Reexamining Island Effects in Japanese Complex NPs with Argument *Wh-in-situ*

Nozomi Tanaka
Indiana University

1 Introduction

The proposal of a factorial definition of island effects has led to recent research revisiting the notion of islands in various languages using formal acceptability judgment experiments. This research has discovered measurable island effects in structures previously thought to elude them, such as complex NP (CNP) island effects with argument *wh-in-situ* in *wh-in-situ* languages (Lee and Park 2015; Lee 2018; Tanaka and Schwartz 2018; Lu et al. 2020; Nguyen 2021).

The current study reexamines the island status of CNPs with argument *wh-in-situ* in Japanese, a structure previously considered to be not island sensitive. The results from three formal acceptability judgment experiments using the factorial definition of island effects show measurable effects with CNPs with argument *wh-in-situ*.[1]

Section 2 provides an overview of CNP island effects in Japanese and other languages, Section 3 presents the logic of the study's factorial design, Sections 4–6 describe the methodology and results of the three experiments, and Section 7 offers a general discussion and conclusions.

2 Background

While *wh*-movement is not clause-bound (1), CNPs such as relative clauses (RCs) (2) and noun complements (NCs) (3) are generally considered to restrict *wh*-movement (Sprouse and Hornstein 2013: 2).

(1) What does Susan think [that John bought _]?
(2) * What did you meet [the scientist who invented _]?
(3) * What do you make [the claim that John bought _]?

In *wh-in-situ* languages, argument *wh-in-situ* is generally believed to be immune to CNP island effects (Huang 1982; Nishigauchi 1986; Choe 1987; Richards 2008). In Japanese, for example, argument *wh-in-situ* in RCs (4) or NCs (5) is possible, while adjunct *wh-in-situ* such as *naze* 'why' within CNPs leads to unacceptability (6).

(4) Takeshi-wa [**nani**-o   katta   onnanohito-o]
     T-TOP   what-ACC   bought   woman-ACC
     tsuretekimashita-ka
     brought-Q
     'What did Takeshi bring [the woman who bought _]?'

(5) Takeshi-wa [onnanohito-ga   **nani**-o   katta-toiu
     T-TOP   woman-NOM   what-ACC   bought-COMP
     hanashi-o]   shimashita-ka?
     story-ACC   tell-Q
     'What did Takeshi tell [the story that a woman bought _]?'

---

[1] The original presentation at the 30th Japanese/Korean Linguistics Conference included a fourth experiment, but this paper focuses on the three experiments due to space limitations.

(6) * Takao-wa [Haruna-o **naze** hometa sensē-o]
       T-TOP   H-ACC   why   praised teacher-ACC
    tsuretekimashita-ka?
    brought-Q
    'Why did Takao bring [the teacher who praised Haruna _]?'

Previous work has sought to explain the lack of CNP island effects in *wh-in-situ* languages by proposing that *wh*-phrases do not move (e.g. Aoun and Li 1993), that covert *wh*-movement is not subject to island effects (e.g. Huang 1982; Lasnik and Saito 1984), or that island effects incurred by covert *wh*-movement can be circumvented by pied-piping the whole CNP (e.g. Nishi-gauchi 1986; Choe 1987).

However, recent formal experimental studies on *wh-in-situ* languages that used the factorial definition of island effects detected island effects previously claimed to be absent. The studies have shown that argument *wh-in-situ* triggers RC island effects in Chinese (Lu et al. 2020), Japanese (Tanaka and Schwartz 2018), and Vietnamese (Nguyen 2021). Island effects have also been observed with argument *wh-in-situ* inside NCs in Korean (Lee and Park 2015; Lee 2018), although not in Japanese (Sprouse et al. 2011).

A finding of (previously unobserved) measurable *wh*-island effects with *wh*-movement in Brazilian Portuguese led Almeida (2014) to propose 'subliminal island effects': measurable island sensitivity effects that do not lead to gross sentence unacceptability, in contrast to traditional, 'supraliminal' island effects, which lead to categorical unacceptability. Subliminal island effects, however, must be carefully examined, as they may reflect participant variability (Kush et al. 2018, 2019) or processing confounds (Keshev and Meltzeer-Asscher 2019).

