An experimental reassessment of complex NP islands with NP-scrambling in Japanese

Shin Fukuda, University of Hawai‘i at Manoa
Nozomi Tanaka, Indiana University
Hajime Ono, Tsuda University
Jon Sprouse, University of Connecticut

Abstract

There is little consensus in the Japanese syntax literature on the question of whether complex NPs with a complement clause headed by to yuu ‘that say’ are islands for NP-scrambling dependencies. To explore this question, we conducted two acceptability judgment experiments using the factorial definition of islands to test the island status of noun complements, relative clauses (which are complex NPs, but uniformly considered islands in the literature), and coordinated NP structures (which are also uniformly considered islands in the literature). Our first experiment yielded clear evidence that relative clauses and coordinated NPs are islands, and that noun complements are not. Our second experiment replicated the relative clause and coordinated NP results, but yielded an inconclusive null result for noun complements. Taken together, our results suggest either that noun complements are not islands, or that noun complements yield a small island effect that cannot be reliably detected at the typical sample sizes of 30-40 participants used here. We also investigated between- and within-participant variability in our results. We observe no evidence of increased between-participant variability for noun complements relative to other islands, and no increase of within-participant variability for noun complements relative to grammatical NP-scrambling, thus corroborating our conclusions. Our results have consequences for a number of issues that have been encoded in current syntactic theories of island effects, including the correlation between syntactic constituent complexity and island status (e.g., number of bounding nodes or phase heads), and the correlation between complementizer deletion and island status (e.g., the complement/adjunct distinction).

Key words: island constraints, scrambling, noun complements, relative clauses, experimental syntax, Japanese

1. Introduction

There is considerable debate in the literature about the island status of complex NPs in Japanese NP-scrambling constructions. For example, Haig (1976) claims that complex NPs with a complement clause headed by to yuu ‘that say’ (henceforth, noun complements) are not islands for NP-scrambling, whereas relative clauses are islands for NP-scrambling. In contrast, Saito (1985) claims that both noun complements and relative clauses are islands for NP-scrambling, but that the island effect of noun complements is smaller than the island effect of relative clauses. Recent experimental work has only seemed to add to this debate. Yano (2019), as part of a broader study of the effect of D-linking on islands in Japanese, tested noun complements (but
not relative clauses) in two acceptability judgment experiments using the factorial definition of island effects. However, the two experiments investigating noun complements with non-D-linked NPs produced contradictory results: the first experiment revealed a (relatively small) island effect, whereas the second experiment revealed no island effect (see section 2 for additional discussion). This suggests a need for additional systematic data collection with both noun complements and relative clauses. Therefore, in this study, we present two additional judgment studies specifically designed to explore complex NP islands with NP-scrambling in Japanese (Experiments 1 and 2). We use the factorial definition of island effects to explore the status of NP-scrambling out of both noun complements and relative clauses, and for additional comparison of the size of the island effects (following Saito’s 1985 suggestion), we also include NP-scrambling out of a coordinate structure, which is uncontroversially considered an island for NP-scrambling in Japanese (Harada 1977).

Our results provide support for previous studies in that relative clauses as well as coordinate structures are shown to yield large island effects with NP-scrambling in Japanese. However, our results for noun complements are less straightforward: while Experiment 1 yielded conclusive evidence that they are not islands, Experiment 2 yielded inconclusive results. These findings suggest either that noun complements in Japanese are not islands or that noun complements in Japanese yield island effects that are too small to reliably detect at these sample sizes and are therefore qualitatively distinct from relative clause island effects, which are easy to detect at these sample sizes. Thus, the results of our experiments show that there is a clear difference between relative clauses and noun complements, at least in Japanese, and more broadly, between noun complements in Japanese and noun complements in other languages that have been tested using the factorial definition, such as English, Italian, and Norwegian (cf. Sprouse et al. 2011; Sprouse et al. 2012; Sprouse et al. 2016; Kush et al. 2018). We argue that our findings have direct consequences for most existing theories of island effects, for theories of the relationship between complementizer deletion and island status, and for theories of the relationship between the complexity of syntactic structure and island status, and we suggest that future studies should probe the properties of relative clauses and noun complements (cross-linguistically) along these dimensions.

2. Complex NPs and NP-scrambling in Japanese

In this section we provide a brief review of previous claims about complex NP islands with NP-scrambling in Japanese. Haig (1976) was one of the first theoretical studies to investigate complex NPs in Japanese, reporting that NP-scrambling out of a noun complement is acceptable (1a), while NP-scrambling out of a relative clause is not (1b). The judgments in (1) are from Haig (1976).


have
‘Mary, I have a desire such that I want to introduce her to Bill.’ (Haig 1976: 369; (25))
Saito (1985; 1987) made the more nuanced claim that both noun complements and relative clauses are islands, with noun complements being relatively more acceptable (2a) than relative clauses (2b). The judgments in (2) are from Saito (1985).

