Dissociating the processing of empty categories in raising and control sentences: a self-paced reading study in Japanese

Koki Yamaguchi[1] and Shinri Ohta[2]*

[1]Department of Linguistics, Graduate School of Humanities, Kyushu University, Fukuoka, Japan,
[2]Department of Linguistics, Faculty of Humanities, Kyushu University, Fukuoka, Japan

Introduction: Theoretical linguistics has proposed different types of empty categories (ECs), i.e., unpronounced words with syntactic characteristics. ECs are a key to elucidating the computational system of syntax, algorithms of language processing, and their neural implementation. Here we examined the distinction between raising and control sentences in Japanese and whether ECs are psychologically real.

Methods: We recruited 254 native speakers of Japanese in the present internet-based experiment. We used a self-paced reading and a probe recognition priming technique. To investigate whether raising and control sentences have different ECs (i.e., Copy and PRO) and whether these ECs cause a reactivation effect, behavioral data were analyzed using linear mixed-effects models.

Results: We found two striking results. First, we demonstrate that the reading times of raising and control sentences in Japanese were better explained by the linear mixed-effects model considering the differences of ECs, i.e., Copy and PRO. Secondly, we found a significant reactivation effect for raising and control sentences, which have ECs, and reflexive sentences without ECs. These results indicate that ECs are processed similarly to reflexive pronouns (e.g., *himself*).

Discussion: Based on these results, we conclude that raising and control sentences in Japanese have different ECs, i.e., Copy and PRO, and that ECs have psychological reality. Our results demonstrate that behavioral experiment based on theoretical linguistics, which is the first step for developing linking hypotheses connecting theoretical linguistics and experimental neuroscience, is indeed necessary for testing hypotheses proposed in theoretical linguistics.

KEYWORDS

syntax, Japanese, self-paced reading (SPR), experimental linguistics, empty category, neurobiology of language, sentence comprehension, reading

Introduction

Theoretical linguistics has proposed different types of empty categories (ECs), i.e., unpronounced words, such as NP-trace (or Copy of a noun phrase) and PRO, each of which has different syntactic characteristics. The reason for incessant attention to ECs in generative syntax (Chomsky, 2021) is that ECs considerably reflect the basic mechanisms underlying linguistic computation. Compared to full nominal expressions such as *John*, ECs by themselves provide considerably few clues about their interpretation. Interpretation of ECs is only derived from syntactic, semantic, or pragmatic operations. Thus, the study of their interpretation provides us with a probe into the computational system of natural language (syntax) and ways to investigate the interfaces between syntax and other language systems (phonology, semantics, and pragmatics). Despite their importance in theoretical linguistics, behavioral and neural mechanisms of ECs are barely discussed in recent experimental linguistics (but see Makuuchi et al., 2013; Ohta et al., 2017; Tanaka et al., 2017 for notable

exceptions). Behavioral and neural mechanisms of ECs are also critical for building linking hypotheses connecting theoretical linguistics and experimental neuroscience. Therefore, it is necessary to examine the property of ECs in detail, using experimental methods [e.g., self-paced reading (SPR)] and statistical analyses [e.g., linear mixed-effects (LME) models].

In the present study, as a first step to examine the behavioral and neural basis of ECs, we conducted an SPR experiment to elucidate the algorithms of language processing related to ECs. Behavioral experiments are crucial for discovering underlying algorithms of language processing, while neurophysiological experiments provide insights into the implementation of the algorithms in the brain (Marr, 1982; Krakauer et al., 2017). We especially focus on ECs of raising and control sentences in Japanese.

Raising and control sentences have been central concerns in theoretical linguistics, especially in generative syntax. A primary motivation for the attention on raising and control sentences is the similarity of the constructions in English, as shown in (1) and (2), which illustrate raising and control sentences, respectively.

(1) Barnett seemed to understand the formula. (Davies and Dubinsky, 2004).
(2) Barnett tried to understand the formula. (Davies and Dubinsky, 2004).

Both sentences have an intransitive matrix clause with an infinitival (non-finite) complement, NP-V-to-VP. The only surface difference is the matrix verb, *seem* vs. *try*. However, it has been pointed out that there are empirical differences between the two constructions since the early generative grammar (Chomsky, 1965, p. 22–24; Rosenbaum, 1967; see Landau, 2013 for review). In a raising sentence (1), the NP *Barnett*, which receives an agentive semantic role (θ-role) from an embedded non-finite predicate "*to understand the formula*", appears in a subject position of the matrix clause, suggesting that the NP has *raised* from the embedded clause to the matrix clause. Importantly, the original and ultimate positions are associated with a *single* argument. By contrast, in a control sentence (2), the NP *Barnett* appears to be associated with two θ-roles from both the matrix verb *tried* and embedded verb *understand*. While its syntactic position corresponds to the matrix θ-role, the interpretation of the sentence indicates that there is an additional argument in the embedded clause, which is coreferential with (or *controlled* by) the NP *Barnett*. Importantly, the two positions are associated with two arguments, which is similar to anaphora expressions (e.g., *myself*).

Raising and control predicates in this study are defined as follows. While both raising and control predicates take an infinitival complement clause, in which subject position is unpronounced, raising predicates assign a θ-role to an internal argument, whereas control predicates assign two θ-roles to an internal and external argument. In other words, the matrix subject of the raising sentence above (e.g., *Barnett*) is an Agent of the embedded verb *understand*, i.e., the person who understands the formula, but not an Agent of the matrix verb *seemed*. In contrast, the matrix subject of the control sentence is an Agent of both the matrix verb *tried*, i.e., the person who tried to do something, and the embedded verb *understand*. The number of θ-roles causes various different syntactic properties between raising and control sentences.

We will explain two representative differences between raising and control, which are observed both in English and Japanese control sentences. First, a feature distinguishing raising and control constructions is selectional restrictions to subject.

(3) The rock seems to be granite.
(4) # The rock tried to be granite.

(3) is a well-formed sentence, while (4) is semantically anomalous. The oddness in (4) results from the semantic requirements of *try*. The control verb *try* requires an agent subject, which needs an entity capable of volition. As the *rock* has no volition, (4) is a semantically odd sentence. Contrarily, the subject verb *seem* in (3) has no semantic sectional restrictions to subject. Thus, raising verbs allow non-animate subjects like *rock*.

Typical conditions in which Japanese raising and control structures appear are syntactic compound verbs. Similar to English, the surface string similarity and the functional differences between Japanese raising and control sentences have been widely reported (Kageyama, 1993; Koizumi, 1999, among others). For example, selectional restriction to subject was also reported in Japanese, as shown in (5) and (6).

(5) Ame-ga furi-*sugi*-ta.
Rain-NOM rain-too much-PST.
"It rained too much."

(6) * Ame-ga furi-*sokone*-ta.
Rain-NOM rain-fail-PST.
"It failed to rain."

The raising verb *sugi-ru* "do too much" allows non-animate subjects, such as *rain*, while the control verb *sokone-ru* "fail" requires the agent (or animate) subject.

Second, another difference is the behavior of idiomatic expressions.

(7) The cat is out of the bag. (Davies and Dubinsky, 2004).

The sentence in (7) has two meanings. One is a situation in which a particular cat is not in a particular container, and the other is that one-time secret is no longer a secret.

(8) The cat seemed to be out of the bag. (Davies and Dubinsky, 2004).
(9) ?The cat tried to be out of the bag. (Davies and Dubinsky, 2004).

With a raising predicate in (8), the expression can retain an idiomatic interpretation. However, with a control predicate in (9), the idiomatic interpretation is no longer possible.

The difference of idiomatic expressions is also reported in Japanese.

(10) Kankodori-ga nak-u.
Cuckoo-NOM sing.

(10) is a Japanese idiom, which has two interpretations. One describes a situation where a cuckoo sings, and the other

is that a store has hardly any customers. The latter is an idiomatic meaning.

(11) Kono mise-de-wa kankodori-ga naki-*kake*-ta.
      This store-DAT-TOP cuckoo-NOM sing-almost-PST.
      "A cuckoo almost sang in this store./This store almost closed down."

(12) Kono mise-de-wa kankodori-ga naki-*wasure*-ta.
      This store-DAT-TOP cuckoo-NOM sing-forget-PST.
      "A cuckoo forgot to sing in this store.

The raising verb *kake-ru* "almost" retains an idiomatic interpretation, while the control verb *wasure-ru* "forget" has no idiomatic interpretation. In addition, there are other empirical differences like θ-roles, passivization, scope ambiguity, and an expletive subject (Davies and Dubinsky, 2004; Landau, 2013).

From Chomsky (1973), referred to as "Extended Standard Theory", to Chomsky (1981), referred to as "Government and Binding Theory", it was established that raising and control sentences have different ECs, Copy (NP-Trace) and PRO as shown in (13) and (14). Chomsky (1981) defined that PRO has anaphoric and pronominal features, while Copy has only anaphoric feature.

(13) John$_i$ seems [Copy$_i$ to be a nice fellow.] (Chomsky, 1973, partially modified).
(14) John$_i$ expected [PRO$_i$ to win.] (Chomsky, 1973).

The syntactic properties of PRO have aroused intense debate within theoretical linguistics, because no full nominal expressions show anaphoric and pronominal features at the same time.

