Re-examining Island Effects with NP-scrambling in Japanese: The Effect of Individual Variation*

SHIN FUKUDA

University of Hawai'i at Mānoa

HAJIME ONO

Tsuda University

NOZOMI TANAKA

Indiana University

JON SPROUSE

New York University Abu Dhabi

1 Introduction

Long-distance dependencies and island effects have been some of the most prolific areas of investigation in recent experimental syntactic studies. A handful of such studies investigated whether long-distance NP-scrambling

* We would like to thank Brian Agbayani, Chung-hye Han, Utako Minai, John Whitman, Kyoko Yamakoshi, Masaya Yoshida, the members of the Experimental Approaches to Theoretical Syntax lab at the University of Hawai'i at Mānoa, and the audiences at the JK 30 for their helpful comments and suggestions that improved this study.

Japanese/Korean Linguistics 30
Edited by Sara Williamson, Adeola Aminat Babayode-Lawal, Laurens Bosman, Nicole Chan, Sylvia Cho, Ivan Fong, and Kaye Holubowsky.

incurs island effects with potential island structures in Japanese and arrived at different conclusions.

Jurka (2010) and Jurka et al. (2011) argue that NP-scrambling out of complex NP subjects as opposed to complex NP objects incurs island effects, while Omaki et al. (2020) argue that there is no subject-object asymmetry in NP-scrambling. Yano (2019) provides experimental evidence for island effects with NP-scrambling out of noun complements and adjunct clauses. In Fukuda et al. (2022), we examined NP-scrambling out of noun complements, relative clauses, and coordinate structures, and argue that NP-scrambling exhibits clear island effects only with relative clauses and coordinate structures.

In Fukuda et al. (2022), we also found that, while a good number of participants gave predictably high ratings to items that involve by-hypothesis grammatical instances of NP-scrambling, e.g., NP-scrambling out of embedded declarative CPs, a non-negligible number of participants gave the same items surprisingly low ratings. In fact, the same tendency can be observed in Yano (2019) and Omaki et al. (2020). In these studies, NP-scrambling out of the intended non-island structures – embedded declarative CPs in Yano (2019) and complex NP objects in Omaki et al. (2022) – received ratings that are below the middle-of-the-scale ratings. These observations illustrate potential difficulties with investigating island effects with NP-scrambling, and suggest something did not go as expected in these previous studies.

This study reports on two acceptability judgment experiments that reexamine NP-scrambling out of three potential island structures in Japanese: subjects, *because*-adjunct clauses, and relative clauses. Experiment 1 examines island effects of NP-scrambling out of subjects and *because*-adjunct clauses, and Experiment 2 tests NP-scrambling out of subject and object relative clauses. The results of these experiments show significant individual differences in the effect size of NP-scrambling out of structures that are by-hypothesis non-islands, i.e., complex NP objects and declarative CPs. Once these individual differences are taken into consideration, our findings provide clear evidence of island effects with NP-scrambling out of *because*-adjunct clauses and relative clauses, and suggestive evidence of island effects with NP-scrambling out of subjects.

The rest of the study is structured as follows. Section 2 introduces the factorial definition of island effects (e.g., Sprouse 2007; Sprouse et al. 2011, 2012), which we adopt to interpret our results. Section 3 critically reviews five previous studies that examined NP-scrambling out of subjects (Jurka 2010; Jurka et al. 2011; Omaki et al. 2020), *because*-adjunct clauses (Yano 2019), and relative clauses (Fukuda et al. 2022), and identifies their potential issues. Section 4 discusses the results of Experiment 1 with subjects and *because*-adjunct clauses, and Section 5 re-evaluates the results of Experiment 1, taking into consideration individual differences in the effect size of NP-

scrambling. Section 6 discusses the results of Experiment 2 with subject and object relative clauses, and Section 7 concludes the study.

2 The Factorial Definition of Island Effects

The goal of the factorial definition of island effects is to *"isolate the acceptability effect that cannot be accounted for by known effects"* (Sprouse et al. 2016: 313). With NP-scrambling, there are two known factors that could affect acceptability: the presence of NP-scrambling and the effect of having a complex syntactic structure, i.e., a potential island. A factorial design acceptability judgment experiment can be constructed with these two factors: SCRAMBLING manipulates the presence or absence of NP-scrambling, and STRUCTURE manipulates the structure of the embedded clause. Fully crossing these two factors leads to the following four conditions.

(1) a. No-scrambling/non-island
b. No-scrambling/island
c. Scrambling/non-island
d. Scrambling/island

The factorial definition isolates the island effect in the interaction between SCRAMBLING and STRUCTURE. If there is no island effect, we expect to see no interaction as illustrated in the leftmost panel of Figure 1, where the two lines that connect the two means for the island condition items, (1b) and (1d), and the non-island condition items, (1a) and (1c), are parallel. If there is an island effect, we expect to see a superadditive interaction as illustrated in the center and rightmost panels, where the two lines are not parallel because the mean for the scrambling/island condition items, (1d), is lower than expected if the effects of the two manipulations are all there are. In statistical terms, this superadditive interaction manifests as a significant interaction between the two factors. We can also look at the size of the interaction as a measure of the size of the island effect; the center panel illustrates a smaller effect, and the rightmost panel illustrates a larger effect.

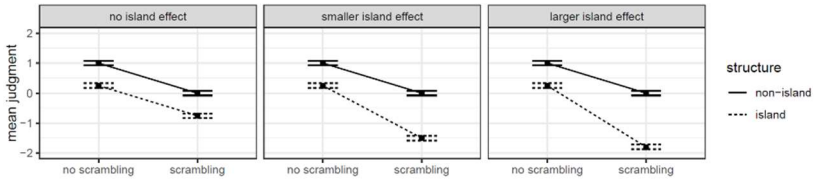

Figure 1. Possible results using the factorial definition of island effects.

3 Previous Experimental Studies on NP-scrambling out of Subjects, *Because*-adjunct Clauses, and Relative Clauses

This section reviews five previous studies that examined NP-scrambling out of subjects (Jurka 2010; Jurka et al. 2011; Omaki et al. 2020), *because*-adjunct clauses (Yano 2019), and relative clauses (Fukuda et al. 2022), and identify their potential issues.

3.1 NP-scrambling out of Subjects

To the best of our knowledge, Jurka (2010) and Jurka et al. (2011) are the first studies to examine NP-scrambling out of subjects in Japanese with acceptability judgment experiments. Jurka (2010) and Jurka et al. (2011) pointed out several issues in the data based on which previous studied claim that subjects are not islands to NP-scrambling in Japanese and examined acceptability of NP-scrambling out of complex NP subjects and complex NP objects with a 2 x 2 factorial design experiment that manipulated SCRAMBLING (no scrambling vs. scrambling) and STRUCTURE (complex NP subject vs. object). Examples of their experimental items in the no-scrambling condition are found in (2). In (2a-b) and all example sentences that follow, the scrambled constituents are outlined with a box.

