

(6) a. * *Taro*$_{\{1\}}$*-wa* [**karera**$_{\{1,2\}}$*-ga* *Ziro*$_{\{2\}}$*-o* *suisen-sita*   *to*]
       Taro-TOP   they-NOM   Ziro-ACC recommend-did C

   itta.
   said

   Lit. 'Taro said that they recommended Ziro.'

   b. * *Taro*$_{\{1\}}$*-wa* [*Ziro*$_{\{2\}}$*-o* **karera**$_{\{1,2\}}$*-ga* t$_{\{2\}}$ *suisen-sita*
       Taro-TOP   Ziro-ACC they-NOM   recommend-did

   to] *itta.*
   C   said

   Lit. 'Taro said that Ziro, they recommended *t*.'

The unacceptability of (6b) is not due to Condition B violation caused by *Ziro* c-commanding *karera* in the local domain. In Japanese, one of the antecedents of a plural pronoun can be in the local domain of the pronoun, as in (7).

(7) *Taro*$_{\{1\}}$*-wa* [*Ziro*$_{\{2\}}$*-ga* **karera**$_{\{1,2\}}$*-o* *suisen-sita*   *to*] *itta.*
    Taro-TOP   Ziro-NOM theyACC   recommend-did C   said

   Lit. '*Taro said that Ziro recommended them.'

Nor is it due to SCO, although the configuration of (6b) is similar to (1). A trace of A-movement is not an R-expression and it is not subject to Binding Condition C.

There are some caveats in order before moving on to our analysis of the data in (6b). First, we must show that the relevant movement operation in (6b) should be analyzed as short-distance scrambling but not as long-distance scrambling to a middle field in the matrix clause. Schematic representations are given in (8).

(8) a. Short Scrambling
       Taro ... [$_{CP}$ Ziro-ACC they t_{Ziro} recommended C] ... said.

     b. Long Scrambling
       *Taro ... Ziro-ACC [$_{CP}$ they t recommended C] ... said.

There are two pieces of evidence that the embedded object phrase moves within the embedded clause. Firstly, it has been observed that Japanese long-distance scrambling exhibits the so-called "radical reconstruction" effect (Saito 1992). (Recall that long-distance scrambling is a scrambling operation across a clausal boundary.) If the object noun in the embedded clause had moved into the matrix clause as in (8b), it would be expected that the scrambled object phrase undergoes reconstruction into the original position, yielding a Condition C violation. To see the nature of the movement operation in (6b), it will be helpful to consider the examples in (9).

(9) a. *$Taro_{\{1\}}$-wa sennsei-ni [mada $karera_{\{1,2\}}$-ga $Ziro_{\{2\}}$-ni
         Taro-TOP   teacher-to yet   they-NOM   Ziro-to
         toohyoo-sitei-nai to] itta.
         vote-do-NEG   C   said
         Lit. 'Taro told a teacher they have not voted for Ziro yet.'

     b. *$Taro_{\{1\}}$-wa sennsei-ni [mada $Ziro_{\{2\}}$-ni $karera_{\{1,2\}}$-ga
         Taro-TOP   teacher-to yet   Ziro-to   they-NOM
         toohyoo-sitei-nai to] itta.
         vote-do-NEG   C   said
         Lit. 'Taro told a teacher that Ziro, they have not voted for yet.'

In (9), the embedded sentence contains sentential negation and the adverb mada 'yet/still'. The adverb mada in (9) behaves like an NPI as in (10).

(10) a. *mada karera-ga Ziro-ni toohyoo-sita.
         yet   they-NOM Ziro-to vote-did
         Lit. 'They have voted for Ziro yet.'

      b. mada karera-ga Ziro-ni toohyoo-sitei-nai.
         yet   they-NOM Ziro-to vote-do-NEG
         Lit. 'They have not voted for Ziro yet.'

The unacceptability of (10a) tells us that mada in (9) cannot be licensed by the matrix affirmative predicate. We take this as evidence that mada has to remain within the embedded clause in (9), and hence it indicates a clause boundary. (Notice also that Japanese phrasal adverbs generally resist long-distance scrambling operation (Miyara 1982; Saito 1985; Bošković and Taka-

hashi 1998.)) The scrambled phrase follows *mada* in (9b). This means that it undergoes clause-internal/short-distance scrambling, which generally does not exhibit the reconstruction effect.