After the Minimalist Program Chomsky (1993), Hornstein (1999) proposed a new analysis to control the phenomena, which is called Movement Theory of Control (MTC). MTC primarily claims that obligatory control[1] is derived via *A-movement*. Thus, MTC considers that the null hypothesis for the derivation of raising and control sentences should resort to the same empty category, Copy. Hornstein's proposal has received crucial criticism (Culicover and Jackendoff, 2001; Landau, 2003; Bobaljik and Landau, 2009; Ndayiragije, 2012; Wood, 2012) and has provoked a great deal of controversy. Over the last few decades, the property of control sentences has been discussed and not settled yet in theoretical linguistics. To reveal whether control structures have PRO or

Copy is important for advancing theoretical linguistics. If control had Copy, it would contribute to removing construction-specific category[2] and construction-specific module. Contrarily, if control had PRO, it would be useful for linguists to consider why PRO is construction specific. Thus, experimental linguistics is necessary to settle this issue.

ECs of control phenomena were widely studied using experimental approaches; however, these studies paid attention to the difference in behavior between *subject control* and *object control* (Sakamoto, 1996, 2001; Witzel and Witzel, 2011) and the process of *agreement* between PRO and its antecedent (Demestre et al., 1999; Betancort et al., 2004; Demestre and García-Albea, 2007). There are only few studies on the comparison between PRO and Copy.

Bever and McElree (1988), McElree and Bever (1989), Featherston et al. (2000), and Featherston (2001) studied the syntactic characteristics between PRO and Copy on sentence processing. Bever and McElree (1988) and McElree and Bever (1989) investigated various kinds of ECs in English. They used a *probe word recognition priming technique*, where a sentence is presented on a screen phrase-by-phrase. At the end of the sentence, a probe word appears on the screen. The participant must decide whether or not the probe word was contained in the presented sentence. They reported that sentences with ECs and pronouns showed significantly faster response times than sentences without ECs, which is called a reactivation effect. They also found that sentences that include Copy evoke significantly faster response times than control sentences that include PRO. Consequently, they claim that PRO should be distinguished from Copy. Featherston et al. (2000) employed ERP to examine the characteristics of ECs in German. It was reported that the comparison between raising and control conditions showed a significantly positive-going ERP for the former in the 600–1,000 ms time windows.

Featherston (2001) conducted a replication study of Bever and McElree (1988) using the SPR paradigm and probe word recognition task, in which English words were replaced by German words and the same conditions were used. However, the results were contrary to Bever and McElree (1988). There was no reactivation effect by ECs and no significant difference between PRO and Copy on the response times for recognizing the probe word.

Whether ECs are psychologically real or not is another controversial topic in experimental linguistics (Pickering and Barry, 1991; Gibson and Hickok, 1993). Bever and McElree (1988) and Miyamoto and Takahashi (2002) showed the evidence for the psychological reality of ECs by presenting the reactivation effect. However, Nakayama (1995) did not find a reactivation effect of ECs in Japanese. Nakano et al. (2002) also did not observe the reactivation effect of ECs in long-distance Japanese scrambling in a low reading span group. As mentioned above, Featherston (2001) also did not find the reactivation effect in German. Therefore, whether or not ECs show the reactivation effect remained unclear.

Although Bever and McElree (1988) and Featherston et al. (2000) contributed to revealing syntactic characteristics of ECs on human sentence processing, their studies are still unsatisfactory. The problem with previous studies is that they do not consider

---

1   Another angle to look at control phenomena was proposed by Williams (1980). He divided control into two categories: Obligatory control (OC) and non-obligatory control (NOC). The examples in (1) and (2) illustrate how these two categories differ.

OC

  *It was expected PRO to shave himself.

NOC

  It was believed that PRO shaving was important. (Hornstein, 2003).

For example, OC PRO needs an antecedent while NOC PRO does not.

---

2   It has been reported that PRO only appears in control constructions.

other factors that may influence the behavioral data and ERPs. As shown in Table 1, control and raising sentences have other differences along with the difference between PRO and Copy. There is a difference between raising and control structures in terms of the number of θ-roles. Second, Pesetsky (1991) reported another difference between raising and control structures regarding the types of θ-roles. Therefore, it is too early to conclude that the observed differences in previous studies were really derived from the difference of ECs. Furthermore, other general factors may influence outputs, as shown below.

> **Factors that affect behavioral data**: Word frequency, Number of characters, Number of morphemes, Clause types, Reactivation effect, and Spillover effect.

There are the frequency of words, number of characters, number of morphemes, clause types (e.g., Mono-clause vs. Bi-clause), reactivation effect, and a spillover effect, where a pre-critical region influences a critical region (Vasishth and Lewis, 2006; Nakatani, 2021). Featherston (2001) reported the opposite results to Bever and McElree (1988), which may reflect the differences of these general factors between English and German.

To solve the problems of previous experimental studies, we used Japanese raising and control sentences, which assign the same θ-role (Proposition). We further introduced causative sentences to control the number of θ-roles and reflexive sentences to control the reactivation effect. As an anaphora expression in the reflexive sentences (e.g., *myself*) takes an antecedent within a sentence, the reflexive sentences clearly cause the reactivation effect. In other words, we used the causative and reflexive sentences as positive control conditions. We further used a mono-clausal sentence without ECs as a baseline condition (i.e., a negative control condition).

We used linear mixed-effects (LME) models, which are effective to examine the influence of each factor on the behavioral data. LME models show how certain independent variables (e.g., frequency and clause type) affect a dependent variable (e.g., reading time for each region), including participants and sets of experimental sentences as random factors. To investigate whether control sentences have PRO or not, two models were created (Table 2, Hypothesis 1). In the first model, the control sentences have PRO and the raising sentences have Copy. In the second model, both the control and raising sentences have Copy. If PRO is to be distinguished from Copy, the former model should show better scores than the latter model. In other words, the former model will show a lower Akaike information criterion (AIC) than the latter model (Akaike, 1974). To further examine whether ECs cause the reactivation effect, two models were created (Table 2, Hypothesis 2). In the first model, the ECs and reflexive cause the reactivation effect. In the second model, only the reflexive causes the reactivation effect. If ECs also cause the reactivation effect, the former model should show lower AIC scores than the latter model. Regarding Hypothesis 2, our research interest was to examine whether ECs showed the reactivation effect, but not to test the difference of reactivation effect between PRO and Copy. Therefore, we did not distinguish Copy from PRO in our analyses. The control and raising sentences are critical for testing Hypothesis

1, while the raising, control, and reflexive sentences are crucial for testing Hypothesis 2.

Materials and methods

Participants

We recruited 254 self-reported native speakers of Japanese through Lancers (https://www.lancers.jp/), a crowdsourcing service in Japan. Following a data trimming procedure, which will be explained in Section 2.4, two participants were removed from the datasets. The final set included 252 participants (146 males) between the age of 20 and 71 years (mean = 42.82, s.d. = 9.72). We used a Latin-square design and divided target sentences into five lists. Following the previous SPR study (Witzel and Witzel, 2011), which included 48 participants, we recruited about 50 participants for each of the five stimulus lists, resulting in 254 participants.

Stimuli

In this experiment, we used five types of sentence materials (see Supplementary material for all materials). Spaces indicate region boundaries for the presentation. All conditions consisted of 6 regions, and PRO and Copy were not presented to the participants (Figure 1). Note that the stimuli, glosses (word-by-word translations), and their English translations are shown here with the Modified Hepburn Romanization system of Japanese, but actual stimuli were presented in a combination of "kanji" and "hiragana". Vowels with a macron ($\bar{a}, i, u, \bar{e}, o$) denote long vowels.

(15) a. Raising condition.
    Nakamura$_i$-ga senshū kayōbi-ni [Copy$_i$ kaisha-de Takahashi-o shikari]-sugi-ta.
    Nakamura-NOM last week Tuesday-DAT at an office Takahashi-ACC scold-too much-PST.
    "Nakamura scolded Takahashi too much at an office last Tuesday."

   b. Control condition.
    Nakamura$_i$-ga senshū kayōbi-ni [PRO$_i$ kaisha-de Takahashi-o shikari]-sobire-ta.
    Nakamura-NOM last week Tuesday-DAT at an office Takahashi-ACC scold-fail to-PST.
    "Nakamura failed to scold Takahashi at an office last Tuesday."

   c. Reflexive condition.
    Nakamura-ga senshū kaisha-de jibunjishin-de Takahashi-o shikat-ta.
    Nakamura-NOM last week at an office myself Takahashi-ACC scold-PST.
    "Nakamura scolded Takahashi by himself/herself at an office last week."

   d. Causative condition.
    Nakamura-ga senshū [kaisha-de Yamashita-ni Takahashi-o shikar]-ase-ta.
    Nakamura-NOM last week at an office Yamashita-DAT Takahashi-ACC scold-CAUSE-PST.

TABLE 1 Different factors among conditions.

Conditions	ECs	Number of morphemes	Number of θ-roles	Clause type	Reactivation effect
Raising	Copy	3	3	Bi-clausal	?
Control	PRO	3	4	Bi-clausal	?
Reflexive	NA	2	2	Mono-clausal	+
Causative	NA	3	4	Bi-clausal	−
Baseline	NA	2	2	Mono-clausal	−

Reflexive, Causative, and Baseline conditions did not contain ECs. The number of morphemes indicates the total number of morphemes in the (compound) verb. The number of θ-roles indicates the total number of θ-roles within a sentence. The reactivation effect under the Raising and Control conditions remained unclear. NA, Not applicable; +, Having the reactivation effect; −, No reactivation effect.

"Nakamura made Yamashita scold Takahashi at an office last week."

e. Baseline condition.

