UNIVERSITY OF CALIFORNIA SAN DIEGO

From Strong to Mild: Experimental and Computational Investigations of the Relative Clause Island Effect in Japanese and English

A Dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Linguistics with a Specialization in Computational Social Science

by

Maho Takahashi

Committee in charge:

Professor Grant Goodall, Chair
Professor Ivano Caponigro
Professor Victor Ferreira
Professor Robert Kluender

2024

The Dissertation of Maho Takahashi is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2024

DEDICATION

Professor Akira Omaki

1980-2018

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

NOM	Nominative case
ACC	Accusative case
DAT	Dative case
GEN	Genitive case
TOP	Topic marker
PST	Past tense
PROG	Progressive
PRS	Present tense
PASS	Passive voice
NEG	Negative marker
COP	Copula
COMP	Complementizer
DEF	Definite marker

ACKNOWLEDGEMENTS

First and foremost, my deepest gratitude goes to my advisor Dr. Grant Goodall. The beginning of my Ph.D. journey was rocky to say the least, which led me to transfer from the University of Washington to UCSD, but he welcomed me to the program as my advisor with open arms. I am forever thankful of helping me make the transition successfully, his willingness to support the line of research that I wanted to do, and being understanding and flexible about my life situations. And thanks to his wonderful academic guidance and expertise in experimental syntax, I have been able to not only uncover so many interesting patterns of sentence acceptability through experimentation, but also make those findings more impactful by thinking about how they relate and how they contribute to syntactic theories. I would also like to thank my dissertation committee members, Dr. Ivano Caponigro, Dr. Victor Ferreira, Dr. Robert Kluender, and Dr. Michelle Yuan for their insightful feedback on my qualifying paper, which forms the basis of this dissertation, as well as the dissertation itself.

I am grateful for the wonderful community of researchers in the UCSD Experimental Syntax Lab, where I have presented my projects in varying stages of development. I especially want to thank Duk-Ho Jung, Alejandro (Alex) Rodriguez, and Joshua Wampler for always giving me helpful comments, even when I felt like I had just rambled for 45 minutes. I am also thankful for my undergraduate research assistants, Shuto Kato, Maho Miura, Sidney Ma, and Kaito Toi for their tremendous help with designing and analyzing my experiments.

I am very fortunate to have had wonderful mentors throughout my academic journey; I would like to thank Dr. Kenichi Namai from Waseda University and Dr. Satomi Mishina-Mori from Rikkyo Univeristy, who recommended that I go study Linguistics in grad school in the United States, which led me to the M.A. Linguistics program in the University of Hawai‘i

at Mānoa. It is safe to say that was the biggest turning point of my life so far. I would also like to thank Dr. William O'Grady, Dr. Kamil Deen, Dr. Amy Schafer, and Dr. Shinichiro Fukuda from UH Mānoa, who have helped me shape my interests in experimental syntax, psycholinguistics, and language acquisition, and decide which Ph.D. program would be right for me.

Even though our time together in person was not as much as it should have due to the COVID-19 pandemic, I am grateful that I am a part of the 2019 cohort, with Catherine Arnett, Matt Carter, Rachel Miles, Shai Nielson, Hagyeong Shin, and Nico Tedeschi. I hope that we all stay in touch.

I would not be writing this today without the unconditional love and support of my parents, Yoji and Mayumi Takahashi, and my brother, Rei Takahashi. I can say that I am happy with absolute certainty because they have given me the freedom to explore and pursue my passion. Last but not least, I would like to thank my husband Ryan P. Shelby for being my biggest cheerleader. I am excited about what the next chapter of our life together will bring, whether that is in Tucson, Seattle, Tokyo, or wherever.

Chapter 2, in part, has been submitted for publication of the material as it may appear in Natural Language and Linguistic Theory, Takahashi, M., and Goodall, G. The dissertation author was the primary researcher and author of this paper.

VITA

2015 Bachelor of Arts in International Liberal Studies, Waseda University

2017 Master of Arts in Linguistics, University of Hawai‘i at Mānoa

2024 Doctor of Philosophy in Linguistics with a Specialization in Computational Social Science, University of California San Diego

PUBLICATIONS

Takahashi, M. and Goodall, G. (2023). Island sensitivity with relativization in Japanese: The case of double relatives. Proceedings of the 58th annual meeting of the Chicago Linguistic Society, 455–467.

Takahashi, M. (2023). Can language models learn constraints on gap-filler dependency? Case of Japanese relative clause islands. Proceedings of the Society for Computation in Linguistics: Volume 6, Article 33.

ABSTRACT OF THE DISSERTATION

From Strong to Mild: Experimental and Computational Investigations of the Relative Clause Island Effect in Japanese and English

by

Maho Takahashi

Doctor of Philosophy in Linguistics with a Specialization in Computational Social Science

University of California San Diego, 2024

Professor Grant Goodall, Chair

This dissertation features a relative clause island, whose status has been known to differ significantly across languages and extraction types. By conducting a series of acceptability judgment experiments with human participants, as well as measuring token-by-token surprisal values among large language models, I demonstrate the following: First, sentences with relativization out of another relative clause (double relative) in Japanese, some of which have been claimed to be well-formed, still display a drop in acceptability indicative of the penalty of an island violation (Chapter 2). Second, the small penalty of relative clause island violation does not appear to be due to aggregating inter-item and/or inter-participant

variability in acceptability judgments (Chapter 3). Third, the fact that most sentences with double relative in English display a much larger drop in acceptability can be accounted for by additional factors pertaining to the structure (Chapter 4). And lastly, the acceptability of Japanese scrambling out of a relative clause, which has been judged as ill-formed without exception, can be improved (Chapter 5). Altogether, I show that the penalty of violating a relative clause island is relatively small regardless of language and extraction type, and what people have perceived as a relative clause island effect should be deconstructed into the penalty per se and a group of additional factors that are processing, syntax, and semantics-oriented.

Chapter 1

Introduction

Relative clauses have long been taken to be islands for extraction (Ross, 1967). Sentences like (1), in which the *wh*-phrase *which clothes* appears to be unable to form a dependency with a gap inside the relative clause, provide evidence for this idea.

(1) *Which clothes$_i$ do you see [$_{NP}$ the man$_j$ [$_{CP}$ that __$_j$ is wearing __$_i$]]?

It is in many ways not surprising that relative clauses seem to behave like islands. First, they are structurally very similar to noun-complement cases such as (2), and indeed, Ross considers relative clauses to be a subcase of the larger Complex Noun Phrase Constraint (CNPC).

(2) *Which clothes$_i$ do you accept [$_{NP}$ the idea [$_{CP}$ that the man is wearing __$_i$]]?

In both (1) and (2), extraction is occurring out of a complex a NP, namely an NP modified by a clause. According to CNPC, placing a gap within such a configuration is disallowed. Second, extraction out of a relative clause involves establishing a filler-gap dependency in which there is an intervening filler of a distinct filler-gap dependency. In (1), for

instance, the object gap of *wearing* is intended to be associated with the filler *which clothes*, but there is another potential filler (*the man*) that intervenes. Intervening fillers like these have long been thought to result in some degree of ill-formedness (e.g., Rizzi 1990, 2018; Villata et al. 2016), as in cases like (3).

(3) *Which clothes$_i$ do you wonder [$_{CP}$ who$_j$ __$_j$ is wearing __$_i$]?

Third, relative clauses are usually analyzed as being modifiers of the head noun and are thus not complements. If one assumes that only complement clauses are transparent for extraction, as in the contrast in (4) and along the lines of Huang (1982) and Chomsky (1986), then one would expect extraction out of relative clauses to be degraded.

(4) a. Which clothes$_i$ do you think [$_{CP}$ the man is wearing __$_i$]?
b. *Which clothes$_i$ did you read the book [$_{CP}$ before the man was wearing __$_i$]?

Given some of the most prominent ideas about what constitutes islandhood, then, relative clauses appear to be prime candidates for being islands, so the fact that they resist extraction, as we saw in (1), does not come as a surprise.

Nonetheless, cases in which extraction out of relative clauses seems to be tolerated have been noted for many years. The most famous cases involve Scandinavian languages (Danish: Erteschick-Shir, 1973; Erteschik-Shir & Lappin, 1979; Swedish: Allwood, 1982; Engdahl, 1982; 1997), but the phenomenon appears to be widespread (Hebrew: Sichel, 2018; Italian: Cinque, 2010; English: Kuno, 1976; McCawley, 1981; Chung & McCloskey, 1983). Some examples are shown in (5).

(5) a. Suppe$_i$ kender jeg mange [$_{RC}$ der kan lide _$_i$].
soup know I many who can like
'Soup, I know many people who like.'
(Danish; Erteschick-Shir, 1973)

b. Al lexem Saxor$_j$, yeS rak gvina axat$_i$
on bread black be only cheese one
[$_{RC}$ Se-keday limraox _$_i$ _$_j$].
that-worthy to.spread
'On black bread, there is only one cheese that's worth spreading.'
(Hebrew; Sichel, 2018)

c. Gianni$_i$, al quale non c'è nessuno [$_{RC}$ che
Gianni for whom NEG there.is nobody who
sia in grado di resistere_$_i$],
is able to resist
'Gianni, whom there is nobody that is able to resist, …'
(Italian; Cinque, 2010)

Relative clauses thus seem to present conflicting evidence with respect to their island status. On the one hand, they seem to behave like clear islands, as in (1), but on the other hand, they sometimes appear to be able to allow extraction, as in (5). In the literature, there have been two main strategies for addressing this conflict. In the first, it is suggested that languages may differ as to which structures are islands (Allwood, 1976, 1982; Maxwell, 1979; Andersson, 1982; Engdahl, 1982; Hawkins, 2004), so (5) represent cases where relative clauses are not islands and extraction is thus allowed. In the second, it is claimed that relative clauses are always islands, but that some form of reanalysis or alternative derivation takes place in cases like (5) that allows extraction to occur without violating the island constraint (Cinque, 2010; Kush et al., 2011; Kush et al., 2013; Sichel, 2018).

In order to evaluate the above strategies, this dissertation will focus on cases of extraction out of relative clause in Japanese, such as in the example in (6).

(6) [$_{RC2}$ [$_{RC1}$ __$_i$ __$_j$ ki-tei-ru] fuku$_j$-ga yogoretei-ru] shinshi$_i$
wear-PROG-PRS clothes-NOM get.dirty-PROG-PRS gentleman
'the gentleman$_i$ [$_{RC2}$ who the clothes$_j$ [$_{RC1}$ that __$_i$ is wearing __$_j$] are dirty]'
(Kuno, 1973)

This case involves relativization out of a (head-final) relative clause, known as a "double relative" structure. The head of the relative clause in (6), *shinshi* 'gentleman', appears to be extracted from the relative clause headed by *fuku* 'clothes' (labeled RC1), contrary to expectations if the relative clause is an island.

Through a series of acceptability judgment experiments with a factorial design (Sprouse, 2007; Sprouse et al., 2012; Sprouse & Hornstein, 2013; Sprouse & Villata, 2021), this dissertation will argue that at least in this case, both of the above strategies about islands are incorrect. That is, it will show that relative clauses are always islands in Japanese and sentences like (6) are island violations, despite initial appearances, against the proposed strategy based on examples like (5). At the same time, it will be shown that relative clauses on their own yield only mild island effects, contrary to what has traditionally been thought based on examples like (1). In cases where the island effects appear to be more severe, as they often do, I will suggest that this is because of independent factors that combine with the island violation to produce a much greater degradation in acceptability, and it is the presence and/or interaction of these independent factors that results in the observed cross-linguistic variation.