In this connection, it should also be noted that Japanese short-distance scrambling can eliminate the violation of Condition C. Relevant examples are given in (11).

(11) a. *kare{1}-*wa* [NP *Aiko-ga*   *Ziro*{1}-*ni* *syookai-sita*
         he-TOP   Aiko-NOM Ziro-to   introduce-did

         otoko]-*o* *suisen-sita.*
         man-ACC recommend-did

         Lit. 'He$_1$ recommended the man that Aiko introduced to Ziro$_1$.'

    b. [NP *Aiko-ga*   *Ziro*{1}-*ni syookai-sita* *otoko*]-*o* *kare*{1}-*wa*
         Aiko-NOM Ziro-to   introduce-did man-ACC he-TOP

         t$_{NP}$ *suisen-sita.*
         recommend-did

         Lit. 'The man that Aiko introduced to Ziro$_1$, he$_1$ recommended.'

Since the object noun phrase in the embedded clause in (6b) undergoes short-distance scrambling as discussed above, it should not cause a Condition C violation, similarly to (11b).

Given these considerations, we assume that the unacceptability of (9b) and the same effect observed in (6b) should be analyzed as crossover effects. In the next section, we propose that the unacceptability of (6b) can be accounted for by amending Rizzi's definition of Chain Condition.

3 Proposal

To explain the unacceptability of (6b), we propose to modify Rizzi's (1986) *Chain Condition*. Rizzi's definition of Chain Condition is given in (12).

(12) Chain Condition: Rizzi (1986)
$(x_1, ..., x_n)$ is a chain iff

    a.   for all i ($1 \leq i < n$), x_i is the local binder of x_{i+1}.
    b. x is a binder of y iff x c-commands y and x and y are coindexed.
    c. x is a local binder of y iff x is a binder of y and there is no intervener z such that
       (i) z is a binder of y, and
       (ii) z is not a binder of x.

This condition is motivated by Italian data in (13). There, one of the objects of the verb undergoes passivization, and the other undergoes cliticization. Though each of these operations alone does not cause ungrammaticality, the combination does. This is because, according to Rizzi (1986), the clitic *se*

intervenes in Chain formation of *Gianni* and the trace t_1. *si* is a local binder of t_1, and the Chain $\langle Gianni, t_1 \rangle$ is ill-formed.

(13) * *Gianni*$_{\{1\}}$ *si*$_{\{1\}}$ *è stato affidato* $t_{\{1\}}$
      Gianni   self is been entrusted

      Intended: 'Gianni has been entrusted to himself.'

Since Rizzi (1986) does not discuss cases with a plural pronoun, *being a binder* is defined in pre-Sportiche's way. In order to extend Chain Condition to account for (6b), we propose to revise the condition as (14). Notice that we adopt Sportiche's definition of *bind* and *disjoint* in (4).

(14)   Chain Condition (Modified Version)
        $(x_1, ..., x_n)$ is a chain iff

        a.   for all i ($1 \leq i < n$), x_i is the local binder of x_{i+1}.
        b. x is a binder of y iff x c-commands y and x and y have the identical index set.
        c. x is a local binder of y iff x is a binder of y and there is no intervener z such that
           (i) z c-commands y,
           (ii) z doe snot c-command x, and
           (iii) z is not disjoint from y.

(15) is a schematic representation of (6b). The trace t_2 needs to form a Chain with *Ziro*, and *Ziro* must be a local binder of t_2 for that. However, this condition is not met there because *they* intervenes between them: *they* is not disjoint from t_2, so *Ziro* is not a local binder of t_2. The sentence is ill-formed because of the failure to form a proper Chain. Notice also that the new definition still accounts for the unacceptability of Rizzi's Italian example.

(15)   Taro$_{\{1\}}$ ... [Ziro$_{\{2\}}$ ... they$_{\{1,2\}}$ ... $t_{\{2\}}$]

The current analysis crucially utilizes indices and Chain, both revoked in recent Minimalist work. We argue that these two notions are still called for in the theory. In order to see this, let us consider a case where $\alpha_{[-F]}$, which has feature F to be checked, 'moves' to a higher position by leaving a copy (α_i). Suppose further that F is checked and deleted in a higher position. Schematically:

(16)   ... $\alpha_{i, \,[-F]}$ ... $\alpha_{j, \,[-F]}$

Now, how can we make sure that F in the lower instance of α, α_j, is also checked? That α_i is a copy of α_j does not help here. By analogy: suppose that you made a file on your laptop and made a copy in a different position. Even after you edit the copied file, the original file does not get any change. Thus, to ensure that feature F is checked in the higher *and* the lower instance of α,

the grammar has to know that these two instances share the same properties, e.g., whether or not their features are checked. Chain is a good candidate to make this connection, and indices are valuable for defining Chain. We also believe that a similar theoretical apparatus has been pursued even in a more recent framework of Minimalism (e.g. Form Copy in Chomsky (2021)).