(2) a.  ?Bill-o John-ga [CNP Mary-ga ti saketeiru-to.yuu uwasa]-o kiita. B-ACC J-NOM [ M-NOM avoid-that.say rumor]-ACC heard
  ‘John heard a rumor that Mary is avoiding Bill’ (Saito 1985: 246; (146b))

  ‘It seems that John is looking for the person who bought that book.’
  (Saito 1985: 246; (146a))

To the best of our knowledge, Yano (2019) is the first and only study to examine the acceptability of NP-scrambling out of complex NPs (specifically, noun complements) with formal acceptability judgment experiments. The goal of Yano (2019) was to examine whether D-linked NPs like *sono shoosetsu ‘that novel’ undergo syntactic movement when they appear in a fronted position. Yano (2019) uses island effects as a diagnostic of movement. To that end, Yano tested two island types: adjunct islands and noun complement islands. Yano (2019) tested both D-linked NPs (with *sono ‘that’) as the target of investigation, and non-D-linked NPs (without *sono ‘that’) as a baseline comparison. Here we focus exclusively on non-D-linked NPs as the effect of D-linked phrases, or lack thereof, is typically considered a separate topic of investigation (see Szabolcsi & Lohndal 2017 for a review of selective islands).

In the first experiment of Yano (2019), the sentences were presented in isolation. In the second experiment, the sentences were presented with a context sentence such as “The novel received the Naoki prize.” to establish the fronted object in the discourse. Yano (2019) used the factorial definition of island effects in which the presence of an island effect appears as a superadditive interaction of two (or more) factors that are themselves independent of the island effects (Sprouse 2007; Sprouse et al. 2011; Sprouse et al. 2012, a.o). For the Yano (2019) experiments these factors were STRUCTURE, manipulating the structure of the embedded clause (either an island or a non-island), and WORD ORDER, manipulating the presence or absence of scrambling out of the embedded clause. An example of the island/scrambling condition is given in (3) for completeness; we review the logic of the factorial design, and provide full examples for our experiments, in Section 3.

(3) Shoosetsu-o [CNP hyooronka-wa kyonen goosutoraitaa-ga ti kaita-toiu novel-ACC commentator-TOP last.year ghost.writer-NOM wrote-that.say hoddoo]-o shinjiteuru.
  news]-ACC believe
  ‘The commentator believes the news that the ghost-writer wrote the novel.’
  (Yano 2009: 5; (9d))

---

1 The Naoki prize is a prestigious Japanese literary award.
The results of the two Yano (2019) experiments are equivocal. In the first experiment (no context), Yano found a small superadditive interaction indicative of a noun complement island effect. In the second experiment (with context), Yano found no superadditive interaction indicative of a noun complement island effect. Similar to the disagreement between Haig (1976) and Saito (1985; 1987), this leaves the status of noun complements under debate. One complicating issue is that the Yano (2019) results also showed very low acceptability even for grammatical scrambling out of non-island embedded clauses (a declarative CP). Yano notes that this could be due to a preference in Japanese that scrambled NPs be longer than the NPs that they are scrambled over; the long-before-short preference (Dryer 1980; Hawkins 1994; Yamashita & Chang 2001; Yamashita 2002; Omaki et al. 2020). The scrambled NPs in the Yano (2019) experiments are single word NPs, which could have pushed the acceptability down. This in turn could have reduced the size of the superadditive interactions (if the long-before-short preference is not additive with island effects, which is itself a potentially interesting observation that might merit future study).

The contradictory results for noun complements between Haig (1976) and Saito (1985; 1987), and between the two experiments in Yano (2019), suggest that additional systematic data collection is needed. To that end, here we report the results of two formal acceptability judgment experiments. We describe the logic and design of these experiments in the next section.

3. The design of the experiments

The logical design of our study is as follows. Both experiments include three island types: noun complements, relative clauses, and coordinated NP structures. By including both types of complex NPs together in the same experiment, we can investigate the Saito’s (1985; 1987) claim that noun complements yield smaller island effects than relative clauses. We included coordinated NP structures because they are uncontroversially considered islands in the literature (Harada 1977), and therefore could serve as a type of baseline comparison for the complex NPs. We use the factorial definition of island effects, both because we believe it matches the logic that has historically been used by syntacticians to define island effects, and because it allows us to eventually integrate our results with the growing cross-linguistic experimental literature using the factorial definition (a.o., Christensen et al. 2013; Almeida 2014; Kim & Goodall 2016; Sprouse et al. 2016; Kush et al. 2018; 2019; Stepanov et al. 2018; Ko et al. 2019; Lu et al. 2019; Tucker et al. 2019; Omaki et al. 2020). As described below, we implement the factorial design completely within participants, allowing us to quantify to what extent each participant reports an island effect, so that we can investigate a conjecture motivated by discussion in Yano (2019) that noun complements may show a high degree of variability across participants (see Section 4.2 for discussion). Experiment 2 also includes two tokens per condition per participants, so we can additionally investigate within-participant variability for each island. Finally, we use relatively long scrambled NPs to satisfy the long-before-short preference.

3.1 The factorial definition of island effects

The factorial design has two factors: SCRAMBLING manipulates the presence or absence of NP-scrambling (no-scrambling/scrambling), and STRUCTURE manipulates the structure of the embedded clause (non-island/island). Fully crossing these two factors in a 2×2 design leads to
four conditions. In (4), we illustrate all four conditions for noun complements. Note that the NPs that are the target of scrambling are outlined with a box.