These experimental studies all conducted formal acceptability judgment experiments using the factorial definition of island effects, as does the current study. As a follow-up to Tanaka and Schwartz 2018, I conducted three acceptability judgment experiments to test the island status of CNPs, specifically RCs and NCs, by manipulating the presence and absence of argument *wh-in-situ*, the former resulting in *wh*-questions and the latter in *yes-no* questions. The results show that there are measurable effects that fit the criteria for subliminal island effects proposed by Almeida (2014).

3 The Logic of the Design

A factorial design crossing two factors, STRUCTURE and QUESTION, each with two levels, resulted in four conditions per experiment. As shown in Table 1, the factor STRUCTURE varied between declarative CPs and RCs in Experiment 1, declarative CPs and NCs in Experiment 2, and NCs and RCs in Experiment

3. The factor QUESTION manipulated the presence of argument *wh-in-situ* in all three experiments. The condition without *wh-in-situ* involved a *yes-no* question and the condition with *wh-in-situ* involved a *wh*-question.

Experiment	STRUCTURE	QUESTION
1	declarative CP	*yes-no* question (no *wh-in-situ*)
	RC	*wh*-question (*wh-in-situ*)
2	declarative CP	*yes-no* question (no *wh-in-situ*)
	NC	*wh*-question (*wh-in-situ*)
3	NC	*yes-no* question (no *wh-in-situ*)
	RC	*wh*-question (*wh-in-situ*)

Table 1. Factorial Designs in Experiments 1–3 (CP = complementizer phrase; NC = noun complement; RC = relative clause).

The factorial design helps isolate island effects as 'superadditive' effects that cannot be explained by the simple combination of effects of multiple experimental factors, thus countering the reductionist view that they are complexity effects arising from processing difficulty (Sprouse et al. 2012). The schematic examples in (7) present the four conditions in Experiment 1.

(7) a. **Declarative CP | *yes-no* question**
   NP-TOP [$_{\text{DECL}}$ NP-NOM   NP-ACC V$_{\text{EMBEDDED}}$COMP]   V$_{\text{MATRIX}}$ Q
   b. **Declarative CP | *wh*-question**
   NP-TOP [$_{\text{DECL}}$ NP-NOM   ***wh*-**ACC V$_{\text{EMBEDDED}}$COMP]   V$_{\text{MATRIX}}$ Q
   c. **RC | *yes-no* question**
   NP-TOP [$_{\text{RC}}$ e$_i$ NP-ACC V$_{\text{EMBEDDED}}$head NP$_i$-ACC]   V$_{\text{MATRIX}}$ Q
   d. **RC | *wh*-question (critical condition)**
   NP-TOP [$_{\text{RC}}$ e$_i$ ***wh*-**ACC V$_{\text{EMBEDDED}}$head NP$_i$-ACC]   V$_{\text{MATRIX}}$ Q

Example (7d) illustrates the condition of interest with RC-internal *wh-in-situ* questions, while (7a) is the baseline condition, with declarative CPs and *yes-no* questions (no *wh-in-situ*). The effect of STRUCTURE can be measured as the difference in acceptability between the two *yes-no* question conditions (7a) and (7c), and the effect of QUESTION can be measured as the difference between the two declarative CP conditions (7a) and (7b). If the total effect, measured as the difference between the baseline (7a) and critical (7d) conditions, is greater than the sum of the effects of STRUCTURE and QUESTION, showing a superadditive relationship, then there are effects that cannot be accounted for by the simple addition of the effects of STRUCTURE and QUESTION: that is, island effects.