Nakamura-ga senshū kayōbi-ni kaisha-de Takahashi-o shikat-ta.

Nakamura-NOM last week Tuesday-DAT at an office Takahashi-ACC scold-PST.

"Nakamura scolded Takahashi at an office last Tuesday."

In addition to the conditions of interest, i.e., the Raising and Control conditions, we included three additional conditions to control the influence of dependent variables. First, we used the Reflexive condition to investigate the influence of the reactivation effect. As we mentioned in the Introduction, whether ECs cause a reactivation effect is not clear. However, anaphora expressions (e.g., *jibunjishin and myself*) in reflexive sentences clearly cause a reactivation effect. The Causative condition was used to examine the influence of the number of θ-roles. The causative morpheme in Japanese assigns "Agent" θ-role to an external argument, e.g., *Nakamura* (Shibatani, 1973), thus, it resulted in the same number of θ-roles with the Control condition. Theoretical linguistics has proposed that causative constructions in Japanese have bi-clausal sentence structures (Shibatani, 1973). Finally, the Baseline condition, which did not contain ECs and therefore did not cause reactivation effect, was used as a baseline for the sentence that may cause reactivation effect (Raising, Control, and Reflexive conditions).

While the types of θ-roles are different between raising and control sentences in English (Pesetsky, 1991), both raising and control constructions assign the same θ-role, i.e., "Proposition", to an internal argument in Japanese (Kageyama, 2016). Therefore, we can control the difference between the θ-role types in the present study.

This experiment used 150 (30 × 5) target sentences and 150 filler sentences. The filler sentences had similar sentence construction to the target sentences, and they also had six regions. All the target sentences and half of the filler sentences were followed by the probe word. Following Bever and McElree (1988) and Featherston (2001), the sentence-initial noun phrase served as the probe word in the target sentences. To prevent the participants from anticipating the probe position, we set the probe word in different regions of the filler sentences. The other half of the filler sentences were followed by a "yes/no" comprehension task; for instance, "It was Yamashita that Takahashi scolded last Tuesday."

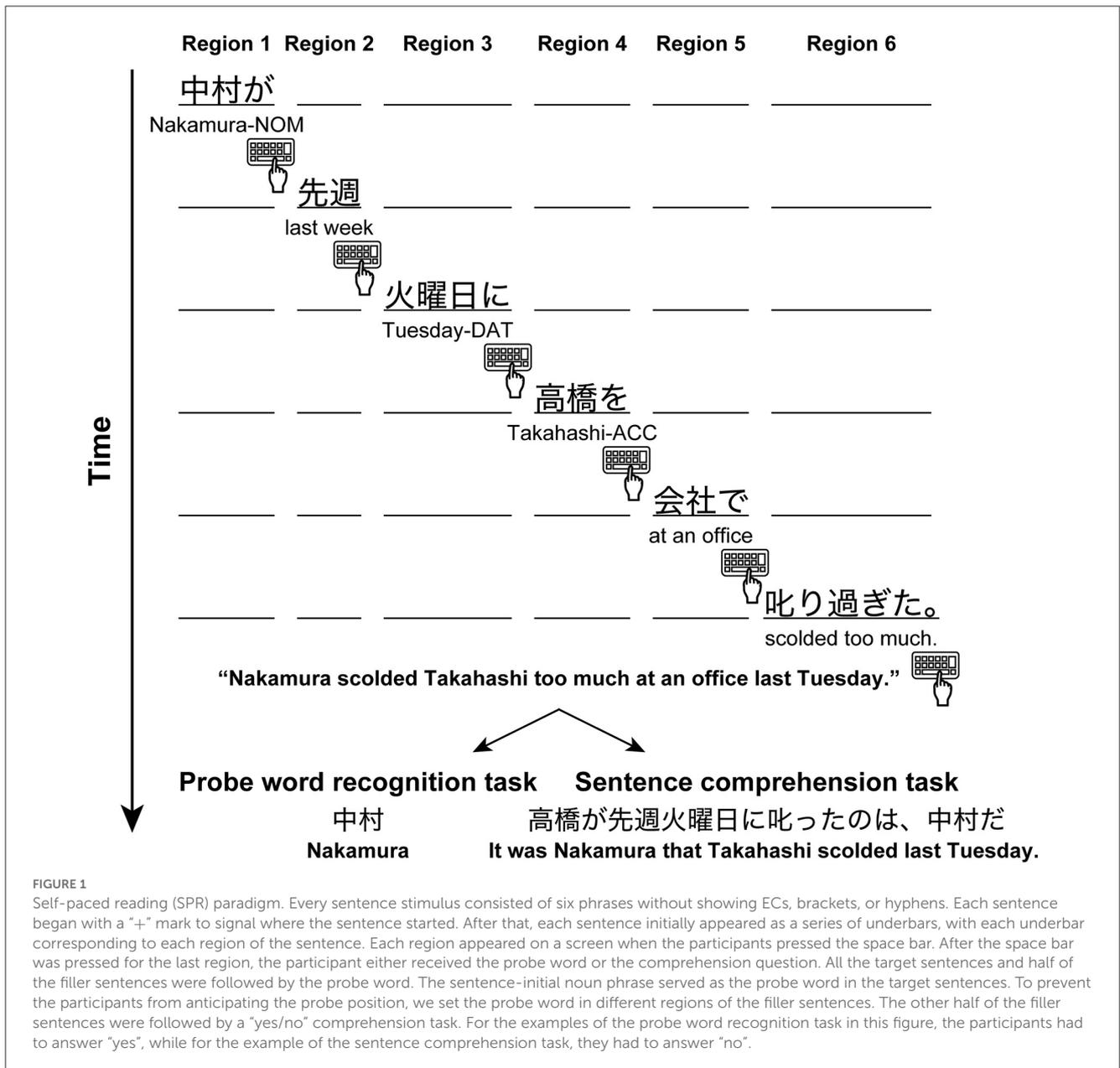TABLE 2 Hypotheses, models, and prediction.

Hypothesis 1:	PRO is distinguished from Copy
Model (PRO + Copy):	Control has PRO and raising has Copy.
Model (Copy + Copy):	Both control and raising have Copy.
Prediction:	Model (PRO + Copy) will show lower AIC.
Hypothesis 2:	**ECs cause the reactivation effect**
Model (ECs + reflexive):	ECs and reflexive cause the reactivation effect (i.e., shorter probe word recognition time).
Model (reflexive):	Only reflexive causes the reactivation effect (i.e., shorter probe word recognition time).
Prediction:	Model (ECs + reflexive) will show lower AIC.

for (15d). All the human-related proper and common noun phrases consist of two characters and four morae in Japanese.

Procedures

The experiment was run on PCIbex (https://doc.pcibex.net/), an online linguistic experiment hosting service (Zehr and Schwarz, 2018). Before the main experiment, participants read the instructions and received a maximum of three practice sessions. Each session consisted of five practice trials. Only participants who answered four or more trials correctly in any sessions were moved to the main experiment. For example, when a certain participant could answer four trials correctly in the first session, the second and third practice sessions were skipped. If a certain participant failed to get four or more scores in any session, the participant was refused to participate in the experiment. The experiment took ∼15–20 min, including the time to read the instructions and the practice parts. As compensation, JPY 120 was paid to each participant.

In this experiment, we used a non-cumulative moving window SPR paradigm and the *probe word recognition priming technique*. A total of 30 × 5 target sentences were distributed into five lists using a Latin-square design. Thirty filler items were added to each list. A total of 60 sentences shuffled in a pseudo-random order, with

FIGURE 1
Self-paced reading (SPR) paradigm. Every sentence stimulus consisted of six phrases without showing ECs, brackets, or hyphens. Each sentence began with a "+" mark to signal where the sentence started. After that, each sentence initially appeared as a series of underbars, with each underbar corresponding to each region of the sentence. Each region appeared on a screen when the participants pressed the space bar. After the space bar was pressed for the last region, the participant either received the probe word or the comprehension question. All the target sentences and half of the filler sentences were followed by the probe word. The sentence-initial noun phrase served as the probe word in the target sentences. To prevent the participants from anticipating the probe position, we set the probe word in different regions of the filler sentences. The other half of the filler sentences were followed by a "yes/no" comprehension task. For the examples of the probe word recognition task in this figure, the participants had to answer "yes", while for the example of the sentence comprehension task, they had to answer "no".

no more than three targets or three fillers, can be presented in a row. After every 20 items, the participants were encouraged to take a short rest. Each sentence began with a "+" mark to signal where the sentence started. After that, each sentence initially appeared as a series of dashes, with each dash corresponding to each region of the sentence (Figure 1). When the space bar was pressed, each region showed up on a screen. The participant continued in this manner until the end of the sentence. After the space bar press for the last region, the item was removed from the screen. Then, the participant either received the probe word or the comprehension question. The probe word and the comprehension question were enclosed by two angles. If the probe word was presented, the participant answered "included" with the "F" key on the keyboard or "not included" with the "J" key. If the comprehension question was presented, the participant answered "yes" with the "F" key on the keyboard or "no" with the "J" key.