(4) Example conditions for noun complements
a. non-island / no-scrambling
Shinjin-no kisha-wa [CP IT-gaisha-no shachoo-ga
novice-GEN reporter-TOP [CP IT-company-GEN CEO-NOM
[chuumoku-no kakkitekina sofutowea-o] daigaku.zaigakuchuu-ni kaihatsushita-to]
popular-GEN innovative software-ACC college.days-in developed-COMP]
kiji-ni kaita.
article-as wrote
‘That novice reporter wrote (as an article) that the CEO of the IT company developed the popular, innovative software while s/he was in college.’

b. non-island / scrambling
[Chuumoku-no kakkitekina sofutowea-o] shinjin-no kisha-wa
popular-GEN innovative software-ACC novice-GEN reporter-TOP
[CP IT-gaisha-no shachoo-ga daigaku.zaigakuchuu-ni kaihatsushita-to]
[CP IT-company-GEN CEO-NOM college.days-in developed-COMP]
kiji-ni kaita.
article-as wrote

For relative clauses (5) and coordinated NP structures (6), we report only the no-scrambling conditions because the scrambling conditions can be constructed by simply fronting the boxed NP. This allows us to save space while simultaneously focusing on the less obvious manipulation of structure. The non-island structure that we chose for relative clauses was a declarative CP. The non-island structure that we chose for coordinated NP structures was an NP-PP sequence.
(5) Example conditions for relative clauses (only no-scrambling examples shown)

a. non-island / 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 han’ee saseteinai-to}
drastic sales-GEN growth-ACC employee-GEN salary-to reflect not.make-COMP

hihanshita.
criticized

‘The union leader criticized that the executives of the company were not making the drastic sales growth reflected in the employees’ salaries.’

b. island / no-scrambling

Roodookumiai-no riidaa-wa
union-GEN leader-TOP

[RC oohabana uriage-no nobi-o juugyooin-no kyuuyo-ni han’ee saseteinai]
drastic sales-GEN growth-ACC employee-GEN salary-to reflect not.make
kaisha-no juuyaku]-o hihanshita.
company-GEN executives-ACC criticized

‘The union leader criticized the company’s executives who were not making the drastic sales growth reflected in the employees’ salaries.’

(6) Example conditions for coordinated NP structures (no-scrambling only)

a. non-island / no-scrambling

Sono kaishain-wa tenkinsuru buka-ga
that employee-TOP transferring subordinate-NOM

[NP kompooshita hikkoshi-no nimotsu-o][PP kattabakari-no shinsha-to]
[boxed moving items-ACC][just.purchsed-GEN new.car-with]
tomon fune-de hikkoshisaki-ni okutta-to setsumeeshita
together ship-by new.address-to sent-COMP explained

‘The employee explained that the subordinate who was being transferred shipped boxes of items together with a new car that s/he just purchased to the new address.’

b. island / no-scrambling

Sono kaishain-wa tenkinsuru buka-ga
that employee-TOP transferring subordinate-NOM

[NP kompooshita hikoashi-no nimotsu-o] kattabakari-no shinsha-o]
[boxed moving items-and just.purchsed-GEN new.car-ACC]
tomon fune-de hikkoshisaki-ni okutta-to setsumeeshita.
together ship-by new.address-to sent-COMP explained

‘The employee explained that the subordinate who was being transferred shipped boxes of items and a new car that s/he just purchased together to the new address.’

The value of the factorial definition is that it isolates the island effect in the interaction between SCRAMBLING and STRUCTURE (while subtracting out the main effects of those factors). If there is no island effect, we expect to see no interaction as illustrated in the left panel of Figure 1, where the two lines that connect the two means for the island condition sentences and the non-island
condition sentences are parallel. If there is an island effect, we expect to see a superadditive interaction as illustrated in the center and right panels, where the two lines are not parallel because the mean for the scrambling/island condition sentence is lower than expected if the effects of the two manipulations are all there are. Crucially, we can also look at the size of the interaction as a measure of the size of the island effect (e.g., to test the claim by Saito 1985; 1987); the center panel illustrates a smaller effect, and the right panel illustrates a larger effect.

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

### 3.2 Materials and survey construction

In Experiment 1, each participant completed an experiment that consisted of 58 items: 6 practice items, 12 experimental items and 40 filler items pseudorandomized to avoid related experimental items appearing in succession. The 12 experimental items consisted of 1 token of each of the 4 conditions for each of the three islands. We chose one judgment per condition per participant to keep the total number of experimental items low to minimize the chance that participants would notice the goal of the experiment. We compensated for the increased risk of noise with one judgment per condition by testing a sample size ($n = 40$) that is likely to yield good statistical power (Sprouse & Almeida 2017). We created 8 lexical sets of items per island. The items were then distributed among 8 experimental lists using a Latin square procedure so that participants saw a unique lexical item in each condition.