A superadditive relationship can be envisioned as nonparallel lines mapping two factors, as shown in the righthand graph in Figure 1. When there are

no superadditive effects, two parallel lines, as shown in the lefthand graph, indicate that any lowering in the acceptability of (7d) can be accounted for by the combination of the STRUCTURE and QUESTION effects. A superadditive relationship can be statistically confirmed as a significant interaction between STRUCTURE and QUESTION.
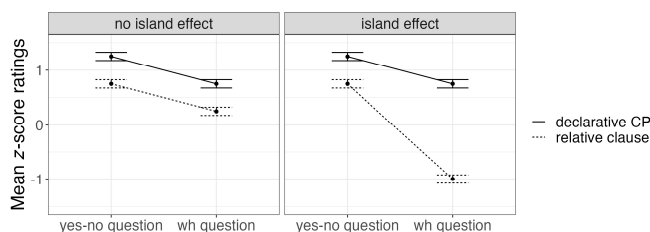

Figure 1. Patterns predicted for Experiment 1 (no island effects vs. island effects).

A superadditive effect can also be measured as a difference-in-difference (DD) score (Maxwell and Delaney 2003), calculated by subtracting the difference between (7a) and (7c) from the difference between (7b) and (7d): DD = (7b – 7d) – (7a – 7c). A DD score of zero indicates the absence of island effects; a DD score larger than zero indicates an island effect. DD scores also quantify the size/strength of the superadditive effect. Hence, by-participant DD scores have been used to assess the relation of the size of the effect to other variables such as working memory (e.g. Sprouse et al. 2012) and language proficiency (e.g. Zenker and Schwartz 2017), and to show between- and within-participant variation (Kush et al. 2018, 2019; Fukuda et al. 2022; see also Fukuda et al. in this volume).

As discussed in Section 2, island effects can be supraliminal or subliminal. The graphs in Figure 2 illustrate both scenarios.
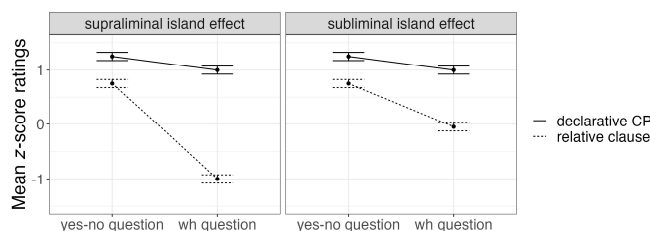

Figure 2. Patterns predicted for Experiment 1 (supraliminal vs. subliminal island effects).

In the graphs, zero on the *y*-axis represents the grand mean. The graph on the left visualizes supraliminal island effects: The critical condition receives a mean *z*-score rating well below zero, signifying that the structure is relatively less acceptable. The graph on the right visualizes subliminal island effects: There is measurable superadditivity, but the mean rating for the critical conditions hovers around zero or is above zero.

Similar logic was used for Experiment 2, which compared declarative CPs and NCs, and for Experiment 3, which compared NCs and RCs. As detailed below, the inclusion of NCs aimed to address the differences in the number of dependencies involved in the embedded structures.

4 Experiment 1

Experiment 1 compared declarative CPs and RCs with and without argument *wh-in-situ* to reexamine the island status of RCs.

4.1 Method

4.1.1 Participants

Self-identified native speakers of Japanese ($n = 91$, 41 female, mean age = 35.7) were recruited at universities in Japan and in North America, through social media, and using a Japanese crowdsourcing platform, CrowdWorks (https://crowdworks.jp). Participants were compensated.

4.1.2 Materials

A factorial design crossing two factors, STRUCTURE (declarative CP, RC) and QUESTION (*yes-no* question, *wh*-question) resulted in four conditions. The example in (8) presents the four versions of a sample lexicalization.