Data analysis

Following Witzel and Witzel (2011), data trimming was conducted in the following way. The data from participants (a) with error rates of 30% or greater on the target and filler sentences or (b) with error rates of 30% or greater on the target sentences were eliminated from the analysis. The data set of one participant was excluded based on these cut-off scores. For the probe word recognition task, the data that were answered incorrectly were eliminated from further analysis. Thus, 4.51% of the data were eliminated in this way. We also excluded trials including the reading time data or the probe word recognition time data longer than 4 s. The data set of 1 participant was excluded based on this procedure. Thus, 4.68% of the data were eliminated in this way. The outlier reading time data for each region and the recognition time data for the probe word were then trimmed as follows: the data that

were two standard deviations above or below a subject's mean for a given region or task were replaced with the value two standard deviations above or below the participants' mean for that region or task. Thus, 4.74% of the data were trimmed in this way.

As for statistical analyses, reading and response time data were converted into natural logarithms. This study aims to compare the ECs of raising and control structures. The property of ECs was decided only when participants read Region 6 (verb position); that is, the raising and control structures from Regions 1 to 5 had no differences. Thus, our regions of interest were Region 6 and the following probe word recognition time. To confirm whether the behavioral data show the main effect of condition (Raising, Control, Reflexive, Causative, and Baseline), we conducted one-way repeated-measures analyses of variance (rANOVAs) for each region, the probe word recognition time, and accuracy. We used the "anovakun" function (version 4.8.7, http://riseki.php.xdomain.jp/index.php?ANOVA%E5%90%9B) in R (version 4.2.1). For *post-hoc* comparisons among the conditions, we applied Shaffer's modified sequentially rejective Bonferroni procedure.

For the above reason, LME models were only fitted for Region 6 and the probe word recognition time, using the *lmerTest* package in R (Kuznetsova et al., 2017). Note that all conditions were included in the following LME analyses. We first created the most complicated models that included all the factors that may affect the behavioral data as a fixed effect (i.e., the number of θ-roles, characters, morpheme, clause types, reactivation effect, and spillover effect) as shown in (16) (see also Introduction and Table 1).

(16) Complex Model < − lmer (Region 6/Probe word recognition time ∼ Frequency + Number of characters + Number of θ-roles + Number of morphemes + Reactivation effect + Spillover effect + Clause type + (1|Participant) + (1|Item), data = Data).

The detailed description of each dependent variable is the following. First, the frequencies of words were collected through the Balanced Corpus of Contemporary Written Japanese (https://chunagon.ninjal.ac.jp/ver. *Chunagon* 2.7.0) (Maekawa et al., 2013). Second, the number of characters and the number of morphemes were based on the (compound) verbs (Region 6) (Table 1). Third, the number of θ-roles was collected from the entire sentence (Raising = 3, Control = 4, Reflexive = 2, Causative = 4, and Baseline = 2). Fourth, we assumed Raising, Control, and Reflexive conditions cause the reactivation effect (these three conditions were assigned the dummy argument 1, and the other two conditions were assigned the dummy argument 0). We also included the logarithmic time of the pre-critical region as a fixed effect (spillover effect). Finally, we assumed raising, control, and causative structures are bi-clausal, and the others are mono-clausal (Raising, Control, and Causative conditions are assigned the dummy argument 1, and the others were assigned the dummy argument 0). Finally, the factor of θ-role types was not included as a fixed effect, because both control and raising constructions assign the same θ-role (2.2 Stimuli).

Both the participants' and items' intercepts were included in the model as random factors (Baayen et al., 2008). We also

attempted to include random slopes as well, but we were forced to simplify the random effects until convergence failure and singularity warnings disappeared. Eventually, it was impossible to include random slopes.

We excluded irrelevant factors in Region 6 and the probe word recognition task by using the *step* function in the *lmerTest* package, resulting in the following fixed and random effects. To select significant factors, we used a stepwise backward elimination method widely used in the model selection of the LME analyses (Baayen, 2008; Matuschek et al., 2017).

(17) Region 6 ∼ Number of characters + Spillover effect (reading times of Region 5) + (1|Participant) + (1|Item).

(18) Probe word recognition time ∼ Spillover effect (reading times of Region 6) + Reactivation effect + Number of morphemes + (1|Participant) + (1|Item).

We fitted the number of characters and spillover effect as fixed effects, which showed significant effects, and added the ECs for the reading time of Region 6, as shown in (19) and (20). It should be noted that the variables of interest in our study were PRO, Copy, Reactivation, and Reflexive, thus, we included these factors without applying the stepwise variable selection in the following analyses.

We assigned the different dummy arguments for the Raising and Control conditions for the model (PRO + Copy) (PRO: Raising condition has dummy argument 1 and the others have dummy argument 0; Copy: Control condition has dummy argument 1 and the others have dummy argument 0). On the other hand, we used the same dummy argument for the Raising and Control conditions for the model (Copy + Copy) (Raising and Control conditions have dummy argument 1, and the others have dummy argument 0).

(19) Model (PRO + Copy) < − lmer (Region 6 ∼ Number of characters + Spillover effect + PRO + Copy + (1|Participant) + (1|Item), data = Data).

(20) Model (Copy + Copy) < − lmer (Region 6 ∼ Number of characters + Spillover effect + Copy + (1|Participant) + (1|Item), data = Data).

We also fitted the reactivation effect, spillover effect, and the number of morphemes as fixed effects and added the ECs for the reaction time of probe word recognition task, as shown in (21) and (22). Same as in the above models (19) and (20), we assigned the different dummy arguments for the model (PRO + Copy) and the same dummy argument for the model (Copy + Copy).

(21) Model (PRO + Copy) < − lmer (Probe word recognition time ∼ Spillover effect + Reactivation effect + Number of morphemes + PRO + Copy + (1|Participant) + (1|Item), data = Data).

(22) Model (Copy + Copy) < − lmer (Probe word recognition time ∼ Spillover effect + Reactivation effect + Number of morphemes + Copy + (1|Participant) + (1|Item), data = Data).

We also fitted the spillover effect and number of morphemes as a fixed effect for the reaction time of probe word recognition task,

as shown in (23) and (24), to reveal the reactivation effect of ECs. We created two models. One was the model which assumed ECs and Reflexive conditions caused the reactivation effect (Raising, Control, and Reflexive conditions have dummy argument 1, and the other conditions have dummy argument 0). The other was the model which assumed only the Reflexive condition caused the reactivation effect (Reflexive condition has dummy argument 1, and the other conditions have dummy argument 0).

(23) Model (ECs + Reflexive) < − lmer (Probe word recognition time ~ Spillover effect + Number of morphemes + Reactivation + (1|Participant) + (1|Item), data = Data).

(24) Model (Reflexive) < − lmer (Probe word recognition time ~ Spillover effect + Number of morphemes + Reflexive + (1|Participant) + (1|Item), data = Data).

Previous studies reported reduced replicability of the results when selecting variables based on their statistical significance (Henderson and Denison, 1989; Mundry and Nunn, 2009). To check whether or not the variable selection caused any problems, we also tested the LME models including all variables hypothesized to be relevant, i.e., models without applying the stepwise variable selection. In the LME analyses comparing PRO and Copy, we could not include the number of θ-roles and the number of morphemes simultaneously due to the convergence failure. Thus, we created two models: One included all the variables except the number of morphemes, while the other included all the variables except the number of θ-roles. In addition, we included all variables without convergence failure when comparing the model that ECs and reflexive cause the reactivation effect with the model that only reflexive causes the reactivation effect.

Results

The main effect of condition

Mean of the raw reading times for each region and the probe word recognition times are summarized in Figure 2. The detailed results of the probe word recognition times are shown in Figure 3A. For Regions 1–5, there was no significant main effect of condition [Region 1: $F_{(4, 988)} = 0.14$, $p = 0.97$; Region 2: $F_{(4, 988)} = 0.21$, $p = 0.93$; Region 3: $F_{(4, 988)} = 0.28$, $p = 0.89$; Region 4: $F_{(4, 988)} = 2.1$, $p = 0.081$; Region 5: $F_{(4, 988)} = 1.8$, $p = 0.14$]. For Region 6, we found a significant effect of condition [$F_{(4, 988)} = 9.4$, $p < 0.0001$]. *Post-hoc* comparisons showed that the reading times of the Control condition were significantly longer than those of the Baseline, Causative, and Reflexive conditions [vs. Baseline: $t_{(247)} = 5.4$, corrected $p < 0.0001$; vs. Causative: $t_{(247)} = 3.3$, corrected $p = 0.0061$; vs. Reflexive: $t_{(247)} = 4.1$, corrected $p = 0.0003$]. In addition, the Raising condition also showed significantly longer reading times than the Baseline and Reflexive conditions [vs. Baseline: $t_{(247)} = 4.8$, corrected $p < 0.0001$; vs. Reflexive: $t_{(247)} = 4.6$, corrected $p < 0.0001$]. These results suggested that the Raising and Control conditions were more demanding. For the probe word recognition time, the effect of condition was also significant [$F_{(4, 988)} = 13$, $p < 0.0001$]. *Post-hoc* comparisons showed that the probe word recognition time of the Causative condition was

significantly longer than other conditions [vs. Raising: $t_{(247)} = 4.4$, corrected $p = 0.0001$; vs. Control: $t_{(247)} = 4.2$, corrected $p = 0.0002$; vs. Reflexive: $t_{(247)} = 6.0$, corrected $p < 0.0001$; vs. Baseline: $t_{(247)} = 5.3$, corrected $p < 0.0001$]. These results indicate that the difference of the sentence conditions changed the processing loads for the probe word recognition task.

The mean accuracy under every condition was higher than 95%, indicating the participants' reliable and consistent judgments on the task (Figure 3B). Furthermore, the rANOVA on the accuracy showed that the effect of condition was not significant [$F_{(4, 988)} = 0.55$, $p = 0.70$], further suggesting that the task difficulty was controlled among conditions.