In Experiment 2, each participant completed an experiment that consisted of 60 items: 16 experimental items and 44 filler items pseudorandomized to avoid related experimental items appearing in succession. The crucial difference between the experiments is that in Experiment 2, participants rated 2 tokens per condition so that we could investigate the consistency of participants’ ratings. To keep the total number of experimental items low (to minimize the risk that participant would identify the goal of the experiment), we used the same non-island (declarative CP) conditions for all three island types, resulting in a total of 8 experimental conditions (two shared, and two unique for each of the three islands). Because the factorial definition subtracts out main effects, the declarative CP island conditions yield exactly the same island effect (the difference between the two designs would be in the interpretation of the main effect of structure). We created 4 lexical sets of items per structure. The items were then distributed among 2 experimental lists using a Latin square procedure so that participants saw a unique lexical item in each trial.

### 3.3 Participants and presentation

Participants were students at a university in Tokyo, Japan. They were all self-reported native speakers of Japanese. They received course credit for their participation. Forty students
participated in Experiment 1, and 32 students participated in Experiment 2. The experiment was administered online using IBEX (Drummond 2013). Each sentence was presented one at a time on its own presentation screen with a 1 (mattaku fushizen ‘completely natural’) to 7 (mattaku shizen ‘completely unnatural’) scale. Participants indicated their rating by clicking on the appropriate number. Because complex NP islands may show variability across participants, we did not exclude any participants from analysis.

4. The results of the experiments

In this section we describe the results of the experiments, with a particular focus on (i) the presence or absence of the superadditive interaction indicative of island effects (and, relatedly, the relative size of the effect), (ii) the variability of island effects across participants, and (iii) the consistency of participants’ ratings across the two tokens of each condition in Experiment 2.

4.1 The presence or absence of island effects

To determine the presence or absence of island effects, we will look for two properties: (i) a visual pattern indicating a superadditive interaction among the four conditions in the factorial design (as illustrated in Figure 1), and (ii) statistical corroboration of the superadditive interaction. To assess the visual patterns in the results, Figure 2 reports the means and estimated standard errors (±1) for each condition in each experiment, arranged in an interaction plot. We z-score transformed the results of each participant prior to analysis to reduce the impact of common forms of scale bias. The top row reports the results for Experiment 1, and the bottom row reports the results for Experiment 2. The columns report each island type.

For statistical corroboration, we conducted two types of analyses: one in a null hypothesis testing framework and one in a Bayesian framework. For the null hypothesis test, we constructed linear mixed effects models with scrambling and structure as fixed effects and participant and item as random effects (intercepts only) for each island type using lme4 package in R (Bates et al. 2015). We calculated p-values using the lmerTest package, which uses the Satterthwaite approximation for degrees of freedom (Kuznetsova et al. 2017). The full set of results are reported in the appendix. We will interpret p-values below the conventional threshold of .05 as evidence against the null hypothesis, and therefore by implication, corroboration of the presence of an island effect. We will interpret p-values above the conventional threshold of .05 as a failure to reject the null hypothesis. For the Bayesian analysis, we calculated Bayes factors for the interaction term for the fixed effects in the linear models using the BayesFactor package (Morey & 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 (H1) that an interaction is present to the likelihood of the data under the null hypothesis (H0) that there is no interaction present. Following Jeffreys (1939/1961), we will interpret a BF_{10} greater than 3 as strong evidence that an interaction is present, as this indicates that the data is at least 3x more likely under a theory in which the interaction is present than one in which the interaction is absent. Similarly, we will interpret a BF_{10} less than 0.33 as strong evidence that there is no interaction, as this indicates that the data is 3x more likely under a theory in which the interaction is absent. We will also interpret Bayes factors near 1 as inconclusive (as the data is equally likely under both theories). We have added the interaction term p-value and interaction BF_{10} to each cell of the plot.
In Experiment 1, we see clear evidence of island effects with coordinated NP structures – the visual pattern suggests a superadditive interaction, and both statistical analyses corroborate the interaction. We also see substantial evidence of island effects for relative clauses – the visual pattern suggests an interaction and the Bayes factor analysis corroborates this, but the null hypothesis test yielded a marginal $p$-value. However, for noun complements, we see no visual pattern of an interaction. The $p$-value is substantially above the conventional threshold, suggesting a failure to reject the null hypothesis. The Bayes factor is 0.41, which is close to the conventional threshold of 0.33, and suggests that the data is about 2.5x more likely under the null hypothesis than under the experimental hypothesis. We thus cautiously conclude that Experiment 1 suggests that noun complements are not islands for NP-scrambling.

In Experiment 2, we again see clear evidence of island effects with coordinated NP structures – the visual pattern suggests a superadditive interaction, and both statistical analyses corroborate the interaction. We also see clear evidence of island effects with relative clauses – this time corroborated by both statistical analyses. However, for noun complements, we have an inconclusive result. There is a visual pattern indicating a small superadditive interaction, but both statistical analyses are inconclusive – the $p$-value is above the conventional threshold of .05, and the Bayes factor is near 1. Thus Experiment 2 suggests either that noun complements are not islands for NP-scrambling, or that noun complements yield a small island effect that cannot be reliably detected at the sample size tested here (32 participants, two judgments per condition).