The structure of the dissertation is as follows. Chapter 2 features the case of Japanese double relatives such as (6), where it shows that the structure always triggers a small relative clause island effect, and discusses why the size seems larger for some of the sentences with double relatives. Chapter 3 examines the nature of island effects that have only a mild impact on acceptability by leveraging large language models. Chapter 4 turns to the case of English double

relatives as in (1), suggesting that the seemingly robust penalty of violating the relative clause island is due to the combination of independent factors. Chapter 5 investigates the acceptability of scrambling out of a relative clause in Japanese, which also appears to incur a strong penalty for island violations. Chapter 6 discusses the implications of the findings as well as future directions.

Chapter 2

Island sensitivity with relativization in Japanese: The case of double relatives

2.1 Introduction

Extraction out of relative clauses has been well documented in languages of East Asia, such as in the Japanese example in (1) (see Aoun & Li 2003, and Han & Kim 2004 for analogous cases in Chinese and Korean, respectively).

(1) [$_{RC2}$ [$_{RC1}$ __$_i$ __$_j$ ki-tei-ru] fuku$_j$-ga yogore-tei-ru] shinshi$_i$
wear-PROG-PRS clothes-NOM get.dirty-PROG-PRS gentleman
'the gentleman$_i$ [$_{RC2}$ who the clothes$_j$ [$_{RC1}$ that __$_i$ is wearing __$_j$] are dirty]'
(Kuno, 1973)

As has been laid out in the last chapter, previous studies have attempted to explain the cross-linguistic variation of the relative clause island effect by proposing that languages may differ as to which structures are islands (Allwood, 1976, 1982; Maxwell, 1979; Andersson, 1982; Engdahl, 1982; Hawkins, 2004), or that relative clauses are always islands, but that some form of reanalysis or alternative derivation takes place (Cinque, 2010; Kush et al., 2011; Kush et al., 2013; Sichel, 2018). According to the first strategy, (1) represent cases where relative clauses are

not islands and extraction is thus allowed. In the second, (1) must have a derivation that allows extraction to occur without violating the island constraint.

This chapter will focus on Japanese double relatives as in (1), and show that despite initial appearances, relative clauses are always islands in Japanese and sentences like (1) are island violations. In doing this, we will argue against two well-known analyses of double relatives in Japanese which propose that there is an alternative derivation in which an island violation is avoided. In one, *shinshi* in (1) is associated with *pro*, rather than a gap, and the dependency between the two is taken to be that of a full DP and a null (resumptive) pronoun, rather than an instance of canonical A'-movement (Perlmutter, 1972; Murasugi, 2000; Fukui & Takano, 2000). Since Japanese arguably allows null pronouns and dependencies between DPs and pronouns are generally not sensitive to islands, this analysis offers a plausible account of why double relatives such as (1) seem to be possible. In the second type of derivation, the gap of *shinshi* is taken to be outside of the relative clause itself (Sakai, 1994; Han & Kim, 2004; Ishizuka, 2009; Han, 2013; see Whitman, 2013 for an overview). Such an analysis is conceivable because Japanese independently allows structures in which a subject (known as a "major subject"; Kuroda, 1986; Yoon, 2007) binds *pro* in an adjacent relative clause, as in the sentence in (2).

(2) Shinshi$_i$-ga [$_{RC}$ *pro*$_i$ __$_j$ ki-tei-ru] fuku$_j$-ga yogore-tei-ru.
gentleman-NOM wear-PROG-PRS clothes-NOM get.dirty-PROG-PRS
'As for the gentleman$_i$, the clothes$_j$ [$_{RC}$ that *pro*$_i$ is wearing __$_j$] are dirty.'

If *shinshi* in (1) is associated with a gap in the position of *shinshi* in (2) (i.e., in a position outside of the relative clause itself), we would then not expect an island effect in (1).

In this study, however, we will see that analyses such as these that allow potential island violations to be circumvented appear to be incorrect; structures as in (1) show small yet significant signs of island sensitivity. That is, despite the fact that relativization in Japanese is superficially able to extract out of relative clauses, it cannot do this without violating an island constraint. In cases where the island effects appear to be more severe, as they often do, we suggest that this is because of independent factors that combine with the island violation to produce a much greater degradation in acceptability, and it is the presence or absence of these independent factors that results in the cross-linguistic variation observed. We demonstrate this through a series of acceptability experiments where structures like (1) are examined not in isolation, but as part of a full factorial design in which island effects can be detected that might not be apparent at first glance.

This chapter is organized as follows. In section 2.2, we give further background on relative clauses in Japanese. We show that in other types of extraction beyond relativization, relative clauses clearly behave like islands in Japanese, and we provide further details about double relativization. In section 2.3, we present an experiment showing that, when double relativization is tested within a full factorial design, we obtain the type of interaction that is the defining property of an island effect. In section 2.4, we conduct a second experiment showing that the interaction disappears when the A'-movement dependency is replaced with a different kind of dependency, suggesting that the interaction results from the movement itself. In section 2.5, we examine and eliminate the possibility that the sign of an interaction effect observed in section 2.3 may disappear once the comprehension of head-final relative clauses is facilitated by providing appropriate contexts. The section then shows that the interaction effect obtains both with double relatives that are relatively acceptable and with those that are relatively unacceptable

and that there is no significant difference in effect size between the two cases, suggesting that we are seeing the same island effect in both cases. We close in section 2.7 with remarks on what our results mean for the analysis of relative clauses in Japanese and for the nature of cross-linguistic variation with respect to relative clause island phenomena.

2.2. Background

2.2.1 Island status of Japanese relative clauses

As we have seen, the existence of double relatives as in (1) in Japanese raises the possibility that Japanese relative clauses are not islands, despite the traditional claims made about relative clauses in other languages. Given a broader range of evidence, however, this possibility does not seem plausible, since the behavior of other extraction types does suggest that relative clauses in fact are islands in Japanese, just as in other languages (and see Kaplan & Whitman, 1995 for evidence that Japanese relative clauses contain a CP, as in other languages). Some of this evidence comes from scrambling, where it has been noted that scrambling out of a relative clause, as in (3), is ill-formed (Haig, 1976; Saito, 1985).

(3) ?*Ano hon$_j$-o John-ga [$_{RC}$ __$_i$ __$_j$ kai-ta] hito$_i$-o
that book-ACC John-NOM write-PST person-ACC
sagashi-tei-ru rasii.
look.for-PROG-PRES seem
'That book, John seems to be looking for the person who wrote (it).'

(Saito, 1985)

Examples like (3) are more informative when they are considered in the context of a factorial design, rather than in isolation, and Fukuda et al. (2022) conducted an acceptability experiment along exactly these lines. They compared long-distance scrambling out of a non-

island, as in (4b), to long-distance scrambling out of a relative clause, as in (4d), in relation to the baseline sentences in (4a) and (4c), respectively.

(4) a. *non-island / no scrambling*

Roodookumiai-no	riidaa-wa	[$_{CP}$ kaisha-no	juuyaku-ga	oohabana
union-GEN	leader-TOP	company-GEN	executives-NOM	drastic
uriage-no	nobi-o	juugyooin-no	kyuuyo-ni	han'ee
sales-GEN	growth-ACC	employee-GEN	salary-to	reflect
sasete-i-nai-to]		hihan-shi-ta.		
make-PROG-NEG-COMP		criticize-do-PST		

'The union leader criticized that the company's executives were not making the drastic sales growth reflected in the employees' salaries.'

b. *non-island / scrambling*

Oohabana	uriage-no	nobi$_j$-o	roodookumiai-no	riidaa-wa
drastic	sales-GEN	growth-ACC	union-GEN	leader-TOP
[$_{CP}$ kaisha-no	juuyaku-ga __$_j$	juugyooin-no	kyuuyo-ni	han'ee
company-GEN	executives-NOM	employee-GEN	salary-to	reflect
sasete-i-nai-to]		hihan-shi-ta.		
make-PROG-NEG-COMP		criticize-do-PST		

'The drastic sales growth, the union leader criticized that the company's executives were not making (them) reflected in the employees' salaries.'

c. *island / no scrambling*

Roodookumiai-no	riidaa-wa	[$_{RC}$ __$_i$	oohabana	uriage-no nobi-o
union-GEN	leader-TOP		drastic	sales-GEN growth-ACC
juugyooin-no	kyuuyo-ni		han'ee	sasete-i-nai]
employee-GEN	salary-to		reflect	make-PROG-NEG
kaisha-no	juuyaku$_i$-o		hihan-shi-ta.	
company-GEN	executives-ACC		criticize-do-PST	

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

d. *island / scrambling*

Oohabana	uriage-no	nobi$_j$-o	roodookumiai-no	riidaa-wa
drastic	sales-GEN	growth-ACC	union-GEN	leader-TOP
[$_{RC}$ __$_i$ __$_j$	juugyooin-no	kyuuyo-ni	han'ee	sasete-i-nai]
	employee-GEN	salary-to	reflect	make-PROG-NEG
kaisha-no	juuyaku$_i$-o	hihan-shi-ta.		
company-GEN	executives-ACC	criticize-do-PST		

'The drastic sales growth, the union leader criticized the company's executives who were not making (them) reflected in the employees' salaries.'

Fukuda et al. observed a decline in acceptability in (4d) that is larger than would be expected given the decline seen with scrambling alone, as in (4b), or with the presence of a relative clause alone, as in (4c). That is to say, they found a significant interaction effect between the manipulated factors (presence of scrambling and relative clause as an extraction domain), such that sentences like (4d) were significantly less acceptable than would be expected given the additive effect of these factors. Such a *superadditive* effect is interpreted as the defining property of an island effect (Sprouse & Hornstein, 2013; Sprouse & Villata, 2021) and thus suggests that with respect to scrambling, relative clauses clearly seem to be islands.

Additional evidence for the island status of relative clauses in Japanese comes from *wh*-dependencies. In Japanese, these dependencies involve an in-situ *wh*-phrase and the interrogative particle *ka,* and they are thus different from the scrambling examples in that the purported movement is covert rather than overt. Tanaka and Schwartz (2018) use a factorial-design acceptability experiment and find that this dependency displays island effects with relative clauses. That is, creating a long-distance *wh*-dependency where the *wh*-phrase is inside a relative clause, as in (5d), is significantly more degraded than expected when compared to a long-distance dependency without a relative clause, as in (5b), or a relative clause without a long-distance dependency, as in (5c).