(2) a. Non-island (complex NP object)/no-scrambling
Sono syouzyo-wa [CP iziwaruna ane1-ga [OBJ PRO1
that girl-TOP mean sister-NOM
kuma-no-nuigurumi-o suteta koto]-o naisyo-ni
teddy.bear-ACC dumped fact-ACC secret-DAT
siteita to] uttaeta.
kept that claimed
'The girl claimed that her mean sister kept as a secret the fact that she dumped her teddy bear.' (Jurka et al. 2011: 130; (7b))

b. Island (complex NP subject)/no-scrambling
Sono syouzyo-wa [CP [SUBJ iziwaruna ane-ga kuma-no-
that girl-TOP mean sister-NOM teddy.bear-
nuigurumi-o suteta koto]-ga kenka-no gen'in da to]
-ACC dumped fact-NOM fight-GEN cause be that
uttaeta.
claimed
'The girl claimed that the fact that her mean sister dumped her teddy bear is the cause of the fight.' (Jurka et al. 2011: 130; (7a))

The experiment presented three tokens per condition to twenty-seven self-claimed native speakers of Japanese and asked them to judge their naturalness with a 7-point scale. While their results showed that there was no significant difference in the mean acceptability judgments between complex NP subjects (2.73) and complex NP objects (2.85) in the scrambling condition, there was a significant difference between the mean acceptability judgments of complex NP subjects (6.93) and complex NP objects (5.79) in the no-scrambling condition. The interaction between SCRAMBLING and STRUCTURE was also significant. Based on these findings, Jurka (2010) and Jurka et al. (2011) concluded that Japanese subjects are islands with respect to NP-scrambling.

Omaki et al. (2020) challenge Jurka and his colleagues' conclusion, arguing that there is a confounding factor in their experiment. In particular, Omaki et al. (2020) point out that the complex NP object items in the non-scrambling condition (3a) are inconsistent with a well-known psycholinguistic constraint in Japanese, *long-before-short preference*, a preference to place longer constituents before shorter constituents (e.g., Dryer 1980; Hawkins 1994; Yamashita and Chang 2001) while the complex NP subject items in the same condition (3b) are consistent with it. In the schematic examples (3a-b), the complex NPs are highlighted in bold.

(3) a. Complex NP Object
NP-TOP [CP NP1-NOM
[OBJ PRO1 NP-ACC **VEMBEDDED fact]-ACC** NP Copula-C]
VMATRIX

b. Complex NP Subject
NP-TOP
[CP **[SUBJ NP-NOM** **NP-ACC** **VEMBEDDED fact]-NOM** NP Copula-C]
VMATRIX

In (3b), the complex NP subject is the first constituent inside the embedded clause, conforming to the long-before-short preference inside the embedded clause. In contrast, in (3a), the complex NP object is in the middle of the embedded clause after a shorter embedded subject, making the embedded clause inconsistent with the long-before-short preference. Omaki et al. (2020) conjecture that the significant difference in the mean acceptability judgments between the complex NP subject items and the complex NP object items in the no-scrambling condition in Jurka (2010) and Jurka et al. (2011) may have been because only the former is consistent with the long-before-short preference. To test their hypothesis, Omaki et al. (2020) conducted a 2 x 2 factorial design experiment manipulated SCRAMBLING (no scrambling vs. scrambling) and STRUCTURE (complex NP object vs. subject). Crucially, in their materials,

complex NP objects were scrambled to the embedded sentence initial position, as in (4), to be consistent with the long-before-short preference.

(4) According to NP,
[[OBJ **NP-NOM** **NP-ACC** **V EMBEDDED fact]-ACC/DAT1** NP-NOM ***t1***
V MATRIX]

Omaki et al. (2020) found a significant main effect of SCRAMBLING, but neither the main effect of STRUCTURE nor the interaction between these two factors was significant. Based on these findings, Omaki et al. (2020) concluded that there is no subject-object asymmetry in NP-scrambling in Japanese.

There is one potential issue in Omaki et al.'s (2020) experiment, however. Following the standard procedure, the study *z*-score transformed the raw scores elicited with a 7-point scale. The *z*-score transformation converts a participant's scores to units that represent the number of standard deviations a particular rating is from that participant's mean rating and corrects for the potential individual biases in treating the scale differently, e.g., using only a subset of the available ratings (Cowart 1997; Schütze and Sprouse 2014; Langsford et al. 2018). Therefore, the *z*-score value of zero represents each participant's middle-of-the-scale rating. As such, a positive *z*-score means that the structure was rated as more acceptable than the average rating, while a negative z-score means that it was rated as less acceptable than the average rating. In Omaki et al.'s (2020) experiment, the mean *z*-scores for the complex NP object items and the complex subject NP items in the scrambling condition were both below −0.25, i.e., they received mean ratings that are below the average. This raises the following question: If there were no island violations with these items, why were they rated so low?

3.2 NP-scrambling out of Adjunct Clauses

To our best knowledge, Yano (2019) is the only study that examined island effects of NP-scrambling out of adjunct clauses in Japanese with formal acceptability judgment experiments. Yano (2019) examined whether D(iscourse)-linked NPs like *sono hon* 'the book' undergo syntactic movement when they appear in a fronted position. Using island effects as a diagnostic of movement, the study tested two island types: adjunct clauses headed by *node* 'because' (*because*-adjunct clauses) and noun complements. Yano (2019) tested both D-linked NPs (with *sono* 'the/that') as the target of investigation, and non-D-linked NPs (without *sono* 'the/that') as a baseline comparison. Here we focus on NP-scrambling of non-D-linked NPs out of *because*-adjunct clauses, which incurs island effects according to Saito (1985).

(5) *Sono hon-o$_1$ John-ga [$_{ADJ}$ minna-ga t_1 kau node]
that book-ACC J-NOM all-NOM buy because
tigau hon-o katta
different book-ACC bought
('John bought a different book because everyone was going to buy that book.') (Saito 1985: 247; (147b))

Yano (2019) conducted two acceptability judgement experiments with a 2 x 2 x 2 factorial design with SCRAMBLING (no-scrambling vs. scrambling), STRUCTURE (declarative CP vs. *because*-adjunct clause), and D-LINKING (non-D-linked vs. D-linked). In both experiments, the same sentences were used and participants were asked to judge the naturalness of sentences with a 5-point scale. The only difference between the two experiments was that the sentences were presented with contexts only in Experiment 2. Below are examples of a non-island item with an embedded declarative CP (6a) and an island condition item with the *because*-adjunct clause (6b) (Yano 2019: 4).

(6) a. Non-island (declarative CP)/no-scrambling
Choonan-wa sakki [$_{CP}$ imooto-ga okashi-o tabeta-to]
brother-TOP a.while.ago sister-NOM snacks-ACC ate-C
omotteiru.
think
'The brother thinks that his younger sister ate snacks a little while ago.'

b. Island (*because*-adjunct clause)/no-scrambling
Choonan-wa sakki [$_{ADJ}$ imooto-ga okashi-o tabeta-node]
brother-TOP a.while.ago sister-NOM snacks-ACC ate-because
okotteiru.
angry
'The son is angry because his younger sister ate the snacks a while ago.'

Forty-two and forty-seven native speakers participated in Experiments 1 and 2, respectively. In both experiments, a significant interaction between SCRAMBLING and STRUCTURE was found within the non-Dinked condition items, providing experimental support to Saito's (1985) claim that NP-scrambling out of *because*-adjunct clauses incurs island effects.

Just like Omaki et al. (2020), the raw scores obtained in Yano's (2019) two experiments were *z*-score transformed, and it is worth noting that the mean *z*-scores for the non-island declarative CP items and the island *because*-adjunct clause items in the scrambling condition were below −0.5, i.e., they received mean ratings that are lower than the average rating.