4.2 Variability in island effects between participants

One possibility raised by Yano (2019) to explain the contradictory results for noun complements across the two experiments is that the island status of noun complements may show more between participant variability than other island types. To investigate this, we calculated the size of the island effect reported by each participant as a differences-in-differences (DD) score
(Maxwell & Delaney 2003): (island/scrambling – non-island/scrambling) – (island/no-scrambling – non-island/no-scrambling). These DD scores will be negative when the participant shows a superadditive interaction indicative of an island effect, with the magnitude indicating the size of the effect; these DD scores will be 0 when the participant shows no interaction, and positive if the participant shows a pattern in which the island-violating condition is more acceptable than the main effects of structure and scrambling would predict (this latter case is not predicted by any theory, so may be indicative of noise in that participant’s responses).

Figure 3 reports the distribution of DD scores for each island type in both experiments using histograms overlaid with probability density estimates.

**Figure 3**: The distribution of island effect sizes by participant, calculated as differences-in-differences scores. Histograms are drawn with a bin width of 0.25 (on the z-score scale). The solid line is an estimate of probability density.

One clear sign that noun complements are more variable than the other islands would be for the distribution for noun complements to be wider than the distributions for the other islands. However, this is not what we see in Figure 3. If anything, noun complements show a narrower distribution, particularly in Experiment 1. What we see instead is that noun complements in Experiment 1 show a relatively normal distribution centered around 0, as expected if there is no island effect, while the other two islands show distributions that are substantially shifted toward the negative range, as expected if there is an island effect. In Experiment 2, noun complements show a relatively normal distribution that is shifted slightly toward the negative; again, this matches the means reported in Figure 2, suggesting the possibility of a small island effect. The other island types in Experiment 2 continue to show distributions that are shifted substantially toward the negative range, again, as expected given the means reported in Figure 2. We therefore conclude that it is unlikely that noun complements show more between-participant variability than the other island types; at least, there is no evidence of it in our experiments. Instead, we see further corroboration from the relatively normal, and relatively narrow, distributions for noun complements that there is either no island effect, or a relatively small island effect that cannot be reliably detected using standard statistical analyses at these sample sizes.
4.3 Variability in island effects within participants

Though there is no evidence of increased variability for noun complements between participants, it is possible that there is increased variability within participants. Recent work by Kush et al. (2018; 2019) in Norwegian has suggested that some island effects that appear relatively small when viewed through the grand means of the sample may in fact be driven by inconsistent judgments within each participant. Though the source of this inconsistency is still an open area of investigation, here we provide a similar analysis for the ratings in Experiment 2.

Figure 4 plots the two judgments that each participant gave for each structure in a scatterplot, with the first judgment along the x-axis and the second judgment along the y-axis. The columns represent each structure, and the rows separate the no-scrambling (top row) and scrambling (bottom row) conditions. We divide each plot into four quadrants. A point in the top right quadrant (Quadrant 1) represents a participant who rated both tokens in the upper half of the scale. For convenience, we will label such a pattern consistent acceptor. A point in the bottom left quadrant (Quadrant 3) represents a participant who rated both tokens in the lower half of the scale. We can label such a pattern consistent rejector. The other two quadrants (Quadrants 2 and 4) represent participants who rated one token in the upper half of the scale and one token the lower half of the scale. We will label this pattern inconsistent. We have added two features to make the plot a bit easier to read: colors representing the three patterns, and two-dimensional (joint) probability density estimates to draw attention to the density of the points in each location. Similar to a topographic map, in a two-dimensional probability density plot, concentric circles that are closer together represent higher density (because, like topographic maps, these plots are looking down on the peaks in the density space from directly above).

Figure 4: Scatterplots of the ratings for the two tokens of each condition for each participant, with two-dimensional (joint) probability density estimates overlaid. The points are colored according to the type of judgment pattern defined by the midpoint (0) of the z-score scales.
The no-scrambling conditions (top row) appear to show the upper bounds of consistency in this experiment – the vast majority of participants show the consistent acceptor pattern (Quadrant 1) for each structure, with a small number of inconsistent patterns mixed in. In contrast, the scrambling conditions (bottom row) reveal potentially relevant patterns. In the first column, for the by-hypothesis grammatical scrambling out of declarative CP, we see that the largest mass of participants is in the consistent acceptor quadrant, while there is also a non-negligible number of participants in each of the other three quadrants. This provides more nuance to the mean rating in Figure 2 – we see now that the middle-of-the-scale mean rating (near 0) was actually driven by a mix of consistent acceptors, consistent rejectors, and inconsistent participants. This sets a baseline expectation for NP-scrambling consistency: the rating of NP-scrambling itself, in the absence of islands, is relatively variable in Japanese. We can then apply this baseline as we look at the potential island structures. For noun complements, we see a shift in the probability mass that moves a bit to the left and down from the declarative CP baseline. The center of mass does not quite fully cross into the consistent rejector quadrant, but instead hovers over the horizontal axis line (indicating a rating near 0 for the second token). This shift from the baseline established by scrambling out of declarative CPs is in line with the equivocal results that we saw for noun complements in the means in Figure 2 and the DD scores in Figure 3 – there is a small trend toward a slightly negative rating, but enough variability that it is still plausibly an effect around 0. For relative clauses, we see a clear shift further into the consistent rejector quadrant, along with a narrowing of the area covered by the dots, consistent with the grand means in Figure 2 and DD scores in Figure 3. Finally, we see an even further shift within the consistent rejector quadrant for coordinated NP structures, again with a narrowing of the area covered by the dots, which is again consistent with the grand means in Figure 2 and the DD scores in Figure 3. Taken together, what we see is that scrambling itself introduces a fair amount of within-participant variability to judgments when compared to no-scrambling conditions. Noun complements appear (visually) to show roughly the same amount of variability; so it is not the case that noun complements introduce additional variability, but rather that they preserve the variability that already exists for scrambling conditions. Instead, what we find is that clear islands like relative clauses and coordinated NP structures reduce the amount of variability in the judgments. Again, this pattern points to noun complements being qualitatively distinct from relative clauses and coordinate structures.