Comparison of PRO and Copy

To investigate whether control and raising sentences have PRO and Copy, respectively, we compared two models, i.e., models (PRO + Copy) and (Copy + Copy), for Region 6 (verb position) (Table 3). The model that hypothesizes PRO is distinguished from Copy showed a lower AIC score[3], suggesting that the participants processed PRO and Copy differently. The summary of the results from the best-fitting model (PRO + Copy) is shown in Table 3. The main effects were found in the number of characters factor and spillover factor, as well. The LME models including all relevant variables, i.e., models without applying the stepwise variable selection, showed similar results (Supplementary Tables S1, S2).

The result of model comparison for the probe word recognition times is presented in Table 4. Contrary to the previous result, the model hypothesizing that PRO is *not* distinguished from Copy showed a lower AIC score. The summary of the results from the best-fitting model (Copy + Copy) is also shown in Table 4. The main effects were found in the spillover effect, the number of morphemes, and Copy. The LME models without applying the stepwise variable selection also showed similar results (Supplementary Tables S3, S4). The number of morphemes was not significant in this model, while the number of characters was significant, reflecting a positive correlation between these variables.

Reactivation effect with ECs

To further examine whether ECs cause the reactivation effect, we compared two models, i.e., models (ECs + Reflexive) and (Reflexive), for the probe word recognition times. The result of model comparison for the probe word recognition times is presented in Table 5. The model which hypothesizes that both ECs and the Reflexive cause the reactivation effect showed a lower AIC score, suggesting the psychological reality of ECs. The summary of the results from the best-fitting model (ECs + Reflexive) is also shown in Table 5. The main effects were found in the spillover effect, reactivation effect, and number of morphemes. The LME models without applying the stepwise variable selection also

---

3 We compared models using Akaike information criterion (AIC), which is a measure of model quality that is based on the log likelihood and number of parameters of the model.
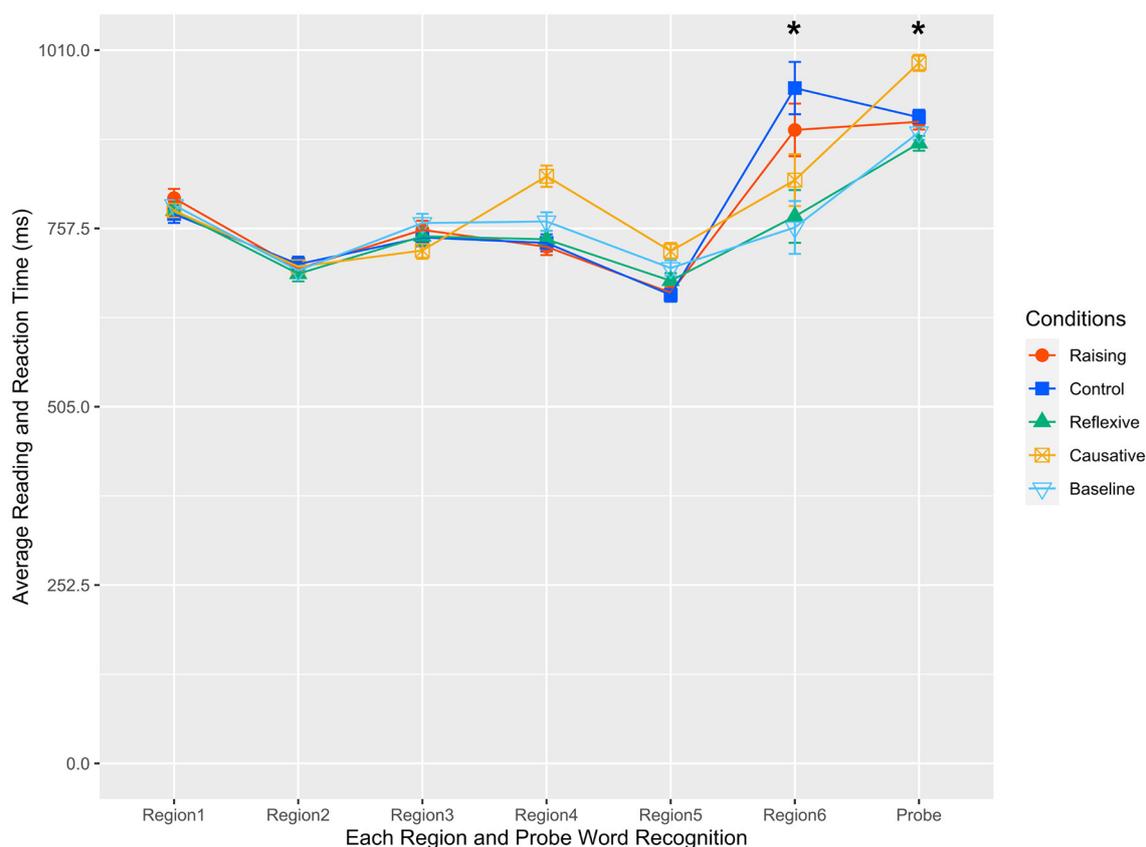
FIGURE 2
Average reading times and probe word recognition times. Error bars represent the standard error of the mean (SEM). *Corrected $p < 0.05$.
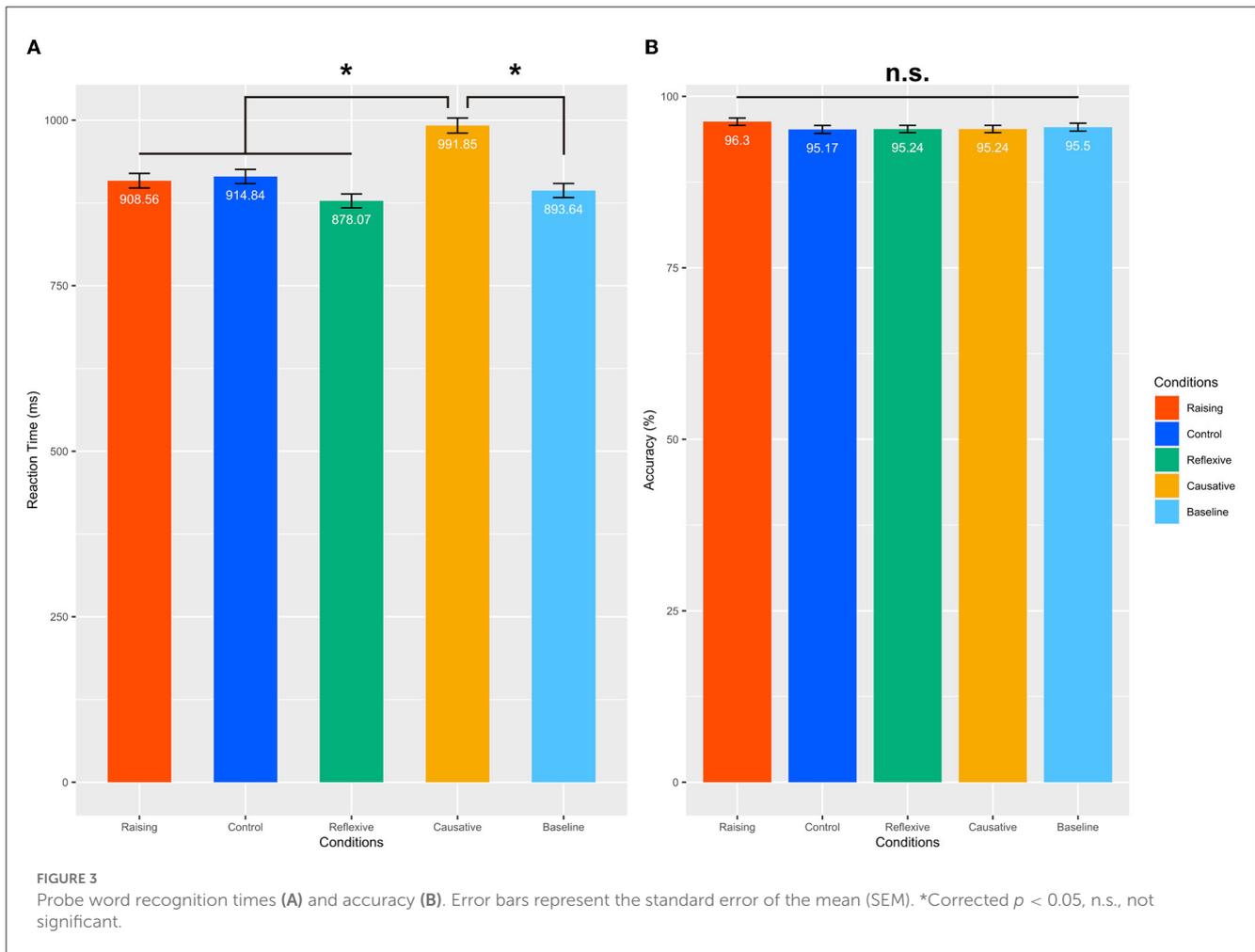
showed similar results (Supplementary Table S5). In this model, the number of morphemes was not significant, while the number of characters was significant, reflecting a positive correlation between these variables.