3.3 NP-scrambling out of Relative Clauses

In Fukuda et al. (2022), we examined island effects with NP-scrambling out of relative clauses, noun complements, and coordinate structures. Early theoretical studies such as Haig (1976) and Saito (1985) agree that NP-scrambling out of relative clauses in Japanese incurs island effects.

(7) a. *Ano hon-o$_i$ watashi-wa [$_{RC}$ ec$_j$ ***t**$_i$* kaita hito$_j$]-ni
that book-ACC I-TOP wrote person-to
aitai.
want.to.meet
('I want to meet the person who wrote that book.') (Haig 1976: 370; (30))

b. ?*Ano hon-o$_i$ John-ga [$_{RC}$ ec$_j$ ***t**$_i$* katta hito$_j$]-o
that book-ACC J-NOM bought person-ACC
sagashiteiru rasii.
look.for seem
('It seems that John is looking for the person who bought that book.')
(Saito 1985: 246; (146a))

We conducted two factorial design acceptability judgment experiments for this study. Focusing on relative clauses, both experiments had a 2 x 2 design with SCRAMBLING (no-scrambling vs. scrambling) and STRUCTURE (declarative CP vs. relative clause).

(8) a. Non-island (declarative CP)/no-scrambling
Roodookumiai-no riidaa-wa [$_{CP}$ kaisha-no juuyaku-ga
union-GEN leader-TOP company-GEN executives-NOM
oohabana uriage-no nobi-o juugyooin-no kyuuyo-ni
drastic sales-GEN growth-ACC employee-GEN salary-to
han'ee saseteinai-to] hihanshi-ta.
reflect not.make-COMP criticize-PST
'The union leader criticized that the executives of the company were not making the drastic sales growth reflected in the employees' salaries.'

b. Island (relative clause)/no-scrambling
Roodookumiai-no riidaa-wa [$_{RC}$ ec$_1$ oohabana uriage-no
union-GEN leader-TOP drastic sales-GEN
nobi-o juugyooin-no kyuuyo-ni han'ee saseteinai]
growth-ACC employee-GEN salary-to reflect not.make
kaisha-no juuyakui$_1$-o hihanshi-ta.
company-GEN executives-ACC criticize-PST
'The union leader criticized the company's executives who were not making the drastic sales growth reflected in the employees' salaries.'

In Experiment 1, eighty-nine self-identified native speakers judged 58 items using a 7-point scale, of which 12 were the experimental items (3 islands x 4 conditions x 1 item per condition). In Experiment 2, the number of conditions was reduced by using declarative CPs as the non-island condition for all three island types, but the number of items was increased to two per condition. This made the number of experimental items 16 (8 conditions x 2 items per condition), which were combined with 44 fillers. A total of sixty items was judged by ninety-three self-identified native speakers using a 7-point scale. In both experiments, the interaction between SCRAMBLING and STRUCTURE was significant with the relative clause items, indicating the presence of island effects with NP-scrambling out of relative clauses.

One issue in the materials we used, like (8a-b), is that they all had subject relative clauses, following the relevant examples discussed the theoretical literature such as (5a-b). The choice of subject relative clauses is a potential confounding factor for the following reason. When the direct object inside a subject relative clause is scrambled, as in the schematic example (9), the scrambled direct object and the matrix topic NP would be directly followed by the VP inside the relative clause, creating a sequence of words that is likely to cause a garden-path effect. The parser might process the direct object NP and the topic NP as arguments of the embedded verb, only to discover later that the verb is inside the relative clause when it encounters the head NP.

(9) [NP ....]-ACC$_2$ NP-TOP [[RC e_1 t_2 $V_{EMBEDDED}$] head NP_1] V_{MATRIX}

Therefore, the observed island effect – the superadditive effect – could have been due to the potential garden-path effect in (9).

Two other observations from Fukuda et al. (2022) are in order. First, in Fukuda et al. (2022), we pointed out that NP-scrambling is an optional syntactic operation, unlike obligatory A-bar dependencies such as *wh*-movement. Because NP-scrambling is optional, if it is not perceived as well-motivated by participants, by-hypothesis grammatical NP-scrambling could be rated poorly. A related observation is that in all the studies on NP-scrambling discussed above, one of the factors is SCRAMBLING, the presence vs. absence of NP-scrambling, while previous studies with obligatory A-bar dependencies manipulated the distance of the dependency, or DEPENDENCY LENGTH (e.g., *wh*-movement that originated in the matrix versus embedded clause). Because of this difference, acceptability judgment experiments with NP-scrambling might show a larger main effect of SCRAMBLING than the main

effect of DEPENDENCY LENGTH in previous studies with obligatory A-bar dependencies.[1] This point is important, as large main effects could make a floor effect more likely with superadditive interaction terms. If NP-scrambling incurs a large main effect, that could bring down the mean acceptability judgment of non-island items close to the lower bound of the scale, leaving no space for the mean of island structure items to go lower.

Second, the results of Fukuda et al. (2022) suggest that there are potentially significant individual differences in the size of the main effect of NP-scrambling in Japanese. Recall that the participants in Experiment 2 judged two items per condition. This allowed us to examine how consistent each participant's judgments were with respect to the two items of the same condition. There, we found that, while the largest group of participants gave predictably high ratings (i.e., positive *z*-scores) to NP-scrambling out of declarative CPs, a non-negligible number of participants gave the same items surprisingly low ratings (i.e., negative *z*-scores). Thus, the rating of NP-scrambling itself, in the absence of islands, is relatively variable in Japanese.

3.4 Section Summary

Our brief review of the five previous experimental studies on NP-scrambling out of three potential island structures, subjects, *because*-adjunct clauses, and relative clauses, identified two potential issues in these studies.

First, the *z*-score means of the non-island items in the scrambling condition were alarmingly low in Omaki et al. (2020) and Yano (2019). suggest that the effect of the mere presence of NP-scrambling was significant enough to considerably lower the acceptability of the non-island items in their experiments, the possibility discussed in Fukuda et al. (2022). This casts doubt on the conclusions in these studies, especially the claim in Omaki et al. (2020) that there is no subject-object asymmetry in NP-scrambling, as the non-significant interaction that the study found could have been due to a floor effect caused by a large main effect of NP-scrambling.

Second, the fact that the experiments in Fukuda et al. (2022) only examined subject relative clauses means that there is an alternative account for their findings: the possible garden-path effect in (9).

[1] In fact, Kluender and Kutas (1993) have shown that the mere presence of a long-distance dependency may cause a significant decrease in acceptability even with obligatory A-bar dependencies such as *wh*-movement.

4 Experiment 1: Reexamining Subject and Adjunct Islands

Experiment 1 was conducted to re-examine NP-scrambling out of subjects and *because*-adjunct clauses with the following goals. First, we wanted to test if an experiment with a larger sample size would still yield the same results that Omaki et al. (2020) and Yano (2019) obtained. To that end, we recruited ninety-three self-identified native speakers, which is a significantly larger sample than those of Omaki et al. (2020) (n = 53) and Yano (2019) (n = 42 and 47). Second, we wanted to improve the overall acceptability of experimental items. While we did not see much room for improvement with Omaki et al.'s (2020) materials, whose materials already took into consideration the relative weight of the relevant constituents, we thought that Yano's (2019) materials could be improved, as the scrambled constituents in Yano's (2019) experiments were bare NPs (e.g., *okashi* 'snack'). As such, we constructed our own materials for the *because*-adjunct clause subexperiment by controlling for the relative weight of the relevant constituents.