(8) Four versions of a sample lexicalization in Experiment 1
    a. **Declarative CP | *yes-no* question**
        Momoko-wa [otokonohito-ga   kaban-o   katta-to]
        M-TOP   man-NOM   bag-ACC   bought-COMP
        iimashita-ka?
        say-Q
        'Did Momoko say [that the man bought a bag]?'
    b. **Declarative CP | *wh*-question**
        Momoko-wa [otokonohito-ga   **nani**-o   katta-to]
        M-TOP   man-NOM   what-ACC   bought-COMP
        iimashita-ka?
        say-Q
        'What did Momoko say [that the man bought _]?'

c. **RC | *yes-no* question**
Momoko-wa [kaban-o   katta   otokonohito-o]
M-TOP   bag-ACC   bought man-ACC
mimashita-ka
saw-Q
'Did Momoko see [the man who bought a bag]?'

d. **RC | *wh*-question**
Momoko-wa [**nani**-o   katta   otokonohito-o]
M-TOP   what-ACC bought man-ACC
mimashita-ka
saw-Q
'What did Momoko see [the man who bought _]?'

Four versions of forty base lexicalizations generated 160 experimental items, which were distributed across four lists using a Latin square design so that each list contained only one version of each lexicalization. The 40 experimental items (4 conditions × 10 tokens) were combined with 40 fillers (20 acceptable, 20 unacceptable) for a total of 80 items in each list.

4.1.3 Procedure

Participants completed a web-based acceptability judgment task administered through Qualtrics. Participants were randomly assigned to one of the four lists, and the order of the items within a list was randomized. Participants were asked to rate the naturalness of each sentence on a scale of 1 (*totemo fushizen* 'very unnatural') to 4 (*totemo shizen* 'very natural'), with an off-scale option, *wakaranai* 'I don't know'.[2]

4.2 Results

After blank or 'I don't know' responses were excluded (0.63%), each participant's acceptance ratings were standardized through *z*-score transformation following Sprouse et al. 2012. The lme4 (Bates et al. 2015) and lmerTest

---

[2] All three experiments were originally designed as a study on second language learners on Japanese. An even-numbered scale was used to avoid a midpoint, which might reflect uncertainty, and the 'I don't know' option was included for instances in which learners were uncertain or unable to make a judgment (Ionin and Zyzik 2014; Ionin 2021). Additionally, forty base lexicalizations were prepared so that each participant saw ten tokens per condition. This differed from most previous studies where participants encountered fewer tokens per condition. While there is a risk of satiation effects in native speakers, a larger number of tokens per condition was necessary due to a large amount of noise typically observed in L2 data. The lexical items used in the stimuli were all taken from elementary-level teaching materials for Japanese as a second language, which significantly restricted the choice of vocabulary that can be used for the experiments.

(Kuznetsova et al. 2017) packages in R version 4.1.1 (R Core Team 2021) were used to fit a linear mixed effects model estimating *z*-score ratings as a function of treatment-coded predictors, STRUCTURE and QUESTION. Random effects included by-participant and by-item varying intercepts. Figure 3 presents the interaction plot, DD score, and *p*-value for the interaction between STRUCTURE and QUESTION in the linear mixed effects model.
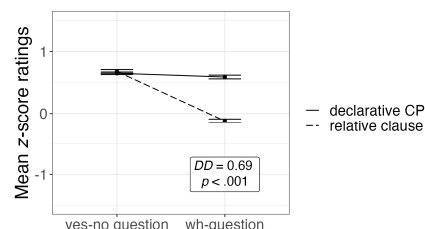
Figure 3. Experiment 1: Interaction plot (points: conditional means; error bars: standard errors), DD score, and *p*-value for the interaction between STRUCTURE and QUESTION in the linear mixed-effects model.

As shown in Figure 3, the two *yes-no* question (no *wh-in-situ*) conditions exhibit very similar mean *z*-score ratings (declarative CPs: 0.65; RCs: 0.68). Declarative CPs with *wh-in-situ* questions received a slightly lower but very similar mean rating (0.59). In contrast, the mean rating for the critical condition (RC | *wh*-question) is much lower (−0.12) than the other three conditions, although it is just below zero. The nonparallel lines, the positive DD score (0.69) and a significant interaction between STRUCTURE and QUESTION (coefficient: −0.73, $SE = 0.07$, $t = -10.44$, $p < .001$) all indicate superadditivity.