Discussion

In the present study, we investigated behavioral data of raising and control sentences, using the SPR paradigm and the *probe word recognition priming technique* with LME models (Figure 1), which shows how certain independent factors affect the behavioral output (Tables 1, 2). We found two striking results. First, we demonstrate that the reading times of raising and control sentences in Japanese were better explained by the LME model considering the differences of ECs (Figure 2; Table 3), suggesting the psychological reality of PRO and Copy. Secondly, we found a significant reactivation effect for raising and control sentences, which have ECs, and reflexive sentences without ECs (Table 5). These results indicate that ECs are processed similarly to reflexive pronouns (e.g., *himself/herself*). Based on these results, we conclude that raising and control sentences in Japanese have different ECs, i.e., Copy and PRO, and that ECs have psychological reality. Our results demonstrate that behavioral experiment based on theoretical linguistics, which is the first step for developing linking hypotheses that meaningfully relate neural circuits to syntactic

processing (Krakauer et al., 2017), is indeed necessary for testing hypotheses proposed in theoretical linguistics. A more formulated hypothesis that compares raising and control structures is needed for further studies.

Previous experimental studies contributed to revealing syntactic characteristics of ECs on human sentence processing, but their studies are still unsatisfactory. The problem with previous studies is that they do not consider other factors that may influence the behavioral data and ERP. As shown in Table 1, control and raising sentences have other differences along with the difference between PRO and Copy. First, control predicates assign two θ-roles to an internal and external argument. Contrarily, raising predicates assign only one θ-role to an internal argument. There is a difference between raising and control structures in terms of the number of θ-roles. Second, Pesetsky (1991) reported that many control predicates assign "Irrealis" to an internal argument, but raising predicates assign "Proposition" to an internal argument. There is another difference between raising and control structures in terms of the types of θ-roles. Therefore, it is too early to conclude that the observed differences in previous studies were really derived from the difference of ECs. To solve the problems of previous experimental studies, LME models were used to show how certain independent variables affect a dependent variable and Japanese stimuli were used because of the same θ-role to an internal argument.

FIGURE 3
Probe word recognition times **(A)** and accuracy **(B)**. Error bars represent the standard error of the mean (SEM). *Corrected $p < 0.05$, n.s., not significant.

TABLE 3 Model comparison (ECs) and results of the LME models for Region 6.

Models	No. parameters	AIC	Log likelihood	Deviance	χ^2	df	p
Model (PRO + Copy)	8	5,539.1	−2,761.6	5,523.1	4.85	1	0.028
Model (Copy + Copy)	7	5,542.0	−2,764.0	5,528.0	0.146	1	0.70
Model (PRO + Copy)	Estimate	SE	t-value		df		p
(Intercept)	4.2	0.10	39.7		1,990		<0.0001
Number of characters	0.072	0.013	5.40		138.1		<0.0001
Spillover effect	0.32	0.014	22.8		6,851		<0.0001
PRO	0.043	0.038	1.15		138.1		0.25
Copy	−0.036	0.042	−0.863		136.9		0.39

As shown in Figure 3B, the results of rANOVA showed no significant difference on Regions 1–5 and the accuracy. As our regions of interest were Region 6 and the probe word recognition time, we investigated them in detail, using LME models. We found that the number of characters and the reading time of the pre-region strongly affect the reading time (Table 3). Furthermore, the spillover effect, number of morphemes, and Copy firmly influenced the probe word recognition time (Tables 4, 5).

Moreover, there is a significant difference between raising and control constructions in the SPR paradigm, as shown in Figure 2 and Table 3. These results support the theory that distinguishes between PRO and Copy. However, Figure 3A and Table 4 seemingly show the opposite result, which may support MTC. It is worth noting that the SPR and the probe word recognition task might reflect different mental processes. Online sentence processing may be strongly reflected in the SPR paradigm. Furthermore, a memory retrieval process may be reflected in the probe word recognition

TABLE 4 Model comparison (ECs) and results of the LME models for the probe word recognition task.

Models	No. parameters	AIC	Log likelihood	Deviance	χ^2	df	p
Model (PRO + Copy)	9	1,507.2	−744.6	1,489.2	0.115	1	0.74
Model (Copy + Copy)	8	1,505.3	−744.7	1,489.3	5.04	1	0.025
Model (Copy + Copy)	**Estimate**	**SEM**	**t-value**	**df**		**p**	
(Intercept)	5.7	0.095	60.7	309.7	<0.0001		
Spillover effect	0.12	0.0085	14.3	6,105	<0.0001		
Number of morphemes	0.10	0.030	3.45	139.1	0.0007		
Reactivation effect	−0.031	0.030	−1.02	138.2	0.31		
Copy	−0.089	0.040	−2.24	138.7	0.027		

TABLE 5 Model comparison (reactivation effect) and results of the LME model for the probe word recognition task.

Models	No. parameters	AIC	Log likelihood	Deviance	χ^2	df	p
Model (ECs + reflexive)	7	1,508.4	−747.2	1,494.4	19.6	2	<0.0001
Model (reflexive)	7	1,523.4	−754.7	1,509.4	0	0	N/A
Model (ECs + reflexive)	**Estimate**	**SEM**	**t-value**	**df**		**p**	
(Intercept)	5.9	0.076	77.6	561.2	<0.0001		
Spillover effect	0.12	0.0085	14.3	6,107	<0.0001		
Reactivation effect	−0.081	0.020	−4.07	140.0	<0.0001		
Number of morphemes	0.053	0.020	2.65	140.6	0.0090		

task. These results do not contain inconsistencies and might merely reflect different mental procedures.

From the results shown in Figure 2 and Table 3, it is natural to conclude that PRO is distinguished from Copy. The behavioral data of the probe word recognition task showed no difference between raising and control sentences; however, it is very likely that neural processes will show the difference as reported in Featherston et al. (2000). Contrastingly, it is unnatural to hypothesize that the different, or contrary, behaviors are derived from the same neural process.

Although MTC has theoretically attractive points, it has many descriptive and theoretical problems. MTC is motivated to eliminate construction-specific PRO and module. Until the Government and Binding Theory, control phenomena were analyzed regarding a construction-specific grammatical primitive, PRO and a construction-specific interpretive system, the control module. Hornstein's analysis was supported in English and other languages (Boeckx and Hornstein, 2006; Fujii, 2006; Takano, 2010). However, Hornstein's proposal has received crucial criticism (Culicover and Jackendoff, 2001; Landau, 2003; Bobaljik and Landau, 2009; Ndayiragije, 2012; Wood, 2012). One of the descriptive problems is how to distinguish the differences of behavior of raising and control sentences, which are explained by the different ECs, i.e., PRO and Copy. For example, Takano (2000) pointed out that cleft sentences can be derived from control constructions, but cannot be derived from raising constructions as shown in (25) and (26). In Hornstein's analysis, it is difficult to explain this difference.

(25) It was [PRO to be frank] that John tried. (Takano, 2000).

(26) *It was [John to be frank] that John seemed. (Takano, 2000, partially modified).

In addition, Hornstein's analysis faces a theoretical issue that deals with adjunct control. Since Ross (1967), the prohibition of extracting from an adjunct is known as adjunct island. Therefore, a simple *A-movement* cannot be applied to adjunct control. To solve this problem, Hornstein (2000) proposed that the operation Copy and Merge should be allowed to apply freely between Workspaces, yielding *Sideward movement*. However, the *Sideward movement* has been widely criticized by researchers (Landau, 2003, 2007), because it over-generates non-existing ungrammatical sentences. In short, Hornstein's proposal faced the descriptive and theoretical issues. Thus, the conclusion of our study is recognized as appropriate in generative syntax.

Our study also indicated that ECs cause the reactivation effect and have psychological reality (Table 5). This conclusion is appropriate, especially in theoretical linguistics. It has been proposed that ECs have no phonetic features, but have the same syntactic features as pronounced constituents.

Bever and McElree (1988) found that sentences with Copy evoked significantly faster response times than control sentences with PRO. However, Featherston (2001), who conducted a replication study of Bever and McElree (1988) using German sentences, reported no significant differences between PRO and Copy on the probe word recognition times, which were the same as ours. Therefore, we assumed these controversial results were derived from word order differences. English takes SVO order in

a complement clause, while both Japanese and German take SOV order in a complement clause.

Regarding the reactivation effect, Bever and McElree (1988) also reported that sentences with ECs and pronouns showed significantly faster response times than sentences without ECs. However, Featherston (2001) reported that sentences with ECs and pronouns did not show significantly faster probe word recognition times than those without ECs. Our results also showed no significant differences between the Raising, Control, and Reflexive conditions, which may cause the reactivation effect, and the Baseline condition on the probe word recognition times. We assumed that the difference in the word order could also explain these different results. English takes SVO order, while the Baseline conditions of our study and Featherston's study used SOV order. Therefore, the information of the object was active in English sentences because of SVO order and caused a stronger intervention effect in a memory retrieval process than in the Japanese and German studies. Contrarily, the information of the object was less active in Japanese and German because of the SOV order, which caused a weaker intervention effect.