Experiment 1 had a 2 x 2 factorial design with SCRAMBLING (no-scrambling vs. scrambling) and STRUCTURE (non-island vs island). For each of the two potential island structures, eight lexicalizations were created with the four conditions (2 x 8 x 4 = 64). These sixty-four experimental sentences were distributed into eight lists, so that each list contained only one condition from each lexicalization group. The resulting 8 lists of 8 experimental sentences were then combined with different sets of 44 fillers (8 + 44 = 52). The experiment was administered online using IBEX (Drummond 2013), and participants were instructed to judge the naturalness of each sentence using a 7-point scale. Examples of the experimental items for the subject and adjunct subexperiments are shown as (10) and (11), respectively.

(10) Subject subexperiment

a. Non-island (complex NP object)/no-scrambling

Sono	seetoo-no	membaa-wa			
that	political.party-GEN	member-TOP			
[OBJ	aru	kookan-ga	iminhoo.ni	kansuru	
[	some	high.ranking.official-NOM	immigration	concerning	
juuyooshorui-o		nakushita-koto]-ni	too-no	riidaa-tachi-ga	
important.document-ACC		lost-koto]-DAT	party-GEN	leader-PL-NOM	
odoroita-to	katatta				
surprised$_{INT}$-C	said				

'That political party's member said that the party leaders were surprised that some high-ranking official lost important immigration documents.'

b. Island (complex NP subject)/no-scrambling
Sono seetoo-no membaa-wa
that political.party-GEN member-TOP
[SUBJ aru kookan-ga iminhoo.ni kansuru
[some high.ranking.official-NOM immigration concerning
juuyooshorui-o nakushita-koto]-ga too-no riidaa-tachi-o
important.document-ACC lost-koto]-NOM party-GENleader-PL-ACC
odorokaseta-to katatta
surprised$_{TR}$-C said
'That political party's member said that the fact that some high-ranking official lost important immigration documents surprised the party leaders.'

(11) *Because*-adjunct subexperiment
a. Non-island (declarative CP)/no-scrambling
Kinjo-no juumin-wa [CP shinai-no gasorinsutando-ga
neighbor-GEN resident-TOP [local-GEN gas.staion-NOM
isseeni gasorin-to keeyu-no nedan-o ageta]-to
at.once gasoline-and diesel.fuel-GEN price-ACC raised]-C
hara-o-tateta
got.angry
'The neighboring residents got angry that the local gas stations raised the prices of gasoline and diesel fuel at once.'
b. Island (*because*-adjunct clause)/no-scrambling
Kinjo-no juumin-wa [ADJ shinai-no gasorinsutando-ga
neighbor-GEN resident-TOP [local-GEN gas.staion-NOM
isseeni gasorin-to keeyu-no nedan-o ageta]-node
at.once gasoline-and diesel.fuel-GEN price-ACC raised]-because
hara-o-tateta
got.angry
'The neighboring residents got angry because the local gas stations raised the prices of gasoline and diesel fuel at once.'

Ninety-three university students in Japan participated (two excluded). The collected judgments were *z*-score transformed before analyzed with linear mixed effects models using R (Bates et al. 2015) with SCRAMBLING and STRUCTURE as fixed factors and participants and items as random factors. We calculated *p*-values using the lmerTest package (Kuznetsova et al. 2017), and also calculated Bayes factors for the interaction term for the fixed factors using the BayesFactor package (Morey and Rouder 2018). The Bayes factors reported here are of the BF_{10} type: they report the ratio of the likelihood of the data under the experimental hypothesis (H_1) that an interaction is present to the likelihood of the data under the null hypothesis (H_0) that there is no

interaction present. Following Jeffreys (1939/1961), a BF_{10} greater than 3 (rounded to a ceiling of 100) is interpreted as strong evidence that an interaction is present, and a BF_{10} less than 0.33 as strong evidence that there is no interaction. Bayes factors between 0.33 and 3 are interpreted as inconclusive.

4.1 Results

The lefthand panel of Figure 2 shows the four condition means in *z*-scores for the subject subexperiment, and the righthand panel shows the four condition means for the *because*-adjunct clause subexperiment.

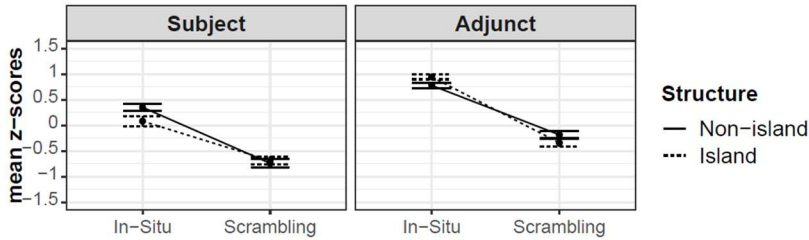

Figure 2. The mean *z*-scores from Experiment 1

Looking at the lefthand panel, the mean *z*-scores for the complex NP subject items and complex NP object items in the scrambling condition are virtually the same, with no indication of an interaction between SCRAMBLING and STRUCTURE. They are also located around −0.75, even lower than those of Omaki et al. (2020), which were around −0.25. The *z*-score means in the righthand panel for the *because*-adjunct clause subexperiment are higher than the *z*-score means in the lefthand panel for the subject subexperiment, suggesting that controlling for potential effects of the long-before-short preference might have improved their acceptability. However, the mean *z*-scores for the scrambling condition of the declarative CP items and the *because*-adjunct clause items are very close to each other, suggesting that there might not be an interaction between the two factors. Importantly, these *z*-score means are below zero, lower than the middle-of-the scale rating. The statistical models show that the interaction between SCRAMBLING and STRUCTURE was not significant within the subject subexperiment ($\beta = 0.32$, $p = 0.73$) and it was marginally significant within the *because*-adjunct subexperiment ($\beta = -0.33$, $p = 0.05$). The Bayes factor for the subject subexperiments is 0.98, within the inconclusive range, while that of the *because*-adjunct clause subexperiment is 3.02, suggesting that the interaction is likely to be present.

4.2 Discussion

At first glance, the null results of Experiment 1 may appear to support Omaki et al.'s (2020) claim that there is no subject-object asymmetry in NP-scrambling in Japanese. However, the results of the subject subexperiment leave the major issue in Omaki et al. unresolved, as the mean *z*-scores for the non-island complex NP object items in the scrambling condition were even lower than the mean *z*-scores of the similar items in Omaki et al. Within the *because*-adjunct clause subexperiment, although there is evidence of an interaction between the two factors, the superadditive effect is barely visible in Figure 2. As such, our results failed to provide clear evidence for island effects with NP-scrambling out of *because*-adjunct clauses. Finally, despite the overall improvement, the *z*-score means for the non-island items in the scrambling condition are below zero, lower than the average rating.

Now recall one of the observations from Fukuda et al. (2022) that we discussed in Section 3.3: the rating of NP-scrambling itself, in the absence of islands, is relatively variable in Japanese. This led us to investigate the possibility that the results of Experiment 1 might reflect acceptability judgments provided by participants with significantly different effect sizes with NP-scrambling. With some participants, the main effect of NP-scrambling might have been significant enough to cause a flooring effect, whereas the main effect of NP-scrambling might have been negligible with other participants. If that was the case, island effects would only be observed with the latter group, since no floor effect is expected with them.