5. General Discussion and Implications

This paper presented two experiments using the factorial definition of island effects to compare three island types: relative clauses, noun complements, and coordinated NP structures. Experiment 1 tested one token per condition, whereas Experiment 2 tested two tokens per condition, allowing for an investigation of within-participant variability. The two experiments unequivocally show that relative clauses and coordinated NP structures are islands in Japanese, corroborating previous studies (e.g., Harada 1977; Haig 1978; Saito 1985; 1987). The results for noun complements, on the other hand, were less clear-cut. While Experiment 1 showed a relatively clear lack of evidence of island effects, the results from Experiment 2 were inconclusive: the visual pattern indicated a small effect, but the null hypothesis test failed to reject the null hypothesis, and the Bayes factor analysis indicated that the data was roughly equally likely under the two hypotheses. We also closely examined data for between- and within-participant variability. Although Yano (2019) raised the possibility that that noun complements are associated with an increase in between-participant variability (compared to adjunct islands),
we did not find evidence of this in Experiment 1 or 2. There was, however, some within-participant variability in Experiment 2, which was suggested in previous studies of island effects in Norwegian (Kush et al. 2018; 2019) to be one potentially interesting cause of the appearance of small island effects in the sample means. However, that within-participant variability appears to be driven by scrambling in general. That variability is preserved in noun complements, but reduced in relative clauses and coordinate structures. Taken together, our findings therefore suggest that noun complements are either not islands, supporting Haig (1978), or are very small islands, supporting Saito (1985; 1987), albeit too small to detect reliably at the sample sizes used here and in Yano (2019).

Although the disjunction appears unsatisfying, it does lead to one strong conclusion: noun complements are qualitatively different from relative clauses in Japanese. Either noun complements are not islands, or they are different types of island – one with an effect size that is substantially smaller than the effect size for relative clauses, and one that does not lead to a reduction in within-participant variability. Though it has occasionally been claimed that island effects for noun complements are smaller than island effects with relative clauses (e.g., Chomsky 1986), to the best of our knowledge, our study is the first to provide experimental evidence either for a difference in effect size, or for the two structures to pattern qualitatively differently. In fact, the only previous studies to our knowledge to directly test both noun complements and relative clauses in the same experiment are Kush et al. (2018; 2019), which tested both islands in Norwegian for wh-questions and topicalization, respectively (cf. Pañeda et al. 2020 for a review of the islands tested in previous experimental studies). The results of these two studies suggest that the island effect size for noun complements and relative clauses are approximately equal (though it is always possible that the difference in effect size is simply too small to detect reliably with the sample sizes used in these studies). Our findings also challenge the claim that noun complements are simply relative clauses in disguise (e.g., Nichols 2003; Kayne 2008; 2010; Arsenijević 2009; Haegeman 2012; cf. de Cuba 2017), and the claim that Japanese lacks English-like relative clauses entirely (e.g., Kuno 1973; Murasugi 2000). Under these analyses, it would be unexpected to find that only relative clauses are islands in Japanese. Finally, it should be noted that our results for noun complements also show that systematic data collection is not a panacea for effects that are small or difficult to detect; instead, systematic data collection often tends to quantitatively corroborate that the effect is small or difficult to detect (as argued in Phillips 2009), as appears to be the case here.

We offer two more observations about the two types of complex NPs in Japanese. First, our results appear to challenge Stowell’s (1981) suggestion that island status correlates with the possibility of complementizer deletion: CP complements of verbs allow complementizer deletion (7), and are not islands, while CP complements of nouns do not allow complementizer deletion, and are islands (8). (The examples in (7) and (8) are our own, with diacritics indicating the pattern discussed by Stowell 1982.)

(7) Jessica claimed that/∅ Lisa invented the app.

(8) Jessica made the claim that/*∅ Lisa invented the app.

Stowell argues that complementizer deletion is possible in (7) because the embedded clause is a true complement of the verb (and therefore the empty category created by the deletion is governed, satisfying the ECP). Stowell further argues that the impossibility of complementizer
deletion in (8) suggests that the embedded clause is, in fact, not a complement, but rather an adjunct (leading to an ECP violation because the empty category created by the deletion is not governed; cf. Grimshaw 1990; Kiss 1990; Takahashi 1994; Sabel 2002; de Cuba 2017). If the embedded clause in noun complement constructions is in fact an adjunct, the island effect observed with English noun complements is expected as a type of adjunct island effect. Interestingly, Fukui (1988) argues that the Japanese noun complements with toyuu also involve an adjunct because toyuu cannot be deleted, as in (9).