4.3 Discussion

The results from Experiment 1 showed that mean rating for the critical condition, which involved RC-internal argument *wh-in-situ* questions, was much lower compared to the mean ratings for the *yes-no* question (no *wh-in-situ*) conditions or the declarative CP conditions. The measurable effects associated with RCs with argument *wh-in-situ* cannot be solely explained by the effects associated with RCs (compared to declarative CPs) or those associated with *wh*-questions (compared to *yes-no* questions). These potential island effects fit the characteristics of subliminal island effects because the mean rating for the critical condition hovered around zero.

This experiment, however, had a confound because the critical condition differed from the other conditions in another way. As schematized in (9), it was the only condition that involved two overlapping dependencies: a gap-

filler dependency in RCs and a dependency between a *wh*-phrase and its Q particle suggested in previous work (e.g. Ueno and Kluender 2003).

(9) NP-TOP [$_{RC}$ e$_i$ *wh*-ACC V$_{EMBEDDED}$ head NP$_i$-ACC] V$_{MATRIX}$ Q

Keshev and Meltzer-Asscher (2019) argued that the processing cost of maintaining two dependencies can create apparent subliminal island effects. Therefore, it is possible that the subliminal effects observed in this experiment reflected the maintenance cost of these two dependencies. The next two experiments address this possibility by adding NCs to the factorial designs.

5   Experiment 2

Experiment 2 compared declarative CPs and NCs with and without argument *wh-in-situ*. As explained in Section 2, NCs are another type of CNPs that are considered islands in *wh*-movement languages, but they do not exhibit island effects with argument *wh-in-situ* in *wh-in-situ* languages. Importantly, NCs do not involve a gap-filler dependency, allowing for a direct comparison between non-islands (declarative CPs) and potential islands (NCs) without the number of dependencies as a confounding factor. In the critical condition involving NCs with *wh-in-situ*, only a dependency between a *wh*-phrase and its Q particle is present, much like in declarative CPs with *wh-in-situ*. If superadditivity is observed in this scenario, it is hypothesized to reflect a structural difference between declarative CPs and NCs, i.e., a difference in island status.

5.1 Method

A factorial design crossing two factors, STRUCTURE (declarative CP, NC) and QUESTION (*yes-no* question, *wh*-question) resulted in four conditions as exemplified in (10).

(10) Four versions of a sample lexicalization in Experiment 2
    a. **Declarative CP | *yes-no/wh*-question**
        Takeshi-wa [onnanohito-ga megane/**nani**-o   katta-to]
        T-TOP   woman-NOM   eyeglasses/what-ACC bought-COMP
        hanashimashita-ka?
        say-Q
        yes-no: 'Did Takeshi say [that the woman bought eyeglasses]?'
        wh: 'What did Takeshi say [that the woman bought _]?'

b. **NC | *yes-no/wh*-question**

Takeshi-wa [onnanohito-ga megane/**nani**-o katta-toiu]
T-TOP woman-NOM eyeglasses/what-ACC bought-COMP
hanashi-o shimashita-ka?
story- ACC told-Q
yes-no: 'Did Takeshi tell [the story that the woman bought eyeglasses]?'
wh: 'What did Takeshi tell [the story that the woman bought _]?'

Four versions of forty base lexicalizations were created and Latin-squared into four lists. The 40 experimental items (4 conditions × 10 tokens) were combined with 40 fillers (20 acceptable, 20 unacceptable) for a total of 80 items in each pseudorandomized list.

Self-identified native speakers of Japanese ($n = 111$, female 58, mean age = 33.5), who were recruited through universities in Japan and CrowdWorks, completed a paper- or web-based acceptability judgment task. Participants were randomly assigned to one of the four lists and asked to rate the naturalness of each sentence using the same scale as Experiment 1.