5 Reexamining the Data from Experiment 1

To pursue the "two-group hypothesis" outlined in Section 4.2, we first calculated the effect size of NP-scrambling for each participant by subtracting the mean *z*-score for the non-island/scrambling condition items from the mean *z*-score for the non-island/no-scrambling condition items. Figure 3 shows the distribution of the individual effect sizes.

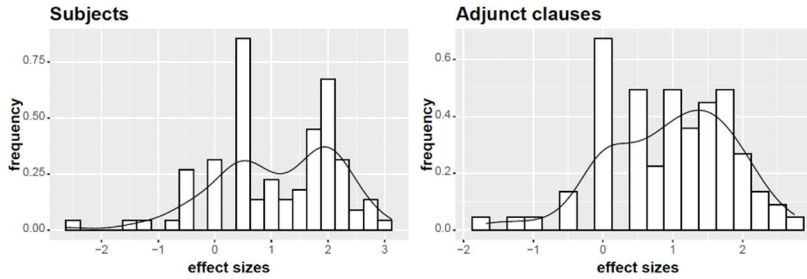

Figure 3. The distribution of the individual effect sizes in Experiment 1

The distribution of the effect sizes provides initial support for the two-group hypothesis. Within the subject subexperiment, the distribution of the effect sizes shows a bimodal distribution with two peaks, one at around the effect size of 0.5 and the other around 2.0. With the *because*-adjunct clause subexperiment, the distribution pattern is more complex, yet the participants seem to be divided into two groups: ones that belong to the most prominent peak at the effect size of 0 and the rest whose effect sizes are distributed between 0.5 and 2.0. Given these observations, we calculated the overall mean effect size of the entire group and divided the participants between two groups using the mean effect sizes: ones whose effect size is greater than the mean effect size and ones whose effect size is less than the mean effect size. For the sake of discussion, we call the former "non-scramblers" and the latter "scramblers". We then analyzed the data from each of the two subexperiments following the same procedure used to analyze the data in Experiment 1.

5.1 Subject Subexperiment

The lefthand panel in Figure 4 shows the *z*-score means for the four conditions for the non-scramblers, whose effect size of NP-scrambling is above average, and the righthand panel shows the *z*-score means for the scramblers, whose effect size of NP-scrambling is below average.

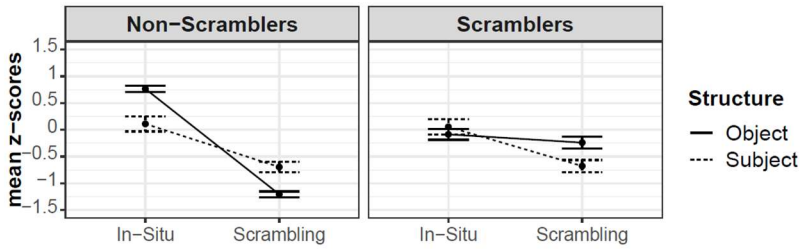

Figure 4. The mean *z*-scores from the subject subexperiment in Experiment 1 divided by GROUP

Looking at the lefthand panel with the non-scramblers, the effect of NP-scrambling is particularly pronounced with the complex NP object items. In a clear contrast, in the righthand panel with the scramblers, the effect of NP-scrambling is negligible within the complex NP object items, while it is more pronounced with the complex NP subject items, suggesting that there may be an interaction between STRUCTURE and SCRAMBLING.

The statistical models reveal that, within the scramblers ($n = 43$), the interaction between STRUCTURE and SCRAMBLING is marginally significant ($\beta = -0.59$, $p = 0.06$). The Bayes factor for the scramblers is 3.08, suggesting

that the interaction is likely to be present. Within the non-scramblers ($n = 46$), the interaction turns out to be also significant ($\beta = 1.15$, $p < 0.01$) and the Bayes factors is significantly above 3.0 (BF > 100). However, the direction of the interaction is opposite, as the *z*-score mean for the non-island complex NP object item in the scrambling condition is lower than the *z*-score mean of the island complex NP subject item in the scrambling condition.

5.2 *Because*-adjunct Clause Subexperiment

Figure 5 shows the *z*-score means for the four conditions for the non-scramblers and the scramblers from the *because*-adjunct clause subexperiment.

Figure 5. The mean *z*-scores from the *because*-adjunct clause subexperiment in Experiment 1 divided by GROUP

The distribution of the four *z*-score means on the lefthand panel for the non-scramblers ($n = 48$) indicates a large effect of NP-scrambling across the non-island and island conditions, with the mean for the non-island items in the scrambling condition numerically lower than that of the island items in the same condition. The mean *z*-scores on the righthand panel, ones from the scramblers ($n = 41$), show a very different distribution. Similar to what we observed with the scramblers in the subject subexperiment, the main effect of NP-scrambling is negligible with the non-island items, while the *z*-score means sharply decrease between the no-scrambling condition and the scrambling condition with the *because*-adjunct clause items.

The statistical models confirm the above visual inspection: there is a significant interaction between STRUCTURE and SCRAMBLING within the scramblers ($\beta = -1.10$, $p < 0.01$) but not within the non-scramblers ($\beta = 0.31$, $p = 0.11$). The Bayes factor of the scramblers is significantly higher than 3.0 (BF > 100), but that of the non-scramblers is 2.05, within the inconclusive range.

5.3 Interim Conclusion

When looking at the overall result, the results of Experiment 1 with both subject and *because*-adjunct clause subexperiments failed to show clear evidence of island effects, and they also suffered from the same issue that Omaki et al.

(2020) and Yano (2019) suffered - i.e., alarmingly low *z*-score means for the non-island items in the scrambling condition. However, once individual differences in the effect size of NP-scrambling were taken into consideration, the results of Experiment 1 turned out to reveal clear evidence for island effects with *because*-adjunct clauses and suggestive evidence for island effects with subjects, but only among the scramblers, or the participants whose judgments were less affected by the mere presence of NP-scrambling.

6 Experiment 2: Reexamining Relative Clause Islands

Experiment 2 was designed to address a potential confounding factor in the experiments in Fukuda et al. (2022). As discussed in Section 2.3, the fact that only subject relative clauses were examined in Fukuda et al. (2022) means that there is an alternative account for the superadditive effect found in the study: the possible garden-path effect (9). Addressing this confounding factor is particularly important, since the original observation that NP-scrambling out of relative clauses in Japanese incurs island effects was based on examples with subject relative clauses such as (5a-b).

Experiment 2 addresses the potential garden-path effect in (9) by examining NP-scrambling out of both subject and object relative clauses with ditransitive verbs. The following schematic examples in (12) show the assumed underlying structures of subject relative clause items.

(12) a. Non-island (declarative CP)/scrambling
[NP ....]-DAT$_1$ NP-TOP [$_{CP}$ NP-NOM ***t*$_1$** NP-ACC V_{EMB}]-C V_{MAT}

b. Island (subject relative clauses)/scrambling
[NP ....]-DAT$_1$ NP-TOP [[$_{RC}$ e$_2$ ***t*$_1$** NP-ACC V_{EMB}] head NP_2]-ACC V_{MAT}

Example (12b) has the same issue that we discussed with (9): It is likely to incur a garden-path effect. However, the situation is different with object relative clause items, which are schematically represented in (13).