(9) Taro-ga sore-o teniireta toyuu/*∅ uwasa
    T-NOM it-ACC obtained toyuu rumor
    ‘The rumor that Taro obtained it’ (Fukui 1998: 513; (26))

If Fukui (1988) is correct, and the embedded clause inside the noun complement is an adjunct, our finding that toyuu noun complements show little to no island effect is puzzling. Fukui’s claim, however, is not supported by other Japanese-internal facts about complementizer deletion. First, the Tokyo dialect does not allow for deletion of the complementizer to/tte even when the embedded clause appears to be a clear case of a verbal complement. (The example in 10 is our own, with diacritic indicating the pattern reported in Fukuda 2000).

(10) Taro-wa Hanako-ga kita to/tte/*∅ itta/omotta
    T-TOP H-NOM came to/tte said/thought
    ‘Taro thought/said (that) Hanako came.’

Thus, we have no reason to expect that complementizers in general can be deleted even when the embedded clause is indeed a complement; therefore it is not surprising that toyuu in (9) cannot be deleted. Second, in some Western dialects of Japanese, the complementizer deletion in (10) is acceptable (Saito 1987; Fukuda 2000; Kishimoto 2006). Yet, crucially, no one has claimed that the CP complement of a verb with to in the Tokyo dialect is an island while it is a non-island in these Western dialects. What is more, according to speakers of these dialects whom we consulted, the same complementizer to/tte can also be deleted in noun complements, as in (11), though we note that this fact should be quantified in future judgment or corpus studies.

(11) Taro-ga sore-o teniireta tte/∅ yuu uwasa
    T-NOM that-ACC obtained tte YUU rumor
    ‘The rumor that Taro obtained that.’

Thus, when we look at the facts of complementizer deletion across both English and Japanese, the apparent correlation between island status and the possibility of complementizer deletion disappears.²

Our second observation is that island status does not appear to correlate with the syntactic complexity of the embedded clause (in terms of number of available positions in the clause, or

---

² We note that we are following Fukui’s assumption that toyuu is an English-style complementizer similar to that. It could be argued that toyuu is not a complementizer, but rather a semi-frozen verbal element similar to the English noun phrase “the rumor that says that Taro obtained it”. In that case, the example in (9) may no longer be relevant to the correlation between complementizer deletion and island status. But, to our knowledge, it is widely agreed that to is a complementizer in Japanese; therefore the examples in (10) and (11) are still relevant to the correlation.
amount of functional structure within the clause). While relative clauses and noun complements in English are analyzed as involving an embedded CP, i.e., embedded clauses with the same complexity, there are many reasons to believe that the structure of relative clauses in Japanese is less complex along certain dimensions than the structure of the embedded clauses inside noun complements: (i) relative clauses cannot be followed by a complementizer while noun complements can (e.g., Murasugi 2000; cf. Tomioka 2015), (ii) relative clauses cannot have a topic inside the clause while noun complements can (e.g., Tomioka 2015), (iii) relative clauses allow the head NP (or D) to license genitive case on the subject while noun complements do not (e.g., Harada 1971; Watanabe 1996; Ochi 2001; Miyagawa 2011), and (iv) relative clauses allow a non-episodic interpretation of change-of-state verbs while noun complements do not (e.g., Ogihara 2004). Despite all these observations indicating that relative clauses are structurally less complex along certain dimensions than noun complements in Japanese, only relative clauses show clear island effects. This observation potentially has implications for bounding-based approaches to island effects (like the classic Subjacency and barriers approaches). Though it clearly depends on the details of the theory, in principle, more complex syntactic constituents like noun complements in Japanese have the potential to host more bounding nodes or barriers than less complex constituents like relative clauses, despite island effects patterning in the opposite direction. This suggests that our results may be relevant for adjudicating among specific implementations of bounding-based approaches to island effects.

While it is beyond the scope of this paper to evaluate the full set of theories of islands in the literature, we would like to mention the consequences of our results for a few prominent theories to illustrate their potential theoretical value. Huang’s (1982) Condition on Extraction Domains, Lasnik & Saito’s (1984) gamma-marking, and Chomsky’s (1986) barriers approach all share the intuition that there is a fundamental distinction between adjunct CPs and complement CPs (in terms of government, gamma-marking, and L-marking respectively), and that this distinction causes (most) adjuncts to be islands and (most) complements to be non-islands. English noun complements create complications for this view, as they appear to be complements, but are nonetheless islands. Under these approaches, our finding that Japanese noun complements are likely not islands (or are extremely small islands) suggests that they are indeed complements, and that it is English noun complements that are exceptional in some way, not noun complement constructions in general (but see Hankamer & Mikkelson (to appear) for a proposal that Dutch and English noun complements involve a CP complement to D or a CP adjunct to DP). The intuition that complements and adjuncts are fundamentally distinct, and that this difference is the source of island effects, is also central to modern phase-based approaches to island effects. For example, Rackowski & Richards (2005) propose that phrases that enter into an Agree relation with a phase head (i.e., v) are transparent to extraction, while phrases that do not enter into an Agree relation are islands. They present evidence that, in Tagalog, complement CPs, which are transparent to extraction, show morphological evidence of this Agree relation, while adjunct CPs, which are islands, do not. Under this approach, our results suggest that, while noun complements in English must not enter into an Agree relation with the next phase head, noun complements in Japanese do, despite the fact that neither language shows a morphological reflex of this relation. Similarly, Müller (2010) proposes that phrases are transparent to extraction as long as they are active and can thus be given an edge feature to accommodate the extraction, in turn circumventing the Phase Impenetrability Condition of Chomsky (2001; 2008). Phase heads are active until their final (i.e., last merged) specifier is merged. Adjuncts are islands because they are the final specifier of a phase head, while complements are not islands because
(by definition) they are not specifiers. Under this approach, English noun complements must be final specifiers, and thus adjuncts, while Japanese noun complements would be true complements. For each of these potential analyses, future studies could explore the properties of English and Japanese noun complements (beyond complementizer deletion) to determine if there is independent evidence (beyond island effects) for postulating these critical differences between the two types of CPs.