(5) a. *no wh-phrase / no relative clause*
Momoko-wa [$_{CP}$ otokonohito-ga kaban-o kat-ta-to] iimashi-ta ka?
Momoko-TOP man-NOM bag-ACC buy-PST-COMP say-PST Q
'Did Momoko say [$_{CP}$ that the man bought the bag]]?'

b. *wh-phrase / no relative clause*
Momoko-wa [$_{CP}$ otokonohito-ga nani-o kat-ta-to] iimashi-ta ka?
Momoko-TOP man-NOM what-ACC buy-PST-COMP say-PST Q
'What did Momoko say [$_{CP}$ that the man bought <what>]]?'

c. *no wh-phrase / relative clause*
Momoko-wa [$_{RC}$__$_i$ kaban-o kat-ta] otokonohito$_i$-o mimashi-ta ka?
Momoko-TOP bag-ACC buy-PST man-ACC see-PST Q
'Did Momoko see the man$_i$ [$_{RC}$ that __$_i$ bought the bag]]?'

d. *wh-phrase / relative clause*
Momoko-wa [$_{RC}$__$_i$ nani-o kat-ta] otokonohito$_i$-o mimashi-ta ka?
Momoko-TOP what-ACC buy-PST man-ACC see-PST Q
'What did Momoko see the man$_i$ [$_{RC}$ that __$_i$ bought <what>]]?'

As with the scrambling case, there is a significant superadditive interaction between the presence/absence of a long-distance dependency and the presence/absence of a relative clause. *Wh*-dependencies thus present clear evidence that relative clauses are islands in Japanese (see Lu et al., 2020 for parallel findings in Chinese).

This conclusion makes the existence of double relatives all the more mysterious, however. As we have seen, these structures appear to involve relativization out of a relative clause, but this should not be possible if relative clauses are islands. We now turn to some of the attempts to address this quandary that have been discussed in the literature.

2.2.2 Analyses of Japanese double relatives and their predictions

Double relatives in Japanese pose an analytical problem because they appear to involve A'-movement across a relative clause, as shown in (6a)[1] (repeated from (1)), despite the independent evidence just reviewed that relative clauses are islands in Japanese. As discussed in Section 1, there have been two main approaches in the literature to resolving this problem. In

[1] We will not be concerned with the exact identity of the empty category tied to the head noun via A'-movement dependency as in (6a), in particular whether it is a "true" gap or a null resumptive pronoun derived by movement, which appears to be a possibility in some languages (Borer, 1981; Koopman, 1984; Engdahl, 1985; see Hewett, 2023 for recent discussions). In either case, (6a) is predicted to show island sensitivity as it involves A'-movement out of a relative clause.

one, the dependency is claimed not to be an A'-movement dependency, as seen in (6b), where *shinshi* 'gentleman' is linked to *pro* or a resumptive pronoun in the relative clause. In the other, the gap of the movement is claimed to be in the major subject position, and not inside the relative clause itself, as in (6c).

(6) a. [$_{RC2}$ [$_{RC1}$ __$_i$ __$_j$ ki-tei-ru] fuku$_j$-ga yogoretei-ru] shinshi$_i$
wear-PROG-PRS clothes-NOM get.dirty-PROG-PRS gentleman
b. [$_{RC2}$ [$_{RC1}$ *pro*$_i$ __$_j$ ki-tei-ru] fuku$_j$-ga yogoretei-ru] shinshi$_i$
c. [$_{RC2}$__$_i$ [$_{RC1}$ *pro*$_i$ __$_j$ ki-tei-ru] fuku$_j$-ga yogoretei-ru] shinshi$_i$
'the gentleman$_i$ [$_{RC2}$ who the clothes$_j$ [$_{RC1}$ that __$_i$ is wearing __$_j$] are dirty]'

In both of these analyses (i.e. in (6b) and (6c), but not in (6a)), the double relative structure can be derived while still maintaining the claim that relative clauses are islands. Both analyses thus predict that there will be no island effect with double relatives and on the face of it, that prediction is correct; double relatives as in (6) are allowed.

Research on islands over the past decade, however, has shown that detecting island sensitivity is not as simple as noting the (un)acceptability of single sentences in isolation, since many factors can influence this acceptability in one direction or the other (Sprouse, 2007; Sprouse et al., 2012; Sprouse & Hornstein, 2013; Sprouse & Villata, 2021). Moreover, we now know that any instance of long-distance extraction, even out of a non-island, leads to a substantial decline in acceptability, so detecting an island effect means finding a case where extraction out of a particular structure leads to significantly greater degradation than one would have expected, given an equivalent instance of long-distance extraction without that structure. Doing this requires a formal sentence acceptability experiment, as exemplified by those of Fukuda et al. (2022) and Tanaka & Schwartz (2018) presented above.

In what follows, we will use sentence acceptability experiments to approach relative clauses in Japanese, with particular attention to double relatives and the fact that at least superficially, they seem to be immune to island effects. We will see that contrary to the predictions made by the analyses sketched in (6b) and (6c), double relatives in Japanese do bear the characteristic properties of island sensitivity when we submit them to experimental scrutiny. Before presenting these experiments, though, we first discuss some further relevant details of double relative structures.

2.2.3 Constraints on double relatives

Double relative structures in Japanese are not always possible, and three main kinds of restrictions have been discussed. First, double relatives seem to be more acceptable when a subject is extracted out of an object relative clause, as in (1)/(6) above, than when an object is extracted out of a subject relative clause, as in (7).

(7) *[$_{RC2}$ [$_{RC1}$ __$_i$ __$_j$ ki-tei-ru] shinshi$_i$-ga koron-da] fuku$_j$
wear-PROG-PRES gentleman-NOM fall-PST clothes
'the clothes$_j$ [that the gentleman$_i$ [that __$_i$ is wearing __$_j$] fell down]'
(Ishizuka, 2009)

Under the major subject analysis of double relatives (Sakai, 1994; Han & Kim, 2004; Ishizuka, 2009; Han, 2013), this is because the object (i.e. *fuku* 'clothes') can only be relativized from inside the relative clause and not from the major subject position. Object-extraction as in (12) thus represents a true island violation under this analysis, unlike the situation in (6)/(11), where an island effect can be avoided.

Second, the head of the outer relative clause is preferentially construed as the possessor of the head of the inner relative clause. This is true in (1)/(6), where *shinshi* 'gentleman' is understood as the possessor of *fuku* 'clothes', but when it is not true, as in (8), a significant decline in acceptability occurs.

(8) *[$_{RC2}$ [$_{RC1}$ __$_i$ __$_j$ shira-nai] syoonen$_j$-ga obore-ta] shinshi$_i$
know-NEG boy-NOM drown-PST gentleman
'the gentleman$_i$ [that the boy$_j$ [that __$_i$ does not know __$_j$] drowned]'
(Ishizuka, 2009)

Third, acceptability increases when the predicate contained in the outer relative clause is an unaccusative, passive, or adjectival predicate. In (1)/(6) above, we take the predicate of the outer relative clause *yogore-tei-ru* 'is dirty' to be unaccusative (since it does not contain a passive suffix), so it is relatively good, while in (9), the relevant predicate is transitive *kan-da* 'bit', so it is much less acceptable.

(9) *[$_{RC2}$ [$_{RC1}$ __$_i$ __$_j$ kawaigatte-iru] inu$_j$-ga tonari-no hito-o kan-da]
love-PRES dog-NOM next.door-GEN person-ACC bite-PST
shoonen$_i$
boy
'the boy$_i$ [that the dog$_j$ [that __$_i$ has taken a good care of __$_j$] bit a neighbor]'
(Ishizuka, 2009)

We will return to the first factor (whether the subject or object undergoes long-distance relativization) in Section 2.5, but for now we will simply use these three factors as guidelines to construct double relatives that are of high acceptability. We will begin by examining double

relatives that have all three of these properties and are thus maximally acceptable, but we will see that when they are explored experimentally, an island effect is nonetheless detected.

2.3 Experiment 1

We have seen that relativization out of a relative clause, resulting in so-called double relatives, appears to be sometimes acceptable in Japanese and that analyses have been proposed that potentially explain why. Under these analyses, acceptable double relatives are derived without direct extraction out of the relative clause, so no island effect is predicted. We will test this prediction here by examining acceptable double relatives within a full factorial design where we independently measure the effects of long-distance extraction and the presence of a relative clause. If there is indeed no island effect, we expect the acceptability of the double relatives to reflect simply the additive degradation associated with long-distance extraction and with a relative clause.

2.3.1 Participants

44 participants, all self-identified native Japanese speakers, were recruited on CrowdWorks, a Japanese crowdsourcing platform. 8 of them were excluded for reasons to be specified below, leaving 36 participants (age range = 20-60, mean = 41.3) whose data were analyzed. All participants reported that Japanese was their first language and that their parents primarily used Japanese to communicate with them. Participants received approximately $2 (in Japanese yen) for participation.

2.3.2 Materials

All stimuli were declarative sentences containing an embedded clause. A 2 x 2 factorial design was employed, crossing EMBEDDED CLAUSE TYPE (relative clause vs. *koto*-clause) and EXTRACTION (relativization) out of that embedded clause (+ vs. -). The embedded clause was always a part of the subject of the main clause, and a complex NP headed by the light noun *koto* 'fact' (a "*koto*-clause") was used as the baseline condition, since much previous work suggests that extracting out of a complex NP headed by *koto* does not lead to an island effect (see Fukuda & Sprouse, 2017; and Omaki et al., 2020 for discussion). *Koto*-clauses were used instead of clauses headed by the complementizer *-to* 'that' in order to facilitate lexical matching across conditions: both *koto*-clauses and relative clauses can be used naturally with the same predicates, but *-to*-clauses cannot.

The +EXTRACTION conditions all involved relativization out of the embedded clause. In order to maximize acceptability of the double relative case, the three generalizations discussed above were followed: (i) the extracted argument corresponded to the subject of the embedded clause, (ii) when the embedded clause was a relative clause, the extracted argument was naturally understood as the possessor of the head of that relative clause, and (iii) the predicate of the outer relative clause was passive (e.g., *tokushu sareta* 'to have been featured').

A sample set of stimuli is provided below.

(10)

Condition 1 EMBEDDED CLAUSE TYPE: *koto*-clause; EXTRACTION: -

[*koto* gakusha-ga	SF-shoosetsu-o	kai-ta-koto]-ga	saikin	shoten-de
professor-NOM	Sci-Fi novel-ACC	write-PST-fact-NOM	recently	bookstore-at
tokushuu-sa-re-ta.				
feature-do-PASS-PST				

'The fact that [*koto* a professor wrote a sci-fi novel] was recently featured in a bookstore.'

Condition 2 EMBEDDED CLAUSE TYPE: *koto*-clause; EXTRACTION: +

[$_{RC}$ [$_{koto}$ __$_i$	SF-shoosetsu-o	kai-ta-koto]-ga	saikin	shoten-de
	Sci-Fi novel-ACC	write-PST-fact-NOM	recently	bookstore-at
tokushuu-sa-re-ta	gakusha$_i$]-wa	hokorashige-da.		
feature-do-PASS-PST	professor-TOP	looks.proud-COP		

'The professor$_i$ [$_{RC}$ who the fact that [$_{koto}$ __$_i$ wrote a sci-fi novel] was recently featured in a bookstore] looks proud.'

Condition 3 EMBEDDED CLAUSE TYPE: relative clause; EXTRACTION: -

[$_{RC}$ gakusha-ga __$_j$	kai-ta]	SF-shoosetsu$_j$-ga	saikin	shoten-de
professor-NOM	write-PST	Sci-Fi novel-NOM	recently	bookstore-at
tokushuu-sa-re-ta.				
feature-do-PASS-PST				

'The sci-fi novel$_j$ [$_{RC}$ that the professor wrote __$_j$] was recently featured in a bookstore.'