(13) a. Non-island (declarative CP)/scrambling
[NP ....]-DAT$_1$ NP-TOP [$_{CP}$ NP-NOM ***t*$_1$** NP-ACC V_{EMB}]-C V_{MAT}

b. Island (object relative clauses)/scrambling
[NP ....]-DAT$_1$ NP-TOP [[$_{RC}$ NP-NOM ***t*$_1$** e$_2$ V_{EMB}] head NP_2]-ACC V_{MAT}

The crucial difference between (12) and (13) is that the object relative clause in (13) has an overt embedded subject. The presence of the overt subject reliably signals the presence of the embedded clause (e.g., Miyamoto 2002), making a garden-path effect unlikely. Thus, if the garden-path effect caused the superadditive effect observed in Fukuda et al. (2022), the effect would be

observed only with the subject relative clause items in (12) and should be absent from the object relative clause items in (13). If a superadditive effect is observed with both subject and object relative clauses, then it is unlikely to be due to the garden-path effect and more likely to be an island effect.

Experiment 2 had two subexperiments: the subject relative clause subexperiment and the object relative clause subexperiment. Each of the subexperiment had a 2 x 2 factorial design with SCRAMBLING (no scrambling vs. scrambling) and STRUCTURE (declarative CPs vs. relative clauses). Eight lexicalizations were created for each of the four conditions for the two types of relative clauses (4 x 8 x 2 = 64). (14a-b) are examples of the subject relative clause items, whereas (15a-b) are examples of the object relative clause items. The scrambled constituents are outlined with a box.

(14) a. Non-island (declarative CP)/no-scrambling
Sushiya-no shujin-wa [$_{CP}$ jooren-no okyakusan-ga
Sushi.restaurant-GEN owner-TOP [regular-GEN customer-NOM
jimoto-no shinbunkisha-ni mise-no koto-o
local-GEN newspaper.reporter-DAT restaurant-GEN thing-ACC
sendenshitekureta]-to hometa
advertised]-C praised
'The owner of the sushi restaurant praised that the regular customer advertised the restaurant to the local newspaper reporter.'

b. Island (subject relative clauses)/no-scrambling
Sushiya-no shujin-wa [$_{RC}$ e$_1$ jimoto-no
Sushi.restaurant-GEN owner-TOP [local-GEN
shinbunkisha-ni mise-no koto-o
newspaper.reporter-DAT restaurant-GEN thing-ACC
sendenshitekureta] jooren-no okyakusan$_1$-o hometa
advertised] regular-GEN customer-ACC praised
'The owner of the sushi restaurant praised the regular customer who advertised the restaurant to the local newspaper reporter.'

(15) a. Non-island (declarative CP)/no-scrambling
Kyuukyuubyootoo-no kangohu-wa[$_{CP}$ shinjin-no ishi-ga
Emergency.room-GEN nurse-TOP [new-GEN doctor-NOM
ishikihumee-no kanja-ni nishurui-no kusuri-o
unconcsious-GEN patient-DAt two.kinds-GEN drug-ACC
tooyoshita]-to tantooi-ni tsutaeta
administered]-C attending.physician-to reported
'The emergency room nurse reported to the attending physician that the new doctor administered two types of drugs to the unconscious patient.'

b. Island (object relative clause)/no-scrambling
Kyuukyuubyootoo-no kangohu-wa[RP shinjin-no ishi-ga
Emergency.room-GEN nurse-TOP [new-GEN doctor-NOM
ishikihumee-no kanja-ni e_1 tooyoshita] nishurui-no kusuri$_1$-o
unconcsious-GEN patient-DAT administered] two.kinds-GEN drug-ACC
tantooi-ni miseta
attending.physician-to showed
'The emergency room nurse showed to the attending physician two types of drugs that the new doctor administered to the unconscious patient.'

The experimental items were distributed into four lists and combined with forty-eight fillers and five practice sentences (16 + 48 + 5 = 69). A total of 100 self-identified Japanese native speakers were recruited via a Japanese crowdsourcing website, CrowdWorks (https://crowdworks.co.jp/en/), and participated in the experiment online using PCIbex (Zehr and Schwarz 2018).

The effect size of NP-scrambling for each participant was calculated using the same procedure used in Experiment 1. Figure 6 shows the distribution of the individual effect sizes in Experiment 2 for the two subexperiments.

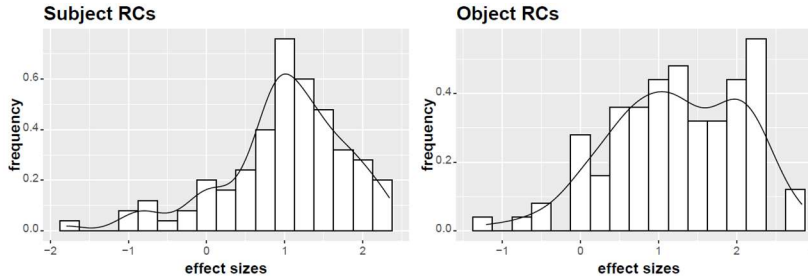

Figure 6. The distribution of the individual effect sizes in Experiment 2

Within the object relative clause subexperiment on the righthand panel, the distribution of the effect sizes shows a bimodal distribution with two peaks, one around the effect size of 1.0 and the other around 2.5. In contrast, there is no indication of a bimodal distribution in the effect sizes with the subject relative clause subexperiment on the lefthand panel, with a single peak at the effect size of 1.0. Although the visual inspection provides no evidence of a bimodal distribution within the subject relative clause subexperiment, we divided the participants in Experiment 2 into two groups, scramblers and non-scramblers, using the same procedure described in Section 4, to be consistent.

6.1 Subject Relative Clauses

The lefthand panel in Figure 7 show the four condition means for the non-scramblers with above-average effect size of NP-scrambling, and the

righthand panel shows the condition means for the scramblers with below-average effect size of NP-scrambling.

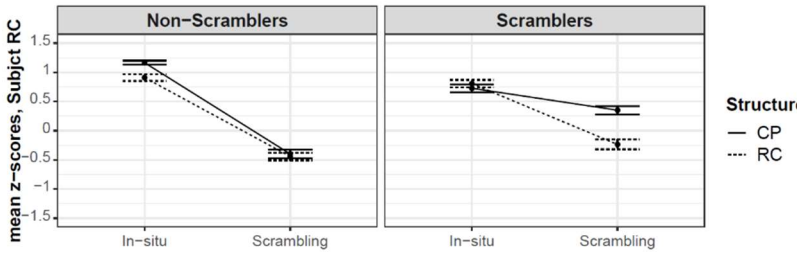

Figure 7. The mean *z*-scores from the subject relative clause subexperiment

With the non-scramblers (n = 52), the *z*-score means for the non-island declarative CP items and the island subject relative clause items with the scrambling condition virtually overlap, showing that the significant main effect of NP-scrambling. With the scramblers (n = 48), the main effect of NP-scrambling with the non-island declarative CPs is negligible while the *z*-score mean for the subject relative clause items with the scrambling condition is lower than the *z*-score mean for the non-island declarative CP items with the same condition, indicating the presence of a superadditive effect.

The statistical models support the observations above. There is a significant interaction between SCRAMBLING and STRUCTURE with the scramblers ($\beta = -0.65$, $p < 0.01$) but not with the non-scramblers ($\beta = 0.19$, $p = 0.10$). The Bayes factor for the scramblers is significantly above 3.0 (BF > 100), providing further support for the presence of the interaction, whereas the Bayes factor for the non-scramblers is 0.66, within the inconclusive range.