5.2 Results

After blank or 'I don't know' responses were excluded (0.63%), the acceptability ratings were analyzed using the same procedure that was used in Experiment 1. As shown in Figure 4, the mean rating for the critical condition with NC-internal *wh-in-situ* questions (0.09) was much lower (albeit nearly zero) than mean ratings for the other conditions (declarative CP | *yes-no* question: 0.60; declarative CP | *wh*-question: 0.48; NC | *yes-no* question: 0.68).
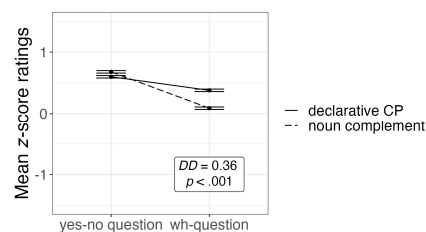

Figure 4. Experiment 2: Interaction plot (points: conditional means; error bars: standard errors), DD score, and *p*-value for the interaction between STRUCTURE and QUESTION in the linear mixed-effects model.

The nonparallel lines, the positive DD score (0.36) and a significant interaction between STRUCTURE and QUESTION (coefficient: −0.36, *SE* = 0.07, *t* = −5.40, *p* < .001) all indicate superadditivity.

5.3 Discussion

Experiment 2 produced similar results to Experiment 1, where significant superadditivity was observed, but with ratings hovering around zero, suggesting subliminal island effects. Unlike in Experiment 1, NCs with *wh-in-situ* in the critical condition only involved one dependency—the same number of dependencies as declarative CPs with *wh-in-situ*. This means that the measurable effects observed in Experiment 2 are likely due to the difference in island status: NCs are (potential) islands, while declarative CPs are not.

6   Experiment 3

Experiment 3 further investigated the source of the effects found in Experiments 1 and 2 by comparing NCs and RCs. As explained previously, both are CNPs; however, RCs involve a gap-filler dependency, unlike NCs. If superadditivity is observed, it suggests that there are effects arising from multiple dependencies or that there is a qualitative difference between NCs and RCs.

6.1 Method

Experiment 3 crossed two factors, STRUCTURE (NC, RC) and QUESTION (*yes-no* question, *wh*-question). The four conditions are exemplified in (11).

(11) Four versions of a sample lexicalization in Experiment 3
   a. **NC | *yes-no/wh*-question**
      Takeshi-wa   [onnanohito-ga   megane/**nani**-o   katta-toiu]
      T-TOP   woman-NOM   eyeglasses/what-ACC bought-COMP
      hanashi-o   shimashita-ka?
      story- ACC   told-Q
      yes-no: 'Did Takeshi tell [the story that the woman bought eyeglasses]?'
      wh: 'What did Takeshi tell [the story that the woman bought _]?'
   b. **RC | *yes-no/wh*-question**
      Takeshi-wa   [megane/**nani**-o   katta   onnanohito-o]
      T-TOP   eyeglasses/what-ACC bought   woman-ACC
      tsuretekimashita-ka
      brought-Q
      yes-no: 'Did Takeshi bring [the woman who bought eyeglasses]?'
      wh: 'What did Takeshi bring [the woman who bought _]?'

Four versions of forty base lexicalizations were created and Latin-squared into four lists, which were mixed with 40 fillers (20 acceptable, 20 unacceptable) to create a total of 80 items in each pseudorandomized list.

The participants in the experiment were self-identified native speakers of Japanese ($n = 107$, female 56, mean age = 30.6) who were recruited through universities in Japan and CrowdWorks. The experiment was conducted following the same procedure as Experiment 2.

6.2 Results

Blank or 'I don't know' responses were excluded from the analysis (0.09%). Participants' acceptance ratings were analyzed using the same procedure as in Experiments 1 and 2. As shown in Figure 5, items with RCs (−0.27) received a lower mean rating than those with NCs (0.31) in the *wh-in-situ* conditions, while the mean ratings for the *yes-no* question (no *wh-in-situ*) conditions did not differ largely between NCs (0.81) and RCs (0.75). The nonparallel lines, the positive DD score (0.5) and a significant interaction between STRUCTURE and QUESTION (coefficient: +0.51, $SE = 0.07$, $t = -6.82$, $p < .001$) all indicate superadditivity.
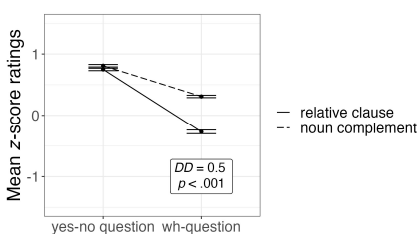

Figure 5. Experiment 3: Interaction plot (points: conditional means; error bars: standard errors), DD score, and *p*-value for the interaction between STRUCTURE and QUESTION in the linear mixed-effects model.