Condition 4 EMBEDDED CLAUSE TYPE: relative clause; EXTRACTION: +

[$_{RC2}$ [$_{RC1}$ __$_i$ __$_j$	kai-ta]	SF-shoosetsu$_j$-ga	saikin	shoten-de
	write-PST	Sci-Fi novel-NOM	recently	bookstore-at
tokushuu-sa-re-ta	gakusha$_i$]-wa	hokorashige-da.		
feature-do-PASS-PST	professor-TOP	looks.proud-COP		

'The professor$_i$ [$_{RC2}$ who the sci-fi novel$_j$ [$_{RC1}$ that __$_i$ wrote __$_j$] was recently featured in a bookstore] looks proud.'

20 lexically-matched sets as in (10) were created using one of 4 passive verbs (*tokushuu-sa-re-ta* 'was featured', *shuzai-sa-re-ta* 'was interviewed', *happyou-sa-re-ta* 'was announced', *kouhyou-sa-re-ta* 'was disclosed'). These verbs all allow a *koto*-clause subject (Condition 1), an animate subject (Condition 2), or an inanimate subject (Conditions 3 and 4) equally naturally. Stimuli were counterbalanced through a Latin-square procedure, resulting in 4 lists (5 items per condition; 20 items per list). 40 fillers were also created, consisting of sentences with varying degrees of acceptability: 10 fillers of expected high acceptability, 10 of intermediate acceptability (e.g., sentences with center-embedding), and 20 of low acceptability (e.g., sentences violating the Coordinate Structure Constraint). Fillers were identical across lists, and each of the lists consisted of 60 items. The order of stimuli was pseudo-randomized such that

two critical items never appeared in a row. The full set of stimuli for the experiments in this chapter is available at the following Open Science Framework page: https://osf.io/g9pmu/.

2.3.3 Procedures

The experiment was hosted on Ibex farm (Drummond, 2013). Participants were instructed to rate how natural each sentence sounded by clicking on a number on a scale from 1 (very unnatural) to 7 (very natural). Participants also completed a brief language background questionnaire.

To screen out participants who were not attending to the task, responses to the 10 filler items with the highest mean acceptability scores across all participants and the 10 with the lowest scores were identified. Participants whose ratings were more than 2 standard deviations away from the mean for 5 or more of these 20 items were excluded from further analysis. Two participants were filtered out in this way. In addition, a server error resulted in over-recruitment of participants for one of the lists, which led us to exclude the last 6 participants in that list (as determined by their submission date), in order to maintain counterbalancing across lists. The final dataset consisted of 9 participants in each of the 4 lists, so 36 in total.

2.3.4 Data analysis

Raw acceptability scores were converted to z-scores prior to analysis, in accord with standard practice (Goodall, 2021). A linear mixed-effects regression model was created using the lmerTest package (Kuznetsova et al., 2017) in R (R Core Team, 2024). Since the maximal model (as recommended in Barr et al., 2013) did not converge, we used random intercepts and random slopes of the two manipulated factors for participant and item, and

random slopes of their interaction only for participant.

2.3.5 Predictions

Long-distance extraction is known to cause a significant decline in acceptability (Frazier & Clifton, 1989; Kluender & Kutas, 1993; Alexopoulou & Keller, 2007; Fanselow, 2021; Goodall, 2021), so we expect the +EXTRACTION cases (Conditions 2 and 4) to be significantly less acceptable than the - EXTRACTION cases (Conditions 1 and 2). If there is no island effect, there should be no interaction between EXTRACTION and EMBEDDED CLAUSE TYPE, and the decline should occur in parallel in both the *koto*-clause and relative clause cases. If there is an island effect, on the other hand, we would expect an interaction between the two factors, with a greater decline in acceptability with extraction out of a relative clause than with extraction out of a *koto*-clause. These two possible outcomes are illustrated in Figure 2.1. The analyses of double relatives that we have examined, one positing *pro* in place of a gap (e.g. (6b)) and the other positing a gap in the major subject position (e.g. (6c)), claim that there is no extraction out of the relative clause, so they do not predict an island effect. We would thus expect results similar to the left panel in Figure 2.1. If extraction is directly out of the relative clause, as in (6a), we then expect results similar to the right panel in Figure 2.1.

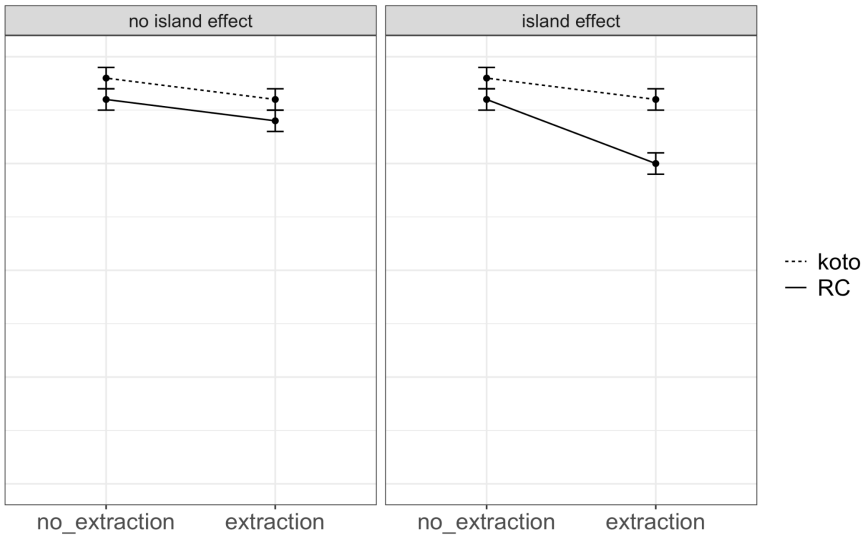

Figure 2.1: Predicted outcome if there is no island effect (left) vs. if there is an island effect (right)

2.3.6 Results

Mean z-scores for the four conditions of Experiment 1 are presented in Figure 2.2.

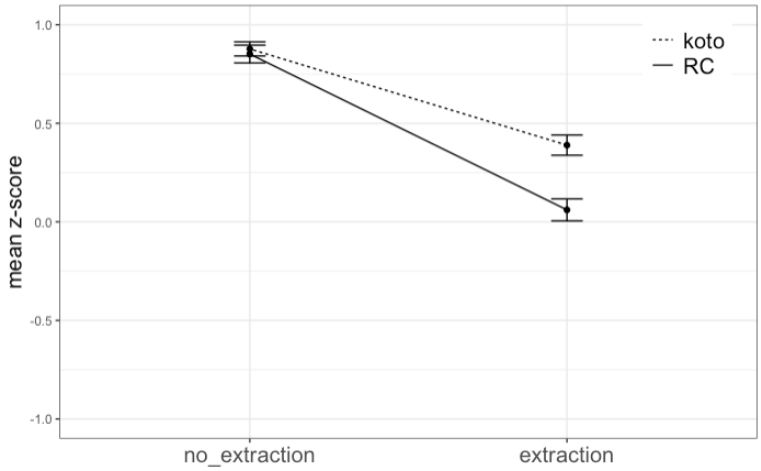

Figure 2.2: Aggregated responses to critical items in Experiment 1

The model revealed a significant main effect of EXTRACTION, such that +EXTRACTION sentences were rated lower than their non-extracted counterparts, as expected (β = -0.49, SE =

0.07, $p < 0.001$). There was no main effect of EMBEDDED CLAUSE TYPE (β = -0.03, SE = 0.08, p = 0.76). Crucially, the interaction between these two factors was significant (β = -0.30, SE = 0.09, $p < 0.01$). That is, the decline in acceptability for extraction out of a relative clause was significantly greater than the decline for extraction out of a *koto*-clause. As seen in the right panel in Figure 2.1, this is the pattern associated with an island effect.

2.3.7 Discussion

The results of Experiment 1 suggest that even when sentences are constructed to make double relatives maximally acceptable, we still find the kind of interaction associated with an island effect. This suggests, at least at first blush, that double relatives in Japanese are what they appear to be: extraction out of a relative clause. Although of relatively high acceptability compared with many other instances of island violation, extraction out of a relative clause incurs an additional penalty over and above extraction out of a *koto*-clause.

The size of the interaction effect may be calculated by means of a differences-in-differences (DD) score (Sprouse et al., 2012), in which the difference between the two -EXTRACTION conditions is subtracted from the difference between the two +EXTRACTION conditions. In this experiment, the DD score is 0.302, which is within the range observed for island effects in Sprouse and Villata (2021), though at the low edge of that range.

A superadditive interaction as observed here is standardly taken to be a signature of A'-movement, given that A'-movement dependencies show this effect and other dependency types do not (e.g., binding relations; Ross, 1967; Yoshida et al., 2014). If this is correct, then non-A'-movement dependencies should not show this same type of interaction even when they are used

in the same environment and under the same experimental conditions. We test this prediction in Experiment 2.

2.4 Experiment 2

In Experiment 1, we saw that relativization in Japanese shows a superadditive interaction between extraction and the type of embedded clause that has commonly been interpreted as indicative of an island effect. This effect is characteristic of A'-movement, but here we explore whether a superficially similar dependency will show a similar effect. We use the reflexive anaphor *jibun* 'self' (also romanized in the literature as *zibun*) since it resembles relativization in two important ways: the dependency between *jibun* and its antecedent can cross clause boundaries and the antecedent can appear to the right of *jibun*. If this dependency, like relativization, shows an interaction when tested experimentally, it would suggest that the superadditivity we observe is perhaps an artifact of the experiment rather than a result of the A'-movement per se. If no such interaction obtains, however, it would strengthen the conclusion that the superadditivity seen in Experiment 1 is a consequence of A'-extraction out of an island.

2.4.1 Participants

A new group of 37 self-identified native Japanese speakers was recruited on CrowdWorks. All were compensated approximately $2 for their participation. The same attention check procedure as in Experiment 1 was used and one participant was excluded in this way, resulting in 36 participants (age range = 19-56, mean = 37.5) whose responses were analyzed.

2.4.2 Materials

The design was very similar to that of Experiment 1, but with a factor of JIBUN (i.e. presence or absence of a backward anaphoric dependency between *jibun* and its referent) instead of EXTRACTION (relativization). For *koto*-clauses containing *jibun*, care was taken in choosing the predicate of which the clause is predicated, given restrictions that have been noted in this area (Oshima, 2009; Kishida, 2011).[2] The referent for *jibun* in some items was modified so that the intended referent was always singular in the experimental items. Sample stimuli are provided in (11).

(11)

Condition 1 EMBEDDED CLAUSE TYPE: *koto*-clause; JIBUN: -

[*koto* gakusha-ga SF-shoosetsu-o kai-ta-koto]-ga saikin shoten-de
professor-NOM Sci-Fi novel-ACC write-PST-fact-NOM recently bookstore-at
tokushuu-sa-re-ta.
feature-do-PASS-PST
'The fact that [*koto* a professor wrote a sci-fi novel] was recently featured in a bookstore.'

Condition 2 EMBEDDED CLAUSE TYPE: *koto*-clause; JIBUN: +

[RC [*koto* jibun$_i$-ga SF-shoosetsu-o kai-ta-koto]-ga saikin shoten-de
self-NOM Sci-Fi novel-ACC write-PST-fact-NOM recently bookstore-at
tokushuu-sa-re-ta] gakusha$_i$-wa hokorashige-da.
feature-do-PASS-PST professor-TOP looks.proud-COP
'The professor$_i$ [RC who the fact that [*koto* self$_i$ wrote a sci-fi novel] was recently featured in a bookstore] looks proud.'