As explained in the Introduction, we expected that the reactivation effect was related to the Raising, Control, and Reflexive conditions (see also Tables 1, 2). The results of LME models for the probe word recognition task demonstrated that the model that assumed the reactivation effect for the above three conditions was better than the model that assumed the reactivation effect for the Reflexive condition alone (Tables 5; Supplementary Table S5), which supported our Hypothesis 2. Moreover, the estimate of the reactivation effect was negative, further indicating that the probe word recognition times became shorter under these conditions. On the other hand, *post-hoc* comparison between the Reflexive and Baseline conditions was not significant. This result may seem odd because the Reflexive condition included a reflexive pronoun, which referred to the probe word, predicting shorter probe word recognition time than the Baseline condition. However, we think a simple mono-clausal construction of the Reflexive condition, which may cause a floor effect, can explain this result. Featherston (2001) also reported similar results, that is, mono-clausal sentences caused faster probe word recognition times than bi-clausal sentences. Moreover, the numbers of morphemes and θ-roles were smaller in the Reflexive condition than in the Raising and Control conditions (Table 1). Furthermore, the reading times of the pre-critical region (Region 6) under the Reflexive condition were shorter than those of the Raising and Control conditions (Figure 2), which may decrease the spillover effect. These factors were also related to the floor effect and reduced the reactivation effect in the Reflexive condition.

We also tested the LME models without applying the stepwise variable selection method (Supplementary Tables S1–S5). The models in which PRO was distinguished from Copy showed lower AIC scores in Region 6 (Supplementary Tables S1, S2). Moreover, the models in which PRO was *not* distinguished from Copy also showed lower AIC scores in the probe word recognition task (Supplementary Tables S3, S4). Finally, the models where both ECs and reflexive caused the reactivation effect showed lower AIC scores in the probe word recognition task (Supplementary Table S5). Taken together, these LME models supported the same conclusion as the models applying the stepwise variable selection methods.

To further investigate the neural processes of ECs, it is necessary to conduct experimental studies using neuroimaging techniques, such as ERP and fMRI. In the early years of experimental research, especially event-related potential (ERP) research, researchers focused primarily on syntactic (Neville et al., 1991; Osterhout and Holcomb, 1992; Friederici et al., 1993) and semantic (Kutas and Hillyard, 1980) violations and their electric indices, the LAN, N400, and P600. Moreover, neuroimaging studies using functional magnetic resonance imaging (fMRI) also focused on the neural basis of syntax (Dapretto and Bookheimer, 1999; Embick et al., 2000; Hashimoto and Sakai, 2002; Friederici et al., 2003; Musso et al., 2003). Recent fMRI studies have further examined the neural basis of a fundamental syntactic operation of human language, *Merge*, i.e., a simple and primitive combinatory operation that takes *n* syntactic objects and forms an unordered set of the syntactic objects (Chomsky, 1995). For instance, we demonstrated that the number of recursive applications of Merge accounted for syntax-selective activations in the left inferior frontal gyrus (L. IFG) (Ohta et al., 2013b; Tanaka et al., 2019; see also Ohta et al., 2013a for review). Other fMRI studies also reported that the L. IFG is crucial for the Merge operation (Zaccarella and Friederici, 2015; Zaccarella et al., 2017; Wu et al., 2019; Trettenbrein et al., 2021). Moreover, a growing body of work uses computational models to predict neural activity during sentence comprehension or production (Brennan et al., 2016; Hale, 2016; Li and Hale, 2019; Oseki and Marantz, 2020). For example, Brennan et al. (2016) reported that the number of nodes predicted the time course of participants' fMRI BOLD signal while they were listening to a natural story. In addition, neurostimulation techniques, such as transcranial electrical stimulation and transcranial magnetic stimulation, are also necessary to reveal causal relationships between language processing and neural activation. Furthermore, other types of control sentences proposed in theoretical linguistics, such as NOC, Adjunct control, and split control, should be examined in future studies.

Conclusion

To investigate the differences between raising and control sentences and whether or not ECs are psychologically real, we used the non-cumulative moving window SPR paradigm and the probe word recognition priming technique. As a result, we found that (1) raising and control sentences in Japanese have different ECs, i.e., Copy and PRO, and that (2) ECs cause the reactivation effect and they have psychological reality.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Institutional Review Board of Kyushu University, Faculty of Humanities. The patients/participants provided their written informed consent to participate in this study.

Author contributions

KY wrote the first draft of the manuscript. All authors contributed to conception and design of the study, prepared the task materials, performed the statistical analysis, contributed to manuscript revision, read, and approved the submitted version.

Conflict of interest

The author SO declared that he was an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/flang.2023.1138749/full#supplementary-material

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19, 716–723. doi: 10.1109/TAC.1974.1100705

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R.* Cambridge: Cambridge University Press.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005

Betancort, M., Meseguer, E., and Carreiras, M. (2004). "The empty category PRO: processing what cannot be seen," in *The On-line Study of Sentence Comprehension: Eyetracking, ERP, and Beyond*, eds. M. Carreiras and C. Clifton Jr. (New York, NY: Psychology Press), 95–118.

Bever, T. G., and McElree, B. (1988). Empty categories access their antecedents during comprehension. *Linguist Inq.* 19, 35–43.Z

Bobaljik, J. D., and Landau, I. (2009). Icelandic control is not A-movement: the case from case. *Linguist Inq.* 40, 113–132. doi: 10.1162/ling.2009.40.1.113

Boeckx, C., and Hornstein, N. (2006). Control in Icelandic and theories of control. *Linguist Inq.* 37, 591–606. doi: 10.1162/ling.2006.37.4.591

Brennan, J. R., Stabler, E. P., Van Wagenen, S. E., Luh, W. M., and Hale, J. T. (2016). Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain Lang.* 157–158, 81–94. doi: 10.1016/j.bandl.2016.04.008

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: The MIT Press.

Chomsky, N. (1973). "Conditions on Transformations," in *A Festschrift for Morris Halle*, eds. S. R. Anderson and P. Kiparsky (New York, NY: Holt, Rinehart, and Winston), 232–286.

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht, Holland; Cinnaminson, NJ: Foris Publications.

Chomsky, N. (1993). "A minimalist program for linguistic theory," in *The View From Building 20*, eds K. Hale and S. J. Keyser (Cambridge, MA: MIT Press), 167-176; 186–195.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: The MIT Press.

Chomsky, N. (2021). Minimalism: Where are we now, and where can we hope to go. *Gengo Kenkyu J. Linguistic Soc. Japan* 160, 1–41. doi: 10.11435/gengo.160.0_1

Culicover, P. W., and Jackendoff, R. (2001). Control is not movement. *Linguist Inq.* 32, 493–512. doi: 10.1162/002438901750372531

Dapretto, M., and Bookheimer, S. Y. (1999). Form and content: dissociating syntax and semantics in sentence comprehension. *Neuron* 24, 427–432. doi: 10.1016/S0896-6273(00)80855-7

Davies, W. D., and Dubinsky, S. (2004). *The Grammar of Raising and Control: A Course in Syntactic Argumentation*. Malden, MA: Blackwell Publishing.

Demestre, J., and García-Albea, J. E. (2007). ERP evidence for the rapid assignment of an (appropriate) antecedent to PRO. *Cogn. Sci.* 31, 343–354. doi: 10.1080/15326900701221512

Demestre, J., Meltzer, S., García-Albea, J. E., and Vigil, A. (1999). Identifying the null subject: evidence from event-related brain potentials. *J. Psycholinguist. Res.* 28, 293–312. doi: 10.1023/A:1023258215604

Embick, D., Marantz, A., Miyashita, Y., O'Neil, W., and Sakai, K. L. (2000). A syntactic specialization for Broca's area. *Proc. Natl. Acad. Sci. U. S. A.* 97, 6150–6154. doi: 10.1073/pnas.100098897

Featherston, S. (2001). *Empty Categories in Sentence Processing*. Amsterdam: John Benjamins Publishing Company.

Featherston, S., Gross, M., Münte, T. F., and Clahsen, H. (2000). Brain potentials in the processing of complex sentences: an ERP study of control and raising constructions. *J. Psycholinguist. Res.* 29, 141–154. doi: 10.1023/A:1005188810604

Friederici, A. D., Pfeifer, E., and Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cogn. Brain Res.* 1, 183–192. doi: 10.1016/0926-6410(93)90026-2

Friederici, A. D., Rüschemeyer, S.-A., Hahne, A., and Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cereb. Cortex* 13, 170–177. doi: 10.1093/cercor/13.2.170

Fujii, T. (2006). *Some theoretical issues in Japanese control* (Ph.D. dissertation), University of Maryland. Available online at: http://hdl.handle.net/1903/4159 (accessed May 13, 2023).

Gibson, E., and Hickok, G. (1993). Sentence processing with empty categories. *Lang. Cogn. Process* 8, 147–161. doi: 10.1080/01690969308406952

Hale, J. (2016). Information-theoretical complexity metrics. *Lang. Linguist Compass* 10, 397–412. doi: 10.1111/lnc3.12196

Hashimoto, R., and Sakai, K. L. (2002). Specialization in the left prefrontal cortex for sentence comprehension. *Neuron* 35, 589–597. doi: 10.1016/S0896-6273(02)00788-2

Henderson, D. A., and Denison, D. R. (1989). Stepwise regression in social and psychological research. *Psychol. Rep.* 64, 251–257. doi: 10.2466/pr0.1989.64.1.251

Hornstein, N. (1999). Movement and control. *Linguist. Inq.* 30, 69–96. doi: 10.1162/002438999553968

Hornstein, N. (2000). *Move!: A Minimalist Theory of Construal.* Malden, MA: Blackwell.

Hornstein, N. (2003). "On Control," in *Minimalist Syntax*, ed. R. Hendrick (Malden, MA: Blackwell), 6–81.

Kageyama, T. (1993). *Bunpo-to gokeisei [Grammar and Word Formation].* Tokyo: Hituzi Syobo.