6.2 Object Relative Clauses

The lefthand panel in Figure 8 shows the four condition means for the non-scramblers and the righthand panel shows those for the scramblers.

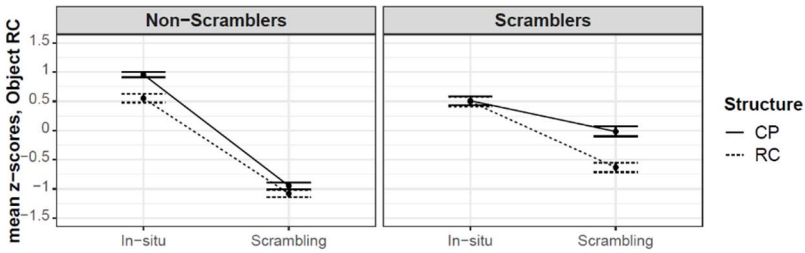

Figure 8. The mean *z*-scores from the object relative clause subexperiment

With the non-scramblers (n = 51), the z-score means for the non-island CP items and the island object relative clauses items with the scrambling condition were very close to each other at closer to the bottom of the scale, showing the indiscriminating effect of NP-scrambling. With the scramblers (n = 49), the effect of NP-scrambling is clearly larger with the island object relative clause items than with the non-island declarative CP items, creating a larger gap between the z-score means on the scrambling side compared to the gap on the no-scrambling side, where the two z-score means completely overlap.

The statistical model indicates a significant interaction between STRUCTURE and SCRAMBLING with the scramblers ($\beta = -0.54$, $p < 0.01$) and the Bayes factor is significantly above 3.0 (BF > 100). With the non-scramblers, there is no significant interaction ($\beta = 0.19$, $p = 0.09$) and the Bayes factor for the scramblers is 1.62, within the inconclusive range.

6.3 Discussion

Just like Experiment 1, the participants in Experiment 2 were divided into scramblers, with whom the effect of scrambling is below the average, and non-scramblers, with whom the effect of scrambling is above the average. With the scramblers, robust island effects of NP-scrambling are observed with both subject and object relative clauses, suggesting that the effects of NP-scrambling out of relative clauses cannot be reduced to the potential garden-path effect or the effects of the overlapping dependencies.

7 Conclusions and Implications

The aim of this study was to reexamine island effects with NP-scrambling in Japanese with three potential island structures: subjects, *because*-adjunct clauses, and relative clauses. Previous studies disagree whether or not there is subject-object asymmetry in NP-scrambling (Jurka 2010; Jurka et al. 2011; Omaki et al. 2020). While Yano (2019) provides experimental evidence for island effects with NP-scrambling out of *because*-adjunct clauses in Japanese, the fact that NP-scrambling out of the non-island declarative CPs was rated surprisingly low in his experiments raises some concern, and the same issue is observed with the experiments in Omaki et al., where the items in the scrambling condition generally received low ratings. In Fukuda et al. (2022), we presented experimental evidence for island effects with NP-scrambling out of relative clauses, but the fact that only subject relative clauses were tested in our experiments leaves open the possibility that the observed effect receives an alternative account. Finally, the second experiment in Fukuda et al. revealed that acceptability judgments of Japanese speakers on by-hypothesis grammatical NP-scrambling out of declarative CPs may vary.

We conducted two acceptability judgment experiments to test island effects of NP-scrambling with subjects and *because*-adjunct clauses (Experiment 1) and subject and object relative clauses (Experiment 2) in Japanese. The results of these experiments show significant individual differences in the effect size of NP-scrambling. Once these individual differences are taken into consideration, our findings provide clear evidence of island effects with NP-scrambling out of *because*-adjunct clauses and relative clauses and suggestive evidence of island effects with NP-scrambling out of subjects.

Island effects with NP-scrambling out of relative clauses have been reliably replicated experimentally. Our contribution is that we have demonstrated the presence of island effects with both subject and object relative clauses. Given that the previous discussions focused only on NP-scrambling out of subject relative clauses, our findings provide novel empirical evidence from object relative clauses and strengthen the claim that Japanese relative clauses are islands with respect to NP-scrambling. It is interesting to note that relative clauses have also been found to incur (subliminal) island effects with *wh*-in-situ (Tanaka and Schwartz 2018; Tanaka this volume) and relativization (Takahashi and Goodall 2021) in Japanese.

Our findings also provide clearer evidence for island effects with NP-scrambling out of *because*-adjunct clauses and suggest that the lower-than-expected *z*-score means for the non-island declarative CP items with NP-scrambling in Yano (2019) could have been due to individual differences in the effect size of NP-scrambling. As the next step, it would be informative to examine island effects of NP-scrambling with different types of adjunct clauses in Japanese. For instance, conditional adjuncts with *moshi* 'if' have been claimed to be a non-island structure (Yoshida 2006). It would be interesting to compare NP-scrambling out of *because*-adjunct clauses and that out of the conditional adjunct clauses.

Finally, Japanese has been argued to lack subject-object asymmetry with subextraction (e.g., Ross 1967; Kuno 1973; Saito 1985, 1992; Nishigauchi 1990; Lasnik and Saito 1992; Watanabe 1992; Takahashi 1994; Ishii 1997; Richards 1997; Stepanov 2007). Given that, the fact that we obtained a superadditive effect with NP-scrambling out of subjects is noteworthy.

Another major empirical contribution of the current study is that we have shown that there are significant individual differences in the effect size of NP-scrambling. The obvious question that this finding raises is: Why is there so much individual variation in the effect size of NP-scrambling? Answering this question requires a careful investigation that identifies and systematically test multiple factors that may potentially affect speakers' judgements of sentences with NP-scrambling. However, as discussed in Section 3.3, one notable characteristic of NP-scrambling that makes it different from the other A-bar dependencies is that it is an optional operation. Because it is optional, if

NP-scrambling in a given sentence is not perceived as well-motivated by participants, such sentence might be rated poorly. If this is a contributing factor to the observed individual differences in the effect size of NP-scrambling, providing contexts might reduce the effect size of NP-scrambling with some speakers. In fact, Koizumi and Imamura (2017) demonstrated that OSV sentences are processed faster when context sentences were provided so that the scrambled object represented old/given information.[2]

We offer two comments about possible effects of contexts on the effect size of NP-scrambling. First, Yano (2019) provided contexts in one of his experiments with NP-scrambling out of *because*-adjunct clauses and noun complements, but the presence of these contexts failed to improve the acceptability of his experimental sentences. Second, while the findings from Koizumi and Imamura (2017) suggest that scrambled constituents typically represent old/given information, as briefly discussed in Section 3.1, previous studies have also shown that NP-scrambling tends to be judged as more acceptable if the scrambled constituent is relatively heavy/long in a given sentence, because of the long-before-short preference among Japanese speakers. These two observations conflict with each other. While heavier/longer constituents are preferred to be placed sentence initially, thus motivating scrambling, scrambled constituents also tend to represent given information, which tend to be shorter (e.g., Arnold et al. 2000). Given these considerations, it would not be a simple task to provide contexts to see if they can improve the acceptability of sentences with NP-scrambling.

While we must leave this question unanswered for now, our findings in this study demonstrate how large-scale acceptability judgment experiments can shed new light on previously unnoticed factors that affect acceptability judgments with different types of long-distance dependencies. We hope that our findings will help improve the future experimental investigations into island effects with NP-scrambling and other A-bar dependencies.