Condition 3 EMBEDDED CLAUSE TYPE: relative clause; JIBUN: -

[RC gakusha-ga __$_j$ kai-ta] SF-shoosetsu$_j$-ga saikin shoten-de
professor-NOM write-PST Sci-Fi novel-NOM recently bookstore-at
tokushuu-sa-re-ta.
feature-do-PASS-PST
'The sci-fi novel$_j$ [RC that the professor wrote __$_j$] was recently featured in a bookstore.'

[2] In sentences with backward anaphoric dependencies involving *jibun* inside a *koto*-clause, the predicate of the *koto*-clause has to denote a mental state such as *hokorashige* 'looks proud' in (16). Such a restriction on the type of predicates does not seem to exist when *jibun* is inside a relative clause. See Section 5.1 of Kishida (2011) for further discussion on the source of the restriction.

Condition 4 EMBEDDED CLAUSE TYPE: relative clause; JIBUN: +

[$_{RC2}$ [$_{RC1}$ jibun$_i$-ga	__$_j$ kai-ta]	SF-shoosetsu$_j$-ga	saikin	shoten-de
	write-PST	Sci-Fi novel-NOM	recently	bookstore-at
tokushuu-sa-re-ta]	gakusha$_i$-wa	hokorashige-da.		
feature-do-PASS-PST	professor-TOP	looks.proud-COP		

'The professor$_i$ [$_{RC2}$ who the sci-fi novel$_j$ [$_{RC1}$ that self$_i$ wrote __$_j$] was recently featured in a bookstore] looks proud.'

Conditions without *jibun* (Conditions 1 and 3) are analogous to Conditions 1 and 3 in Experiment 1 (i.e., those without extraction), and a similar relationship exists between Conditions 2 and 4 in the two experiments (they contain extraction in Experiment 1 and a backward anaphoric dependency with *jibun* in Experiment 2). The current experiment used the same set of fillers, as well as the pseudo-randomization method, as Experiment 1.

2.4.3 Procedures

The same experimental procedures were used as in Experiment 1.

2.4.4 Data analysis

As in Experiment 1, a linear mixed-effect model with random effects of subject and item was used to fit the data and test for significance. The model predicts the acceptability (in z-scores) of sentences as a function of backward anaphoric dependency across an embedded clause and whether the embedded clause is a relative clause or a *koto*-clause. The maximal random model did not converge, and thus the resulting model takes into account the random intercepts and random slopes of the two factors for participant and item, and random slopes of their interaction only for participant.

2.4.5 Predictions

If dependencies involving *jibun* are immune to islands, then we should find no interaction between JIBUN and EMBEDDED CLAUSE TYPE, in a way similar to the left panel in Figure 2.1. If this type of dependency is sensitive to islands, on the other hand, then there should be an interaction between the two factors, where a backward anaphoric dependency with *jibun* across a relative clause would be judged as less acceptable than it would be given the simple combination of relative clause and the dependency, along the lines of the right panel in Figure 2.1.

2.4.6 Results

The mean z-scores for the four conditions are presented in Figure 2.3. The model revealed a main effect of JIBUN (β = -0.46, SE = 0.09, $p < 0.001$), i.e. sentences with a dependency involving *jibun* were significantly less acceptable than those without such a dependency. There was no main effect of EMBEDDED CLAUSE TYPE (β = -0.07, SE = 0.08, p = 0.35). Most notably, however, the interaction between EMBEDDED CLAUSE TYPE and JIBUN was not significant (β = 0.02, SE = 0.08, p = 0.79). This is different from what we saw in Experiment 1, where there was a significant interaction between EMBEDDED CLAUSE TYPE and EXTRACTION. The pattern illustrated on Figure 2.3 thus resembles the left panel of Figure 2.1, which represents the expected results when there is no island violation.[3]

[3] A reviewer pointed out that testing two structures (double relatives with a gap and *jibun*) in a single experiment would have certain advantages, compared with doing so in two separate experiments, as we have done here. However, we ensured that the results from the two experiments are comparable by reusing the identical set of fillers from Experiment 1, so that any difference in the judgments of the critical items in Experiment 2 would not be due to different fillers. Moreover, to confirm that participants in the two experiments evaluated the fillers similarly, we ran a Pearson's correlation test and found that the judgments of fillers in the two experiments were strongly correlated, $r(35) = .99$, $p < .001$. Hence, we are confident that we would obtain parallel results in an experiment where double relatives with a gap and *jibun* are evaluated by the same group of participants.

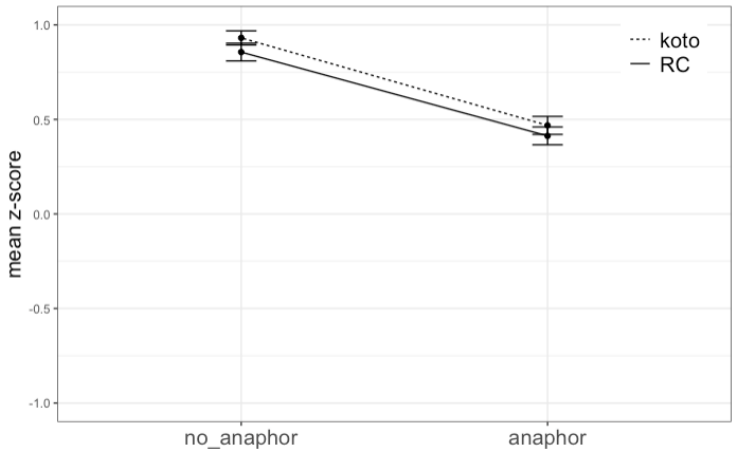

Figure 2.3: Aggregated responses to critical items in Experiment 2

2.4.7 Discussion

The aim of Experiment 2 was to test whether the superadditive interaction that we saw in Experiment 1 is found more generally in non-A'-movement dependencies when they are used in the same environment and under the same experimental conditions. We tested this by examining backward anaphoric dependencies with *jibun* and the results show that this dependency does not exhibit the type of interaction that we saw in Experiment 1. That is, in Experiment 1, there was a significant interaction between EMBEDDED CLAUSE TYPE and EXTRACTION, while in Experiment 2, there is no such interaction between EMBEDDED CLAUSE TYPE and JIBUN. Put more simply, relativization shows island-sensitivity and backward anaphoric dependencies with *jibun* do not. Given that island sensitivity is a known property of A'-movement dependencies, this strongly suggests that relativization in Japanese results from A'-movement.[4]

[4] As a reviewer points out, *jibun* in Condition 4 could be analyzed as an overt resumptive pronoun that is A'-bound by the head of the outer relative clause. If that is the case, though, Experiment 2 suggests that this type of resumption is not sensitive to islands. Another possible analysis is that *jibun* is a long-distance anaphor, as is widely assumed, and that it is taking the head of the outer relative clause as its antecedent in Condition 4. This would mean that the relative clause has no gap, but this is widely taken to be a general property of Japanese, as in examples such as (ii).

(ii)	[atama-ga	yoku-nar-u]	hon
	head-NOM	better-become-PRS	book

The results from Experiments 1 and 2 thus differ in terms of the interaction between the two factors, but they nonetheless show an important similarity: The main effect observed for JIBUN in Experiment 2 is reminiscent of the main effect for EXTRACTION seen in Experiment 1. As noted earlier, the latter effect is presumably due to the increased processing cost associated with long-distance extraction. The effect seen here for JIBUN plausibly has a very similar source, given that predicting the antecedent of *jibun* across a long distance has been claimed to increase the processing burden in a detectable way (Aoshima et al., 2009). In both experiments, then, there is sensitivity to dependency distance, but only in Experiment 1 is there sensitivity to islands.

The issue of whether or not island effects can be observed with backward binding dependencies has been explored in previous studies; Yoshida et al. (2014) discovered that a backward dependency between a pronoun and its referent is not sensitive to relative clause islands in English, so long as it does not violate the Binding Conditions (Chomsky, 1981). In contrast, Keshev and Meltzer-Asscher (2019) demonstrate that a backward dependency in Hebrew does result in a superadditive effect when it is formed across a *wh*-island. The question of the possible island sensitivity of backward binding dependencies is clearly worth exploring more, but our results suggest that backward anaphora does not show such sensitivity.

As we saw in section 2.3, double relatives are highly sensitive to structural factors that increase or decrease their acceptability, but so far, we have only examined those that are configured to maximize acceptability. We will now turn to double relatives that are less acceptable, looking in particular at long-distance relativization of the subject versus the object, and the extent to which

---

'the book (by reading which) one becomes smarter' (Mikami, 1963)

Regardless of the exact analysis of *jibun*, though, the lack of an interaction effect in Experiment 2 clearly suggests that the dependency between *jibun* and its antecedent is not based on A'-movement.

this affects the size of the island effect. In addition to the structural factors, we will also feature a contextual factor and examine whether presenting our stimuli with a context that would facilitate the processing of relative clauses would affect the acceptability of double relatives.

2.5 Experiment 3

We found evidence in Experiment 1 that relative clauses seem to be islands in Japanese with respect to relativization. Long-distance relativization induces some degradation even out of a non-island, but relativization out of a relative clause results in the superadditive interaction that is associated with an island effect, though superadditivity by itself is not a satisfactory condition. This was *prima facie* evidence against analyses in which double relatives do not involve extraction out of the relative clause, since no standard island effect is predicted by these analyses.

Nevertheless, the major subject analysis might lead one to expect some amount of superadditivity. This analysis was seen in (6c), repeated here as (12).

(12) [$_{RC2}$ __$_i$ [$_{RC1}$ *pro*$_i$ __$_j$ ki-tei-ru] fuku$_j$-ga yogore-tei-ru]
wear-PROG-PRS clothes-NOM get.dirty-PROG-PRS
shinshi$_i$
gentleman
'the gentleman$_i$ [$_{RC2}$ who the clothes$_j$ [$_{RC1}$ that__$_i$ is wearing __$_j$] are dirty]'

In this analysis, *shinshi* in (12) is relativized directly from the major subject position, outside of the relative clause. Unlike a regular subject that is licensed by *v*, a major subject is licensed via an aboutness relation with the rest of the sentence (Saito, 1982; Heycock, 1993; Vermeulen, 2005). This additional layer of structure could degrade acceptability relative to the

other conditions in the experiment, which do not have this structure. Structural additions like this typically have only a small effect on acceptability compared to the much larger effects induced by long-distance dependencies, as will be discussed in more detail upon presenting the results, but it is possible that it would be detectable. And if it is, this would result in some superadditivity.

Could structural addition, rather than extraction out of an island, account for the superadditivity in Experiment 1? One way to test this is to compare the amount of superadditivity we obtain when extracting a subject versus extracting an object. Under the major subject analysis, extracting a subject is often possible (since it involves extraction from the major subject position rather than from inside the relative clause), while extracting an object is not (since it would involve extraction from inside the relative clause, inducing an island effect), as discussed in section 2.3. According to this analysis, then, a double relative with long-distance subject extraction could lead to a small amount of superadditivity, due to the additional structure needed to have a major subject, but a double relative with long-distance object extraction should lead to a larger amount of superadditivity, of the type typically seen with island effects.[5] In contrast, an analysis that posits extraction from inside the relative clause for both types of double relative (henceforth *uniform extraction analysis*), means that a double relative with long-distance subject extraction and one with long-distance object extraction should lead to a similar degree of superadditivity.