6.3 Discussion

In Experiment 3, superadditivity was observed when comparing NCs and RCs. While both structures are CNPs and potential islands, RCs involve a gap-filler dependency that is not present in NCs. Although the two structures did not differ in terms of acceptability in the *yes-no* question conditions, there was a nonnegligible difference between NCs and RCs in the *wh*-question conditions. One possibility is that this difference reflects the difference in the number of dependencies and that the maintenance of two overlapping dependencies does incur superadditive effects, as suggested by Keshev and Meltzer-Asscher (2019). However, the results from Experiment 2 indicate that the maintenance cost alone does not account for the superadditivity. It is also possible that there exists a qualitative difference between NCs and RCs beyond the number of dependencies. This interpretation aligns with previous claims made by Chomsky (1986) and Fukuda et al. (2022) that NCs exhibit

weaker island effects than RCs with respect to *wh*-movement in English and scrambling in Japanese.

7 General Discussion and Conclusions

This paper presented results from three formal acceptability judgment experiments that reexamine Japanese CNP islands with argument *wh-in-situ* using the factorial definition of island effects. The results revealed potential subliminal island effects with argument *wh-in-situ* in RCs (Experiment 1), where *wh-in-situ* inside these islands produced superadditive effects but was not completely unacceptable. However, Experiment 1 compared declarative CPs and RCs, which differ not only in terms of the possible island status but also in terms of the number of dependencies. To address the confound, Experiment 2 compared declarative CPs and NCs, which differ only in terms of the possible island status, and Experiment 3 compared NCs and RCs, both of which are potential islands but differ in the number of dependencies involved. Superadditivity was observed in both Experiments 2 and 3. Thus, possible subliminal island effects were observed with both NCs and RCs with *wh-in-situ*, while there was also a difference in the size of the effects between NCs and RCs.

Future research should include other types of dependencies that are by-hypothesis non-island sensitive in the factorial design. This would help further investigate the source of the subliminal island effects and examine the status of complex NP islands. Additionally, limitations in the current study should also be addressed, including the position of the embedded clauses in the experimental items. The items used in the three experiments contained an embedded clause in the center following the matrix subject. In the RC conditions, this pattern created a sequence in which the matrix subject was followed by the embedded object and the embedded verb. The parser might initially misanalyze the embedded object and the embedded verb as the matrix object and the matrix verb, causing a garden-path effect upon encountering the RC head. This additional effect might have influenced the results and should therefore be addressed.[3]

While the findings from the current study do not completely eliminate the possibility of other extrasyntactic factors causing superadditivity, they contribute new data that would further our understanding of crosslinguistic variation of island effects and help develop a theory that accounts for it (Sprouse and Hornstein 2013).

---

[3] The fourth experiment, which is not reported here, compared *wh-in-situ* with dependencies that are hypothesized to be non-island sensitive. The experiment also included items designed to reduce garden-path effects.

References

Almeida, D. 2014. Subliminal *Wh*-islands in Brazilian Portuguese and the Consequences for Syntactic Theory. *Revista da ABRALIN* 13 (2): 55–93.

Aoun, J. and Y.-h. A. Li. 1993. *Wh*-elements in situ: Syntax or LF? *Linguistic Inquiry* 24 (2): 199–238.

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.

Choe, J. W. 1987. LF Movement and Pied-Piping. *Linguistic Inquiry* 18 (2): 348–53.

Chomsky, N. 1986. *Barriers*. Cambridge, MA: MIT Press.

Fukuda, S., H. Ono, N. Tanaka, and J. Sprouse. This volume. Re-examining Island Effects with NP-scrambling in Japanese: The Effect of Individual Variation. *Japanese/Korean Linguistics* Vol. 30, ed. S. Williamson, A. A. Babayode-Lawal, N. Chan, S. Cho, and I. Fong. Stanford, CA: CSLI Publications.