References

Arnold, J. E., A. Losongco, T. Wasow, and R. Ginstrom. 2000. Heaviness vs. Newness: The Effects of Structural Complexity and Discourse Status on Constituent Ordering. *Language* 76.1: 28–55.

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67.1: 1–48.

Cowart, W. 1997. *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. Thousand Oaks, CA: Sage Publications.

Drummond, A. 2013. Ibex Farm. https://spellout.net/ibexfarm/.

[2] Otsu (1994) also demonstrated that Japanese children's comprehension of scrambling sentences greatly improved when contexts that introduce relevant referents were provided. I would like to thank John Whitman and Kyoko Yamakoshi for directing our attention to this study.

Dryer, M. S. 1980. The Positional Tendencies of Sentential Noun Phrases in Universal Grammar. *The Canadian Journal of Linguistics* 25: 123–95.

Fukuda, S., N. Tanaka, H. Ono, and J. Sprouse. 2022. An Experimental Reassessment of Complex NP Islands with NP-Scrambling in Japanese. *Glossa: A Journal of General Linguistics* 7.1: 1–35.

Haig, J. H. 1976. Shadow Pronoun Deletion in Japanese. *Linguistic Inquiry* 7: 363–71.

Hawkins, J. A. 1999. Processing Complexity and Filler-Gap Dependencies across Grammars. *Language* 75: 244–85.

Ishii, T. 1997. An Asymmetry in the Composition of Phrase Structure and Its Consequences. Doctoral dissertation, University of California, Irvine.

Jeffreys, H. 1939/1961. *Theory of Probability*. Oxford, UK: Oxford University Press.

Jurka, J. 2010. The Importance of Being a Complement: The CED Effects Revisited. Doctoral dissertation, University of Maryland, College Park.

Jurka, J., C. Nakao, and A. Omaki. 2011. It's Not the End of the CED as We Know It: Revisiting German and Japanese Subject Islands. *Proceedings of the 28th West Coast Conference on Formal Linguistics*, ed. M. B. Washburn, K. McKinney-Bock, E. Varis, A. Sawyer, and B. Tomaszewicz, 124–32. Somerville, MA: Cascadilla Press.

Kluender, R. and M. Kutas. 1993. Subjacency as a Processing Phenomenon. *Language and Cognitive Processes* 8: 573–633.

Koizumi, M. and S. Imamura. 2017. Interaction between Syntactic Structure and Information Structure in the Processing of a Head-Final Language. *Journal of Psycholinguistic Research* 46: 247–60.

Kuno, S. 1973. *The Structure of the Japanese Language*. Cambridge, MA: MIT Press.

Kuznetsova, A., P. B. Brockhoff, and R. H. B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82.13: 1–26.

Langsford, S., A. Perfors, A. T. Hendrickson, L. A. Kennedy and D. J. Navarro. 2018. Quantifying Sentence Acceptability Measures: Reliability, Bias, and Variability. *Glossa: A Journal of General Linguistics* 3.1: 37. 1–34.

Lasnik, H. and M. Saito. 1992. *Move α: Conditions on Its Application and Output*. Cambridge, MA: MIT Press.

Miyamoto, E. T. 2002. Case Markers as Clause Boundary Inducers in Japanese. *Journal of Psycholinguistic Research* 31.4: 307–47.

Morey, R. D. and J. N. Rouder. 2018. BayesFactor: Computation of Bayes Factors for Common Designs. R package version 0.9.12. https://CRAN.R-project.org/package=BayesFactor

Nishigauchi, T. 1990. *Quantification in the Theory of Grammar*. Dordrecht, Netherlands: Kluwer Academic Publishers.

Omaki, A., S. Fukuda, C. Nakao, and M. Polinsky. 2020. Subextraction in Japanese and Subject-Object Symmetry. *Natural Language and Linguistic Theory* 38: 627–69.

Otsu, Y. 1994. Early Acquisition of Scrambling in Japanese. *Language Acquisition Studies in Generative Grammar: Papers in Honor of Kenneth Wexler from the 1991 GLOW Workshop*, ed. T. Hoekstra and B. D. Schwartz, 253–64. Amsterdam: John Benjamins.

Richards, N. 1997. What Moves Where When in Which Languages? Doctoral dissertation, MIT.

Ross, J. R. 1967. Constraints on Variables in Syntax. Doctoral dissertation, MIT.
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.
Schütze, C. T. and J. Sprouse. 2014. Judgment Data. *Research Methods in Linguistics*, ed. R. J. Podesva and D. Sharma, 27–50. Cambridge, UK: Cambridge University Press.
Sprouse, J. 2007. A Program for Experimental Syntax: Finding the Relationship between Acceptability and Grammatical Knowledge. Doctoral dissertation, University of Maryland, College Park.
Sprouse, J., I. Caponigro, C. Greco, and C. Cecchetto. 2016. Experimental Syntax and the Variation of Island Effects in English and Italian. *National Language and Linguistic Theory* 34: 307–44.
Sprouse, J., S. Fukuda, H. Ono, and R. Kluender. 2011. Reverse Island Effects and the Backward Search for a Licensor in Multiple *Wh*-questions. *Syntax* 14.2: 179–203.
Sprouse, J., M. Wagers, and C. Phillips. 2012. A Test of the Relation between Working-Memory Capacity and Syntactic Island Effects. *Language* 88.1: 82–123.
Stepanov, A. 2007. The End of the CED? Minimalism and Extraction Domains. *Syntax* 10: 80–126.
Takahashi, D. 1994. Minimality of Movement. Doctoral dissertation, University of Connecticut.
Takahashi, M. and G. Goodall. 2021. Gap-Filler Dependencies Are Sensitive to Islands: The Case of Japanese Relative Clauses. The 34th CUNY Conference on Human Sentence Processing, 4–6 March, University of Pennsylvania.
Tanaka, N. This volume. Reexamining Island Effects in Japanese Complex NP with Argument *Wh*-in-situ. *Japanese/Korean Linguistics* Vol. 30, ed. S. Williamson, A. A. Babayode-Lawal, N. Chan, S. Cho, and I. Fong. Stanford, CA: CSLI Publications.
Tanaka, N. and B. D. Schwartz. 2018. Investigating Relative Clause Island Effects in Native and Nonnative Adult Speakers of Japanese. *Proceedings of the 42nd Annual Boston University Conference on Language Development*, ed. A. B. Bertolini and M. J. Kaplan, 750–63. Somerville, MA: Cascadilla Press.
Watanabe, A. 1992. Subjacency and S-Structure Movement of *Wh*-in-situ. *Journal of East Asian Linguistics* 1: 255–91.
Yamashita, H. and F. Chang. 2001. 'Long before Short' Preference in the Production of a Head-Final Language. *Cognition* 81: B45–B55.
Yano, M. 2019. On the Nature of the Discourse Effect on Extraction in Japanese. *Glossa: A Journal of General Linguistics* 4.1: 90. 1–17.
Yoshida, M. 2006. Constraints and Mechanisms in Long-Distance Dependency Formation. Doctoral dissertation, University of Maryland, College Park.
Zehr, J. and F. Schwarz. 2018. PennController for Internet Based Experiments. (IBEX). https://doi.org/10.17605/OSF.IO/MD832