[5] Another possible analysis of the two types of double relatives, as suggested by a reviewer, is based on resumption; a double relative with long-distance subject extraction could involve a null resumptive pronoun co-indexed with the outer relative clause head, which leads to a small amount of superadditivity. This is because cross-linguistically, resumption is known to cause some amount of degradation (Alexopoulou & Keller, 2007; Heestand et al., 2011; Han et al., 2012; Polinsky et al., 2013). In contrast, a double relative with long-distance object extraction involves a true gap, which leads to a standard island effect and a larger amount of superadditivity. This analysis, however, is challenged by the fact that null pronouns are available in both subject and object positions in Japanese, which makes it unclear why resumption would be unavailable in the case of object-extraction double relatives.

Aside from evaluating the two analyses, we also examine whether the acceptability of two types of double relative is susceptible to the effect of context that precedes them. It is possible that the superadditive effect observed in Experiment 1 was due to the difficulty of identifying head-final relative clauses as such in sentence processing, given the gap-filler order and the lack of an overt relative marker in Japanese relative clauses. Some studies have attempted to facilitate the comprehension of both head-initial and head-final relative clauses, by presenting participants with a certain type of context that takes into account the pragmatic function of relative clauses. Lin and Bever (2010) claim that the content of relative clauses is typically something that is already familiar to the hearer, and it is supposed to aid the parser in identifying the referent (i.e., head noun of a relative clause) among all the referents previously mentioned. Therefore, a context that motivates the parser to expect a relative clause would be one where "a small set of referents compete to be selected as the topic of the text that follows" (p.290). As a test case, Lin and Bever introduce Hsu et al.'s (2008) study on the comprehension of Chinese relative clauses, which are head-final just like Japanese relative clauses. Hsu et al. found that, when participants were presented with a context where referents of the same kind were contrasted as in (13), their comprehension of a subsequent relative clause was faster than when a context did not involve multiple referents of the same kind.

(13) a. *Context with two referents of the same kind*: The college student upstairs has two motorcycles. He does not maintain one of the motorcycles, but he maintains the other motorcycle very carefully.

b. *Target sentence:* This semester, the motorcycle that the college student maintained carefully was stolen.

It is therefore worth testing whether a context that motivates a relative clause in a subsequent sentence (henceforth *contrastive context* as such a context contrasts multiple referents of the same kind) mitigates the cost of reanalyzing a simple clause as a relative clause, and whether the resulting increase in acceptability is enough to eliminate the superadditive effect caused by double relativization.

Experiment 3 presented next is divided into two parts. Experiment 3a seeks to confirm that it has to be specifically the contrastive context, instead of any kind of context, that improves the acceptability of double relatives with long-distance subject and object extraction. Experiment 3b then measures the size of the interaction (i.e., the amount of superadditivity) with long-distance subject and object extraction in double relatives, presented along with the contrastive context.

2.5.1 Experiment 3a

2.5.1.1 Participants

42 self-identified native Japanese speakers were recruited on CrowdWorks. Data from 6 of them were excluded based on two criteria: First, their responses to the fillers, as in Experiment 1 and 2. Second, the rate of correct responses to the comprehension questions following some of the trials, the details of which are provided in the Materials section. Data from 36 participants (age range = 22-63, mean = 39.3) were thus included in the data analysis. The amount of payment was the same (approx. $2) as in Experiments 1 and 2.

2.5.1.2 Materials

The stimuli were created using a 2 x 2 factorial design, crossing EXTRACTION TYPE (long-distance subject vs. object extraction) and CONTEXT TYPE (contrastive vs. non-contrastive). With "subject-extraction," the subject is extracted out of a relative clause whose head originates in the object position (as we already saw in Experiment 1), and with "object-extraction," the object is extracted out of a relative clause whose head originates in the subject position. A contrastive context introduces four referents of two kinds (e.g., two professors, two novels), as a subsequent sentence with double relativization involves two relative clauses. A non-contrastive context describes an event that is semantically and temporally coherent with the event of the subsequent sentence. Critically, a non-contrastive context neither mentions nor contrasts the referents introduced in its contrastive counterpart. An example of each type of the contexts is the following:

(14) a. *Contrastive context*: Two professors in a college in Tokyo wrote a sci-fi novel. One's sci-fi novel was featured in a bookstore, but the other's wasn't.

b. *Non-contrastive context*: College students from a regional university who are members of a literature club went to a book fair, where they could meet lots of best-selling authors in person.

The subsequent sentence to be evaluated by participants includes either a subject- or object-extraction double relative. Sentences with subject-extraction double relatives are identical to the stimuli constructed for Experiment 1. Sentences with object-extraction double relatives were derived from their subject-extraction counterparts by switching the head nouns and adjusting the main clause predicate accordingly (e.g., "(the professor) looks proud" to "(the novel) is popular among young people"). The two types of double relative are exemplified below.

(15) a. *subject-extraction double relative (=Condition 4 of (10))*
[$_{RC2}$ [$_{RC1}$ __$_i$ __$_j$ kai-ta] SF-shoosetsu$_j$-ga saikin
write-PST Sci-Fi novel-NOM recently
shoten-de tokushuu-sa-re-ta] gakusha$_i$-wa hokorashige-da.
bookstore-at feature-do-PASS-PST professor-TOP looks.proud-COP
'The professor$_i$ [$_{RC2}$ who the sci-fi novel$_j$ [$_{RC1}$ that __$_i$ wrote __$_j$] was recently featured in a bookstore] looks proud.'

b. *object-extraction double relative*
[$_{RC2}$ [$_{RC1}$ __$_i$ __$_j$ kai-ta] gakusha$_i$-ga saikin shoten-de
write-PST professor-NOM recently bookstore-at
tokushuu-sa-re-ta] SF-shoosetsu$_j$-wa daigakusee-ni ninki-da.
feature-do-PASS-PST Sci-Fi novel-TOP college.students-DAT popular-COP
'The sci-fi novel$_j$ [$_{RC2}$ who the professor$_i$[$_{RC1}$that __$_i$ wrote __$_j$] was recently featured in a bookstore] is popular among college students.'

20 sentences with object-extraction double relatives lexically matching the subject-extraction were created. In addition, while the same 40 fillers from the previous experiments were used, a context was generated for each of them so as to make them similar to the critical items. Contexts for the fillers were similar to the non-contrastive context for the critical items in that they depict an event or a background that is coherent with the subsequent filler. (16) below exemplifies the context for a filler.

(16) *Context*: The job of a primary school teacher involves not only teaching, but also intervening in class fights and stopping students from acting up.
Sentence: A naughty boy switched his classmate's pencil case with his.

Critical items with 4 unique combinations of CONTEXT TYPE and EXTRACTION TYPE (contrastive context -> subject-extraction double relative, contrastive -> object-extraction, non-contrastive -> subject-extraction, non-contrastive -> object-extraction) were divided into 4 lists using a Latin square design (with 5 items from each combination), and the same set of fillers was used across lists. The resulting list of stimuli consists of 20 critical items + 40 fillers = 60 in total

(with each item consisting of a context and a test sentence).

2.5.1.3 Procedures

This experiment was hosted on PCIbex (Zehr & Schwarz, 2018), as the original Ibex Farm that hosted Experiment 1 and 2 permanently suspended its service in September 2021. Each trial in the experiment started with the presentation of a context, and a button to proceed appeared with a 3-second delay to encourage the participants to read the context. After clicking the button, the context disappeared and the test sentence with a 7-point scale was presented. On 8 out of 60 trials, a comprehension question (Yes/No) appeared after participants provided their acceptability judgment. The question could be answered easily if they were paying attention to the context, as (17) exemplifies.

(17) [After the presentation of (16)] Does the story mention a college professor? (correct answer: No)

Those who answered less than 5 out of the 8 questions correctly (i.e., less than 62.5% correct response rate) were excluded from analysis. 2 participants were detected with this criterion, and their data were excluded along with 4 participants whose performance deviated significantly from the average on the selected fillers. After recruiting additional participants to make up for the lost data, we obtained the final dataset of 36 participants.

2.5.1.4 Data analysis

A linear mixed-effect model with random effects of participant and item was built to fit the data and test for significance. The model predicted the acceptability (converted to z-scores) of sentences as a function of CONTEXT TYPE and EXTRACTION TYPE. The maximal random model

did not converge, and thus the model took into account the random intercepts for participant and item, random slopes of the two manipulated factors as well as their interaction for participant, and random slopes of double relative type for item.

2.5.1.5 Results

The z-scores for the fillers followed our prediction (-0.70 for bad, -0.15 for intermediate, 1.32 for good). As illustrated in Figure 2.4, the model revealed a main effect of EXTRACTION TYPE (β = 0.62, SE = 0.12, $p < 0.001$), and critically, a main effect of CONTEXT TYPE (β = -0.24, SE = 0.08, $p < 0.01$) such that the average acceptability of sentences with double relatives was higher when preceded by a contrastive context in comparison to a non-contrastive context. The interaction effect between the two factors was not significant (β = 0.16, SE = 0.12, p = 0.18), suggesting that the extent to which a contrastive context improves acceptability was comparable among sentences with subject- and object-extraction double relatives.

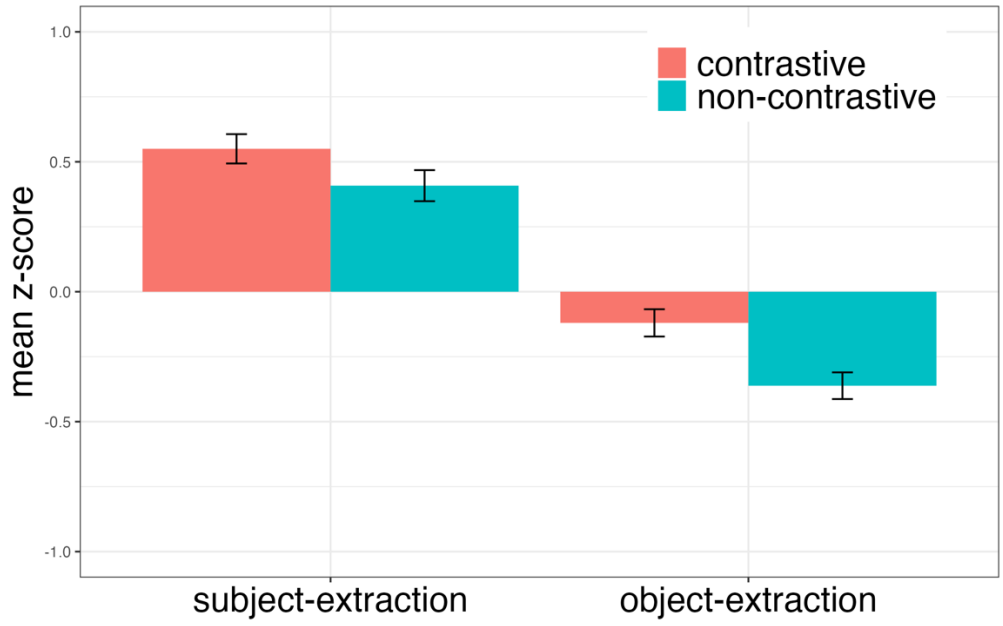

Figure 2.4: Aggregated acceptability of critical items from Experiment 3a

2.5.1.6 Discussion

The present experiment validates that providing a certain type of context that motivates a relative clause in a subsequent sentence, makes it easier to parse a subsequent sentence with double relative (and hence results in a higher acceptability of the sentence). Furthermore, our results are in line with previous judgments that sentences with object-extraction double relatives are lower in acceptability than the subject-extraction ones.