Kageyama, T. (2016). "Verb-compounding and verb-incorporation," in *Handbook of Japanese Lexicon and Word Formation*, eds. T. Kageyama and H. Kishimoto (Berlin, Boston, MA: De Gruyter Mouton), 273–310.

Koizumi, M. (1999). *Phrase Structure in Minimalist Syntax.* Tokyo: Hituzi Syobo.

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., and Poeppel, D. (2017). Neuroscience needs behavior: correcting a reductionist bias. *Neuron* 93, 480–490. doi: 10.1016/j.neuron.2016.12.041

Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science.* 207, 203–205. doi: 10.1126/science.7350657

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13

Landau, I. (2003). Movement out of control. *Linguist Inq.* 34, 471–498. doi: 10.1162/002438903322247560

Landau, I. (2007). "Movement-resistant aspects of control," in *New Horizons in the Analysis of Control and Raising*, eds. W. D. Davies and S. Dubinsky (Dordrecht: Springer Netherlands), 293–325.

Landau, I. (2013). *Control in Generative Grammar: A Research Companion.* New York, NY: Cambridge University Press.

Li, J., and Hale, J. (2019). "Grammatical predictors for fMRI time-courses," in *Minimalist Parsing*, eds. R. C. Berwick and E. P. Stabler (Oxford: Oxford University Press), 159–173.

Maekawa, K., Yamazaki, M., Ogiso, T., Maruyama, T., Ogura, H., Kashino, W., et al. (2013). Balanced corpus of contemporary written Japanese. *Lang. Resour. Eval.* 48, 345–371. doi: 10.1007/s10579-013-9261-0

Makuuchi, M., Grodzinsky, Y., Amunts, K., Santi, A., and Friederici, A. D. (2013). Processing noncanonical sentences in Broca's region: reflections of movement distance and type. *Cereb. Cortex* 23, 694–702. doi: 10.1093/cercor/bhs058

Marr, D. (1982). *Vision : A Computational Investigation into the Human Representation and Processing of Visual Information.* San Francisco, CA: W.H. Freeman.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2017). Balancing type I error and power in linear mixed models. *J. Mem. Lang.* 94, 305–315. doi: 10.1016/j.jml.2017.01.001

McElree, B., and Bever, T. G. (1989). The psychological reality of linguistically defined gaps. *J. Psycholinguist. Res.* 18, 21–35. doi: 10.1007/BF01069044

Miyamoto, E. T., and Takahashi, S. (2002). "Antecedent reactivation in the processing of scrambling in Japanese," in *The Proceedings of HU-MIT 2001, MIT Working Papers in Linguistics*, eds. T. Ionin, H. Ko, and A. Nevins (The MIT Press), 123–138. Available online at: https://cir.nii.ac.jp/crid/1574231875408074112 (accessed May 13, 2023).

Mundry, R., and Nunn, C. L. (2009). Stepwise model fitting and statistical inference: turning noise into signal pollution. *Am. Nat.* 173, 119–123. doi: 10.1086/593303

Musso, M., Moro, A., Glauche, V., Rijntjes, M., Reichenbach, J., Büchel, C., et al. (2003). Broca's area and the language instinct. *Nat. Neurosci.* 7, 774–781. doi: 10.1038/nn1077

Nakano, Y., Felser, C., and Clahsen, H. (2002). Antecedent priming at trace positions in Japanese long-distance scrambling. *J. Psycholinguist. Res.* 31, 531–571. doi: 10.1023/A:1021260920232

Nakatani, K. (2021). "Locality effects in the processing of negative-sensitive adverbials in Japanese," in *The Joy and Enjoyment of Linguistic Research: A Festschrift for Takane Ito*, eds. R. Okabe, J. Yashima, Y. Kubota, and T. Isono (Tokyo: Kaitakusha), 462–472.

Nakayama, M. (1995). "Scrambling and probe recognition," in *Japanese Sentence Processing*, eds. R. Mazuka and N. Nagai (New York, NY: Psychology Press), 257–273.

Ndayiragije, J. (2012). On raising out of control. *Linguist. Inq.* 43, 275–299. doi: 10.1162/LING_a_00086

Neville, H., Nicol, J. L., Barss, A., Forster, K. I., and Garrett, M. F. (1991). Syntactically based sentence processing classes: evidence from event-related brain potentials. *J. Cogn. Neurosci.* 3, 151–165. doi: 10.1162/jocn.1991.3.2.151

Ohta, S., Fukui, N., and Sakai, K. L. (2013a). Computational principles of syntax in the regions specialized for language: integrating theoretical linguistics and functional neuroimaging. *Front. Behav. Neurosci.* 7, 204. doi: 10.3389/fnbeh.2013.00204

Ohta, S., Fukui, N., and Sakai, K. L. (2013b). Syntactic computation in the human brain: the degree of merger as a key factor. *PLoS ONE* 8, e56230. doi: 10.1371/journal.pone.0056230

Ohta, S., Koizumi, M., and Sakai, K. L. (2017). Dissociating effects of scrambling and topicalization within the left frontal and temporal language areas: an fMRI study in Kaqchikel Maya. *Front. Psychol.* 8, 748. doi: 10.3389/fpsyg.2017.00748

Oseki, Y., and Marantz, A. (2020). Modeling human morphological competence. *Front. Psychol.* 11, 513740. doi: 10.3389/fpsyg.2020.513740

Osterhout, L., and Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *J. Mem. Lang.* 31, 785–806. doi: 10.1016/0749-596X(92)90039-Z

Pesetsky, D. (1991). *Zero Syntax.* Vol. 2. Infinitives. Available online at: http://lingphil.mit.edu/papers/pesetsk/infins.pdf (accessed May 28, 2023).

Pickering, M., and Barry, G. (1991). Sentence processing without empty categories. *Lang. Cogn. Process* 6, 229–259. doi: 10.1080/01690969108406944

Rosenbaum, P. S. (1967). *The Grammar of English Predicate Complement Constructions.* Cambridge, MA: The MIT Press.

Ross, J. R. (1967). "*Constraints on variables in syntax,*" (Ph.D. dissertation). Massachusetts Institute of Technology. Available online at: http://hdl.handle.net/1721.1/15166 (accessed May 13, 2023).

Sakamoto, T. (1996). *Processing Empty Subjects in Japanese: Implications for the Transparency Hypothesis.* Fukuoka: Kyushu University Press.

Sakamoto, T. (2001). "Processing filler-gap constructions in Japanese: the case of empty subject sentence," in *Sentence Processing in East Asian Languages*, ed. M. Nakayama (Stanford, CA: CSLI Publications), 189–221.

Shibatani, M. (1973). Semantics of Japanese causativization. *Found. Lang.* 9, 327–373.

Takano, Y. (2000). Illicit remnant movement: an argument for feature-driven movement. *Linguist. Inq.* 31, 141–156. doi: 10.1162/002438900554325

Takano, Y. (2010). Scrambling and control. *Linguist. Inq.* 41, 83–110. doi: 10.1162/ling.2010.41.1.83

Tanaka, K., Nakamura, I., Ohta, S., Fukui, N., Zushi, M., Narita, H., et al. (2019). Merge-generability as the key concept of human language: evidence from neuroscience. *Front. Psychol.* 10, 2673. doi: 10.3389/fpsyg.2019.02673

Tanaka, K., Ohta, S., Kinno, R., and Sakai, K. L. (2017). Activation changes of the left inferior frontal gyrus for the factors of construction and scrambling in a sentence. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* 93, 511–522. doi: 10.2183/pjab.93.031

Trettenbrein, P. C., Papitto, G., Friederici, A. D., and Zaccarella, E. (2021). Functional neuroanatomy of language without speech: an ALE meta-analysis of sign language. *Hum. Brain Mapp.* 42, 699–712. doi: 10.1002/hbm.25254

Vasishth, S., and Lewis, R. L. (2006). Argument-head distance and processing complexity: explaining both locality and antilocality effects. *Language* 82, 767–794. doi: 10.1353/lan.2006.0236

Williams, E. (1980). Predication. *Linguist. Inq.* 11, 203–238.

Witzel, J. D., and Witzel, N. O. (2011). "The processing of Japanese control sentences," in *Processing and Producing Head-final Structures*, eds. H. Yamashita, Y. Hirose, and J. L. Packard (Dordrecht: Springer Netherlands), 23–47.

Wood, J. (2012). Against the movement theory of control: another argument from Icelandic. *Linguist. Inq.* 43, 322–330. doi: 10.1162/LING_a_00089

Wu, C. Y., Zaccarella, E., and Friederici, A. D. (2019). Universal neural basis of structure building evidenced by network modulations emerging from Broca's area: the case of Chinese. *Hum. Brain Mapp.* 40, 1705–1717. doi: 10.1002/hbm. 24482

Yamaguchi, K., and Ohta, S. (2023). Dissociating the processing of empty categories in raising and control sentences: a self-paced reading study in Japanese. *PsyArXiv.* doi: 10.31234/osf.io/zksb6

Zaccarella, E., and Friederici, A. D. (2015). Merge in the human brain: a sub-region based functional investigation in the left pars opercularis. *Front. Psychol.* 6, 1818. doi: 10.3389/fpsyg.2015.01818

Zaccarella, E., Meyer, L., Makuuchi, M., and Friederici, A. D. (2017). Building by syntax: the neural basis of minimal linguistic structures. *Cereb Cortex.* 27, 411–421. doi: 10.1093/cercor/bhv234

Zehr, J., and Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX).* Available online at: https://doi.org/10.17605/OSF.IO/MD832 (accessed May 28, 2023).