4 Consequences

The present proposal has further consequences. Under the definition in (14c), z is an intervener for x being a local binder of y if and only if z c-commands y. The proposed analysis predicts that the resulting sentence should be grammatical if we embed the intervening plural pronoun in (6b) to another nominal phrase. This prediction is borne out. (17) and (18) are grammatical.

(17) *Taro*$_{\{1\}}$-*wa* [[*karera*$_{\{1,2\}}$-*no zyoosi*]-*ga Ziro*$_{\{2\}}$-*o suisen-sita*
Taro-TOP   they-GEN   boss-NOM Ziro-ACC recommend-did
to] *itta.*
C   said
Lit. 'Taro said that their boss recommended Ziro.'

(18) *Taro*$_{\{1\}}$-*wa* [*Ziro*$_{\{2\}}$-*o* [*karera*$_{\{1,2\}}$-*no zyoosi*]-*ga* t$_{\{2\}}$
Taro-TOP   Ziro-ACC   they-GEN   boss-NOM
suisen-sita   *to*] *itta.*
recommend-did C   said
Lit. 'Taro said that Ziro, their boss recommended t.'

Furthermore, the proposal can be extended to a novel observation. Consider the contrast below. (19a) shows that coreference between *Taro* and *kare* 'he' does not induce a Condition B violation, as long as the latter is embedded in coordination. However, (19b) is ungrammatical with coreference between *Taro* and *kare*. The configuration there is similar to the cases we have discussed so far.

(19) a. *Taro*$_{\{1\}}$-*ga* [*kare*$_{\{1\}}$ *to Hanako*$_{\{2\}}$]$_{\{1,2\}}$-*o suisen-sita.*
Taro[1]-NOM he   and Hanako[2]]-ACC recommend-did
Lit. 'Taro[1] recommended [him[1] and Hanako].'

b. #*Taro*$_{\{1\}}$-*o* [*kare*$_{\{1\}}$ *to Hanako*$_{\{2\}}$]$_{\{1,2\}}$-*ga* t$_{\{1\}}$
Taro-ACC   he   and Hanako-NOM
suisen-sita.
recommend-did
Lit. 'Taro, [he and Hanako] recommended t.'

The paradigm is explained away with the assumption that indices of coordinated DPs percolate. As a result, the subject in (20b) has as its index set the

union of index sets of its conjuncts, namely $\{1,2\}$. As a result, the subject is counted as an intervener for Chain formation, hence ungrammatical.

It also explains the contrast between (5b), repeated in (20a), and (20b) (Koizumi 1995; Miyagawa 1996). The latter is ungrammatical because the reciprocal pronoun *otagai* 'each other' intervenes Chain formation of *Taro to Ziro* 'Taro and Ziro' and its trace.

(20) a. [*Taro*$_{\{1\}}$ *to* *Hanako*$_{\{2\}}$]$_{\{1,2\}}$-*o* [*otagai*$_{\{1,2\}}$-*no*
      Taro   and   Hanako-ACC   each.other-GEN

      sensei]-*ga*   t$_{\{1,2\}}$ *hihansita.*
      teacher-NOM   criticized

      Lit. 'Each other's teacher criticized Taro and Hanako.'

    b. *[*Taro*$_{\{1\}}$ *to* *Ziro*$_{\{2\}}$]$_{\{1,2\}}$-*o otagai*$_{\{1,2\}}$-*ga*   t$_{\{1,2\}}$ *hinansita.*
       Taro   and Ziro-ACC   each.other-NOM   criticized.

      Lit. '[Taro and Ziro], each other criticized *t*.'

Summarizing this section, we proposed the revised Chain Condition to account for (6b). The success of the analysis implies the necessity of indices and Chain in the syntactic theory.