Having confirmed the positive effect of contrastive context on the acceptability of double relatives, we will move on to test whether such contexts are sufficient to eliminate a superadditive effect exhibited by the two types of double relatives.

2.5.2 Experiment 3b

2.5.2.1 Participants

A new group of 51 self-identified native Japanese speakers was recruited on CrowdWorks. Data from 3 participants were excluded using the same attention check procedure using fillers as in Experiments 1 and 2, resulting in 48 participants whose data were analyzed (age range = 24-67, mean = 43.3). Compensation to the participants was increased to approximately $3 (from $2 in the previous experiments) given the larger number of trials in the current experiment.

2.5.2.2 Materials

A 2 x 2 x 2 factorial design was employed, crossing EMBEDDED CLAUSE TYPE (*koto*-clause vs. relative clause), EXTRACTION (relativization) out of that embedded clause (+ vs. -), and EXTRACTION TYPE (subject-extraction vs. object-extraction). Compared to Experiment 1, the

number of conditions doubled (4 to 8) as a result of having both extraction types. Another difference from Experiment 1 is that each test sentence is now preceded by a context. All the contexts for critical items were contrastive. We used the identical set of fillers as well as their accompanying contexts from Experiment 3a.

The next few paragraphs walk through each condition and its corresponding context. The first condition, which involves double relatives, is exemplified below.

(18) Condition 4 EMBEDDED CLAUSE TYPE: relative clause; EXTRACTION: +
Context: Two professors in a college in Tokyo wrote a sci-fi novel. One's sci-fi novel was featured in a bookstore, but the other's wasn't.

a. EXTRACTION TYPE: subject-extraction[6]
[$_{RC2}$ [$_{RC1}$ __$_i$ __$_j$ kai-ta] SF-shoosetsu$_j$-ga saikin shoten-de
write-PST Sci-Fi novel-NOM recently bookstore-at
tokushuu-sa-re-ta] gakusha$_i$-wa hokorashige-da.
feature-do-PASS-PST professor-TOP looks.proud-COP
'The professor$_i$ [$_{RC2}$ who the sci-fi novel$_j$ [$_{RC1}$ that __$_i$ wrote __$_j$] was recently featured in a bookstore] looks proud.'

b. EXTRACTION TYPE: object-extraction
[$_{RC2}$ [$_{RC1}$ __$_i$ __$_j$ kai-ta] gakusha$_i$-ga saikin shoten-de
write-PST professor-NOM recently bookstore-at
tokushuu-sa-re-ta] SF-shoosetsu$_j$-wa daigakusee-ni
feature-do-PASS-PST Sci-Fi novel-TOP college.students-DAT
ninki-da.
popular-COP
'The sci-fi novel$_j$ [$_{RC2}$ who the professor$_i$[$_{RC1}$that __$_i$ wrote __$_j$] was recently featured in a bookstore] is popular among college students.'

As established in Experiment 3a, contexts in Condition 4, which involves double relatives, introduce and contrast four entities of two kinds. The two conditions without an extraction out of a relative clause are displayed next.

[6] The labels "subject-extraction" and "object-extraction" mean that the "subject-extraction" condition corresponds to sentences in Condition 4 with subject-extraction out of an object relative clause, and the "object-extraction" condition corresponds to the ones with object-extraction out of a subject relative clause.

(19) Condition 3 EMBEDDED CLAUSE TYPE: relative clause; EXTRACTION: -

Context: Two sci-fi novels were published. One of the sci-fi novels was written by a professor, but the other was written by an actor.

a. EXTRACTION TYPE: subject-extraction

[RC gakusha-ga __j kai-ta] SF-shoosetsu$_j$-ga saikin shoten-de
professor-NOM write-PST Sci-Fi novel-NOM recently bookstore-at
tokushuu-sa-re-ta.
feature-do-PASS-PST
'The sci-fi novel$_j$ [RC that the professor wrote __j] was recently featured in a bookstore.'

Context: There are two professors in a college in Tokyo. One of them wrote a sci-fi novel, and the other one wrote an academic journal.

b. EXTRACTION TYPE: object-extraction

[RC __i SF-shousetsu-o kai-ta] gakusha$_i$-ga saikin shoten-de
Sci-Fi novel-ACC write-PST professor-NOM recently bookstore-at
tokushuu-sa-re-ta.
feature-do-PASS-PST
'The professor$_i$ [RC who __i wrote a sci-fi novel] was recently featured in a bookstore.'

The next conditions exemplified below involve relativization across a *koto*-clause.

(20) Condition 2 EMBEDDED CLAUSE TYPE: *koto*-clause; EXTRACTION: +

Context: Two professors each wrote a sci-fi novel. The fact that one of them wrote one was featured in a bookstore, but the fact that the other wrote one wasn't.

a. EXTRACTION TYPE: subject-extraction

[RC [*koto* __i SF-shoosetsu-o kai-ta-koto]-ga saikin shoten-de
Sci-Fi novel-ACC write-PST-fact-NOM recently bookstore-at
tokushuu-sa-re-ta] gakusha$_i$-wa hokorashige-da.
feature-do-PASS-PST professor-TOP looks.proud-COP
'The professor$_i$ [RC who the fact that [*koto* __i wrote a sci-fi novel] was recently featured in a bookstore] looks proud.'

Context: Two sci-fi novels were published. The fact that a professor wrote one of them was featured in a bookstore, but the fact that an actor wrote the other wasn't.

b. EXTRACTION TYPE: object-extraction

[RC [*koto* gakusha-ga __j kai-ta-koto]-ga saikin shoten-de
professor-NOM write-PST-fact-NOM recently bookstore-at
tokushuu-sa-re-ta] SF-shousetsu$_j$-wa daigakusee-ni
feature-do-PASS-PST Sci-Fi novel -TOP college.students-DAT

ninki-da.
popular-COP
'The sci-fi novel$_j$ [*koto* that the fact that [$_{RC}$ the professor wrote __$_j$] was recently featured in a bookstore] is popular among college students.'

Condition 2a (=20a) corresponds to Condition 4a (=18a) in that both involve long-distance relativization of the subject (*professor*), whereas Condition 2b (=20b) corresponds to Condition 4b (=18b) in that both involve long-distance relativization of the object (*sci-fi novel*). Similar to Condition 3 (=19), contexts for these conditions contrast two referents of the same kind, differentiating them using a *koto*-clause. Lastly, the condition with a *koto*-clause without an extraction out of it is presented below.

(21) Condition 1 EMBEDDED CLAUSE TYPE: *koto*-clause; EXTRACTION: -
Context: There is a professor in a college in Tokyo who plays an active role in the literary world.
EXTRACTION TYPE: subject-extraction / object-extraction
[*koto* gakusha-ga SF-shoosetsu-o kai-ta-koto]-ga saikin
professor-NOM Sci-Fi novel-ACC write-PST-fact-NOM recently
shoten-de tokushuu-sa-re-ta.
bookstore-at feature-do-PASS-PST
'The fact that [*koto* a professor wrote a sci-fi novel] was recently featured in a bookstore.'

As the condition without any extractions, (21) corresponds to both Condition 2a and 2b (=20); the only difference between this condition and 2a,b is the relativization out of *koto*-clause in the latter. This is why EXTRACTION TYPE of this condition is labeled "subject-extraction / object-extraction," although stimuli in this condition did get divided into "subject-extraction" and "object-extraction" groups for the sake of creating a Latin square and conducting statistical analyses.

32 lexically-matched sets of conditions (18) through (21) were created. Stimuli were counterbalanced through a Latin-square procedure, resulting in 8 lists (4 items per condition; 32 items per list). Each list was mixed with the same set of 40 fillers as in the previous experiments, resulting in 72 items in total. As in the previous experiments, the order of stimuli was pseudo-randomized such that two critical items never appeared sequentially.

2.5.2.3 Procedures

The same experimental procedures were used as in Experiment 3a. One modification made in the current experiment is that feedback was provided in response to a comprehension question (e.g., (17)), in hopes of encouraging participants to pay more attention once they learn that they have answered incorrectly. Perhaps thanks to this modification, no participant was excluded due to a poor performance on comprehension questions. 3 participants whose performance was significantly different from the average on the selected fillers were excluded, and we obtained the final dataset of 48 participants after recruiting additional participants.

2.5.2.4 Data analysis

In order to test for the existence of island effects in the two types of double relatives (subject-extraction and object-extraction), we first built two separate linear mixed-effect models with random effects of participant and item. Each of the models predicted acceptability based on EMBEDDED CLAUSE TYPE, EXTRACTION, and their interaction, just as in Experiment 1. The maximally random model converged only for the subject-extraction group, and the other model based on the object-extraction group converged by eliminating the random interaction of the two factors for item.

Another model was then made based on the entire dataset, predicting the acceptability based on EMBEDDED CLAUSE TYPE, EXTRACTION, EXTRACTION TYPE, and their interactions. Such a three-way interaction model would help us find out whether the size of the interaction effect is significantly different between the two types of double relatives (in spite of the contrastive context). This model was created under two statistical frameworks, one frequentist and the other Bayesian. Under the former, we calculated a *p*-value for the interaction of the three factors, in parallel to the other models we have constructed. Importantly, while a *p*-value smaller than 0.05 is conventionally taken as the evidence to reject the null hypothesis (in our case, that there is no difference in the size of island effect between the two types of double relatives), a *p*-value larger than 0.05 does not necessarily mean that the null hypothesis is correct; such a *p*-value only lets us conclude that the model failed to find an effect (Nicenboim & Vasishth, 2016, Vasishth & Nicenboim, 2016). And yet, evaluating the null hypothesis is of importance here as it helps us evaluate the two analyses (major subject analysis or uniform extraction analysis). The Bayesian framework allows us to do exactly this, as it computes the probability of each hypothesis given the observed data, according to Bayes' theorem. Thus, we constructed a Bayesian mixed-effects model using the brms package (Bürkner, 2017) in R with weakly informative priors, and calculated Bayes Factors for hypothesis testing, which is the ratio of likelihoods of the two hypotheses under comparison. We adopt the standard threshold (Jeffreys, 1939, 1961) that $BF_{10} > 3$ is evidence for the hypothesis that a three-way interaction is present (i.e., the size of the island effect between the two types of double relatives is different), $BF_{10} < 0.33$ is evidence for the hypothesis that the interaction is absent (i.e., there is no difference in the size of the island

effect between the two types of double relatives), and $0.33 < BF_{10} < 3$ is inconclusive (both hypotheses are equally likely).[7]

The model that converged under the frequentist framework controls for random intercepts of participant and item, random slopes of EMBEDDED CLAUSE TYPE, EXTRACTION, and their interaction for participant, and random slopes of EMBEDDED CLAUSE TYPE for item. The Bayesian model with maximal random effects was able to converge.