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.

Huang, C.-T. J. 1982. Logical Relations in Chinese and the Theory of Grammar. Doctoral dissertation, MIT.

Ionin, T. 2021. Acceptability Studies in L2 Populations. In *The Cambridge Handbook of Experimental Syntax*, ed. by G. Goodall, 373–93. Cambridge: Cambridge University Press.

Ionin, T. and E. Zyzik. 2014. Judgment and Interpretation Tasks in Second Language Research. *Annual Review of Applied Linguistics* 34: 37–64.

Keshev, M. and A. Meltzer-Asscher. 2019. A Processing-Based Account of Subliminal *Wh*-island Effects. *Natural Language & Linguistic Theory* 37 (2): 621–57.

Kush, D., T. Lohndal, and J. Sprouse. 2018. Investigating Variation in Island Effects: A Case Study of Norwegian *Wh*-extraction. *Natural Language & Linguistic Theory* 36: 743–79.

Kush, D., T. Lohndal, and J. Sprouse. 2019. On the Island Sensitivity of Topicalization in Norwegian: An Experimental Investigation. *Language* 95 (3): 393–420.

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.

Lasnik, H. and M. Saito. 1984. On the Nature of Proper Government. *Linguistic Inquiry* 15 (2): 235–89.

Lee, Y.-h. 2018. Scrambling and Island Constraints in Korean: An Experimental Approach. *Linguistic Research* 35 (3): 483–511.

Lee, Y.-h. and Y. Park. 2015. Complex-NP Islands in Korean: An Experimental Approach. *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation,* ed. H. Zhao, 240–49. Shanghai, China: Association for Computational Linguistics.

Lu, J., C. K. Thompson, and M. Yoshida. 2020. Chinese *Wh-in-situ* and Islands: A Formal Judgment Study. *Linguistic Inquiry* 51 (3): 611–23.

Maxwell, S. E. and H. D. Delaney. 2003. *Designing Experiments and Analyzing Data: A Model Comparison Perspective.* Mahwah, NJ: Lawrence Erlbaum.

Nguyen, T. Q. 2021. Island Effects in Vietnamese Relativization: A Formal Judgment Study. *Proceedings of the 35th Pacific Asia Conference on Language, Information and Computation*, ed. K. Hu, J.-B. Kim, C. Zong, and E. Chersoni, 89–98. Shanghai, China: Association for Computational Linguistics.

Nishigauchi, T. 1986. Quantification in Syntax. Doctoral dissertation, University of Massachusetts, Amherst.

R Core Team. 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Richards, N. 2008. *Wh*-questions. *The Oxford Handbook of Japanese Linguistics*, ed. S. Miyagawa, 349–71. Oxford, UK: Oxford University Press.

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. and N. Hornstein. 2013. Experimental Syntax and Island Effects: Toward a Comprehensive Theory of Islands. *Experimental Syntax and Island Effects*, ed. J. Sprouse and N. Hornstein, 1–17. Cambridge, UK: Cambridge University Press.

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.

Tanaka, N. and B. D. Schwartz. 2018. Investigating Relative Clause Island Effects in Native and Nonnative Adult Speakers of Japanese. *BUCLD 42: Proceedings of the 42nd Annual Boston University Conference on Language Development*, ed. A. B. Bertolini and M. J. Kaplans, 750–63. Somerville, MA: Cascadilla Press.

Ueno, M. and R. Kluender. 2003. Event-Related Brain Indices of Japanese Scrambling. *Brain and Language* 86 (2): 243–71.

Zenker, F. and B. D. Schwartz. 2017. Topicalization from Adjuncts in English vs. Chinese vs. Chinese-English Interlanguage. *BUCLD 41: Proceedings of the 41st Boston University Conference on Language Development*, ed. M. Lamendola and J. Scott, 806–19. Somerville, MA: Cascadilla Press.