5 Two Definitions of Disjointness

In this section, we discuss a further theoretical issue raised by the binding of plural pronouns. Recall the definition of *disjointness* in Sportiche (1985), repeated below as (21). This definition should also be adapted to the definition of Binding Condition B.

(21)   *x* is *disjoint* from *y* iff the intersection of the index set *x* and *y* is the empty set.

(22)   Binding Condition B (First Version)
     Pronouns must be *disjoint* from any C-commanding elements in its local domain.

(22) seems to make a correct prediction for English—coreference in (23) is ill-formed.

(23)   *John$_1$ said Mary$_2$ criticized them$_{1,2}$.

However, the definition makes a wrong prediction for Japanese. (7), repeated below as (24), shows that having one of the two antecedents in the local domain of a plural pronoun does not result in ungrammaticality.

(24)   *Taro*$_{\{1\}}$-*wa* [*Ziro*$_{\{2\}}$-*ga* **karera**$_{\{1,\,2\}}$-*o suisen-sita*   *to*] *itta.*
     Taro-TOP   Ziro-NOM   they-ACC   recommend-did   C   said

     Lit. '*Taro said that Ziro recommended them.'

In order to account for these differences, we propose two definitions of disjointness – *strong disjointness* and *weak disjointness*.

(25) a. α_A is *weakly disjoint* from β_B iff $A \neq B$.

    b. α_A is *strongly disjoint* from β_B iff $A \cap B = \emptyset$.

As seen in Section 2, Chain Condition adopts strong disjointness to define a local binder. Strong disjointness should also be adopted for Condition B in English. On the other hand, Condition B in Japanese should adopt weak disjointness. It means that Condition B is subject to cross-linguistic variation in a way that has never been discussed.

(26) Binding Condition B (Final Version)

    a. English
Pronouns must be *strongly disjoint* from any C-commanding elements in its local domain.

    b. Japanese
Pronouns must be *weakly disjoint* from any C-commanding elements in its local domain.

Notice further that both strong and weak disjointness are necessary even in one language. In Japanese, for example, Condition B adopts weak disjointness, and Chain Condition adopts strong disjointness. Further variation within Japanese can be observed by considering Condition C. (27) suggests Condition C adopts strong disjointness in Japanese.

(27) *$Karera_{\{1,2\}}$-ga $Taro_{\{1\}}$-o hihansita.
    They-NOM   Taro-ACC criticized

    'They criticized Taro.'

It suggests that even within one language, different grammatical principles may adopt different definitions of disjointness. With this discussion, we argue that a thorough reinvestigation of existing grammatical principles is called for, especially for principles that are/use to be defined with indices. By considering plural pronouns, we observe a new inter- or intra-linguistic variation.

6 Conclusion

In this study, we made one empirical and two theoretical claims. Empirically, we claimed that the novel data in (6b) should be accounted for with the revised Chain Condition in (14). Theoretically, we claimed that (i) indices and Chain are still necessary theoretical devices in the syntactic theory, and (ii) plural pronouns reveal a two-way definition of disjointness, suggesting the necessity of thorough reinvestigation of grammatical principles with plural pronouns.

References

Bošković, Ž. and D. Takahashi 1998. Scrambling and Last Resort. *Linguistic Inquiry* 29(3):347–366.

Chomsky, N. 1995. *The Minimalist Program*. MIT press, Cambridge, MA.

Chomsky, N. 2021. Minimalim: Where Are We Now, and Where Can We Hope to Go. *Gengo Kenkyu* 160:1–41.

Koizumi, M. 1995. Phrase Structure in Minimalist Syntax. Doctoral dissertation, MIT.

Miyagawa, S. 1996. Word Order Restrictions and Nonconfigurationality. In *Proceedings of Formal Approaches to Japanese Linguistics 2*, 117–141.

Miyara, S. 1982. Reordering in Japanese. *Linguistic Analysis* 9:307–340.

Rizzi, L. 1986. On Chain Formation. In Borer, H., ed., *The syntas of pronominal clitics: Syntax and semantics 19*, 65–95. Academic Press, New York.

Saito, M. 1985. Some Asymmetries in Japanese and Their Theoretical Implications. Doctoral dissertation, MIT.

Saito, M. 1992. Long Distance Scrambling in Japanese. *Journal of East Asian Linguistics* 1:69–118.

Sportiche, D. 1985. Remarks on Crossover. *Linguistic Inquiry* 16:460–469.