6. Conclusion

This article presented two studies that contribute to the discussion of complex NP islands in Japanese. While there is little contention among previous studies that relative clauses are islands, there is no consensus regarding noun complements: Haig (1976) argued they are not islands, Saito (1985; 1987) claimed they are, and Yano’s (2019) two experiments yielded contradictory results, with one showing a small island effect and the other showing no island effect. We presented two acceptability judgement experiments to compare relative clauses and noun complements with coordinated NP structures, which are uncontroversially islands. Both experiments were designed using the factorial definition of island effects. Participants saw one token per condition per participant in Experiment 1, while participants saw two tokens per condition in Experiment 2, thus allowing us to explore possible between- and within-participant variability, which has been raised as a possible complicating factor in previous studies of island effects (Kush et al. 2018; 2019; Yano 2019). Our results corroborated previous studies in that relative clauses as well as coordinated NP structures yield large island effects in Japanese. Our results for noun complements were less straightforward: while Experiment 1 yielded conclusive evidence that they are not islands, Experiment 2 yielded inconclusive results. A closer look at the results revealed no between-participant variability in the judgments on noun complements in Experiments 1 or 2, and the same level of within-participant variability (in Experiment 2) for both scrambling out of verb complements and noun complements. We conclude either that noun complements in Japanese are not islands or that noun complements in Japanese yield island effects that are too small to reliably detect at these sample sizes, and are therefore qualitatively distinct from relative clause island effects, which are easy to detect at these sample sizes.

While our studies cannot adjudicate between the disjunction above, they do attest to a clear difference – whether the difference lies in the island status or in the size of island effects – between relative clauses and noun complements, at least in Japanese, and more broadly, between noun complements in Japanese and noun complements in other languages that have been tested using the factorial definition, such as English, Italian, and Norwegian (Sprouse et al. 2011, Sprouse et al. 2012, Sprouse et al. 2016, Kush et al. 2018; 2019). We have argued that these findings have direct consequences for most existing theories of island effects, for theories of the relationship between complementizer deletion and island status, and for theories of the relationship between the complexity of syntactic structure and island status. This in turn suggests that future studies should probe the properties of relative clauses and noun complements (cross-linguistically) along these dimensions.

Our study used scrambling dependencies to evaluate island effects, because wh-questions, which are typically used to investigate island effects in wh-movement languages, do not move overtly in Japanese. There are, however, studies that suggest that wh-in-situ also plays an important role in the theories of islands in these languages. Studies such as Tanaka & Schwartz (2017) on Japanese and Lu et al. (2020) on Mandarin Chinese found that argument wh-phrases
that stay *in situ* within a relative clause give rise to island effects, contrary to previous claims that only certain *wh*-adjuncts invoke island effects in these languages. (cf. Kim & Goodall 2016, whose experimental investigation of island effects involving *wh*-in-*situ* in Korean also observed island effects; but cf. Sprouse et al. 2011, which found no island effects for subject, adjunct, *whether*, and noun complement islands in Japanese.) The complex picture emerging from these studies suggests a need for comprehensive comparisons across island effects and dependency types within *wh*-in-*situ* languages.

**Abbreviations**

ACC = accusative, CNP = complex noun phrase, COMP = complementizer, CP = complementizer phrase, DAT = dative, GEN = genitive, NOM = nominative, TOP = topic,

**References**

Christensen, Ken Ramshøj, Johannes Kizach & Anne Mette Nyvad. 2013. Escape from the island: Grammaticality and (reduced) acceptability of *wh*-island violations in Danish. *Journal of Psycholinguistic Research* 42. 51–70. DOI: http://dx.doi.org/10.1007/s10936-012-9225-3
Drummond, Alex. 2013. Ibex Farm. https://spellout.net/ibexfarm/.

18
Müller, Gereon. 2010. On deriving CED effects from the PIC. *Linguistic Inquiry* 41(1). 35–82. DOI: https://doi.org/10.1162/ling.2010.41.1.35


Tomioka, Satoshi. 2015. Embedded wa-phrases, predication and judgment theory. *Natural Language and Linguistic Theory* 33. 267–305. DOI: https://doi.org/10.1007/s11049-014-9258-4