2.5.2.5 Predictions

For both types of double relatives (subject-extraction vs. object-extraction), if the contrastive context improves acceptability to the point that their associated superadditivity disappears, we should not find an interaction effect between EMBEDDED CLAUSE TYPE and EXTRACTION. In contrast, if the superadditivity persists despite the context, there should be an interaction effect for subject-extraction (replicating Experiment 1) and/or object-extraction double relatives.

Furthermore, if the major subject analysis is correct and only object-extraction involves an island violation, then we would expect to find a smaller effect size with subject-extraction (as a result of additional structure only in the double relative condition) and a larger effect size with object-extraction (as a result of the island effect induced by extraction of the object out of the relative clause island). That is, subject-extraction should induce a structure effect (presumably relatively small) and object-extraction should induce an island effect (presumably relatively large). If, on the other hand, double relatives are true island violations and both subjects and objects are extracted directly out of the relative clause island, as the uniform extraction analysis

[7] BF_{10} is the ratio of the likelihood of data under H1 to the likelihood of data under H0. It is also possible to calculate BF_{01}, where the thresholds will be the opposite of those of BF_{10}, (e.g., $BF_{01} > 3$ is in support of H0).

proposes, then we should find a significant interaction (i.e., island effect) with both subject-extraction and object-extraction, and the size of this interaction should not significantly differ between the two extraction types.

2.5.2.6 Results

The mean z-scores for the eight conditions are presented in Figure 2.5.

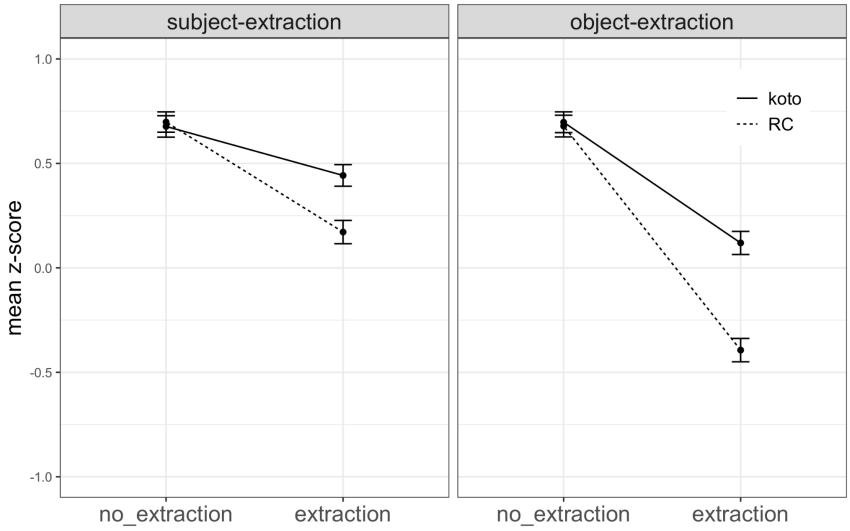

Figure 2.5: Aggregated responses to critical items in Experiment 3b

Focusing first on the subject-extraction conditions (left), the model revealed significant main effects of EMBEDDED CLAUSE TYPE (β = -0.27, SE = 0.09, $p < 0.01$) and EXTRACTION (β = 0.23, SE = 0.11, $p = 0.03$). Crucially, the interaction between the two factors was also significant (β = 0.28, SE = 0.1, $p < 0.01$), replicating the results of Experiment 1. Next, the model based on the object-extraction conditions (right) also revealed significant main effects of EMBEDDED CLAUSE TYPE (β = -0.51, SE = 0.11, $p < 0.001$), EXTRACTION (β = 0.58, SE = 0.12, $p < 0.001$), and a significant interaction between them (β = 0.5, SE = 0.13, $p < 0.001$).

The model based on the full dataset made under the frequentist framework revealed a

significant main effect of EXTRACTION TYPE, such that sentences belonging to the subject-extraction conditions were rated higher than the ones in the object-extraction conditions (β = 0.32, SE = 0.06, $p < 0.001$). Importantly, however, the three-way interaction did not reach significance (β = -0.21, SE = 0.13, $p = 0.1$), indicating that there is no evidence for a significant difference between the two extraction types in the size of the interaction effect between EMBEDDED CLAUSE TYPE and EXTRACTION.

The results of the Bayesian model align with those of the frequentist model just presented. The Bayes Factor of the main effect of EXTRACTION TYPE was 738.65, clearly showing the significance of the effect ($BF_{10} > 3$). The Bayes Factor of the three-way interaction was 0.18, which is under the $BF_{10} < 0.33$ threshold and suggests that the obtained data is about five times more likely under the null hypothesis, which posits that there is no difference between subject-extraction and object-extraction in the size of the interaction, than under the alternative hypothesis in which there is a difference. Crucially, while the *p*-value larger than 0.05 in the frequentist model only lets us conclude that we failed to find a three-way interaction effect, we can take the fact that the Bayes Factor is smaller than 0.33 to indicate that there is no three-way interaction effect.

We also obtained the DD scores for the two types of double relative, as in Experiment 1. The score for the subject-extraction type was 0.25 (which is close to 0.3 from Experiment 1), and it was 0.5 for the object-extraction type. There is thus a numerical difference in the size of the interaction, but this difference is not significant, as we have just seen.

2.5.2.7 Discussion

The results from Experiment 3b show that both subject-extraction and object-extraction out of a relative clause induce a superadditive effect on acceptability, and while double relatives of the former type have higher acceptability than those of the latter type, there is no significant difference between the two in the size of the superadditive effect. Moreover, despite the outcome of Experiment 3a, which found a significant improvement in acceptability of double relatives as a result of accompanying them with contrastive context, the superadditive effect triggered by the structures persisted.

As we saw above, the major subject analysis predicts that the superadditive effect will be smaller with subject extraction than it will be with object extraction. This is because with subject extraction, there is no island effect under this analysis, and any superadditivity is due to the increased structural complexity associated with the major subject structure. With object extraction, in contrast, superadditivity could only be due to an island effect. This prediction in the size of the superadditivity effect is strengthened further by the results seen here in which the addition of a relative clause, which entails a considerable increase in structural complexity, has no effect on acceptability. Specifically, Condition 3 (EMBEDDED CLAUSE TYPE: relative clause; EXTRACTION: -) has essentially the identical level of acceptability as Condition 1 (EMBEDDED CLAUSE TYPE: *koto*-clause; EXTRACTION: -), even though Condition 3 contains a relative clause and Condition 1 does not. If complex structures like relative clauses have little to no effect on acceptability, then a major subject structure is likely to be similar, and we would expect only a small amount of superadditivity, if any, in the subject-extraction case. Superadditivity in the object-extraction case, however, is expected to be larger, since this necessarily involves extraction out of an island, which is known to be associated with robust superadditivity.

The results of Experiment 3b reveal superadditivity with both subject extraction and object extraction and crucially, there is no significant difference in the size of these two effects. The two effects are numerically different when they are measured as DD scores, as seen above, but one cannot draw statistical inferences from this comparison, given that each DD score is, in effect, a single observation. Instead, testing for a three-way interaction under a frequentist model shows no evidence for a difference in the size of superadditivity between subject extraction and object extraction, while the Bayesian model provides evidence that there is no such difference.

This lack of a significant difference between the two extraction types in the size of the interaction effect suggests strongly that the major subject analysis is not correct. That analysis claims that there is additional structure in the subject-extraction case, which we would expect to result in little to no superadditivity, and that there is an island violation in the object-extraction case, which we would expect to result in much larger superadditivity. As we have seen, though, Experiment 3 found no significant difference between the two in the amount of superadditivity. This result is exactly what is expected under what we have termed the uniform extraction analysis. Under that view, both extraction types would involve extraction directly out of the relative clause, and therefore island violations. Since both extraction types operate in essentially the same way, there is no reason to expect a difference in the amount of superadditivity in the two cases, and this is just what Experiment 3 found.

The uniform extraction analysis thus captures an important finding of Experiment 3; that both subject-extraction and object-extraction show superadditivity and there is no significant difference between them in the size of the effect. Nevertheless, the two extraction types do not behave identically in Experiment 3. Specifically, there is a significant main effect of EXTRACTION TYPE, driven by the fact that Conditions 2 and 4 are significantly more acceptable

with subject-extraction (seen in the left panel of Figure 2.5) than they are with object-extraction (seen in the right panel of Figure 2.5). This result is not predicted by the uniform extraction analysis since there is no reason to expect that extracting an object out of an embedded clause would be systematically less acceptable than extracting a subject. It is also not predicted by the major subject analysis. In that analysis, we would expect this asymmetry in Condition 4 (where subject extraction involves extraction of a major subject and object extraction involves extraction out of an island), but not in Condition 2 (where both extraction types involve extraction out of the embedded clause).

What, then, could be causing this difference between subject- and object-extraction? To see this, let us look more closely at these two extraction types in these two conditions. Condition 2 is presented in (22) and Condition 4 is presented in (23), with subject-extraction in the (a) examples and object-extraction in the (b) examples (repeated from (20) and (18), respectively).

(22) Condition 2

a. [$_{RC}$ [*koto* __$_i$ SF-shoosetsu-o kai-ta-koto]-ga saikin
Sci-Fi novel-ACC write-PST-fact-NOM recently
shoten-de tokushuu-sa-re-ta gakusha$_i$]-wa hokorashige-da.
bookstore-at feature-do-PASS-PST professor-TOP looks.proud-COP
'The professor$_i$ [$_{RC}$ who the fact that [*koto* __$_i$ wrote a sci-fi novel] was recently featured in a bookstore] looks proud.'

b. [$_{RC}$ [*koto* gakusha-ga __$_j$ kai-ta-koto]-ga saikin
professor-NOM write-PST-fact-NOM recently
shoten-de tokushuu-sa-re-ta SF-shoosetsu$_j$]-wa
bookstore-at feature-do-PASS-PST Sci-Fi novel-TOP
daigakusee-ni ninki-da.
college.students-DAT popular-COP
'The sci-fi novel$_j$ [*koto*that the fact that [$_{RC}$ the professor wrote __$_j$] was recently featured in a bookstore] is popular among college students.'

(23) Condition 4

a. [RC2 [RC1 __i __j kai-ta] SF-shoosetsu$_j$-ga saikin
write-PST Sci-Fi novel-NOM recently
shoten-de tokushuu-sa-re-ta gakusha$_i$]-wa hokorashige-da.
bookstore-at feature-do-PASS-PST professor-TOP looks.proud-COP
'The professor$_i$ [RC2 who the sci-fi novel$_j$ [RC1 that __i wrote __j] was recently featured in a bookstore] looks proud.'

b. [RC2 [RC1 __i __j kai-ta] gakusha$_i$-ga saikin
write-PST professor-NOM recently
shoten-de tokushuu-sa-re-ta SF-shoosetsu$_j$]-wa
bookstore-at feature-do-PASS-PST Sci-Fi novel-TOP
daigakusee-ni ninki-da.
college.students-DAT popular-COP
'The sci-fi novel$_j$ [RC2 that the professor$_i$[RC1 who __i wrote __j] was recently featured in a bookstore] is popular among college students.'

All the sentences above involve long-distance dependencies from one clause to another, specifically the extraction of the subject or object inside the relative clause to a position outside the relative clause. They contrast with Condition 3 presented below (repeated from (19)), which involve subject- and object-extraction at a shorter distance, within the relative clause.

(24) Condition 3

a. [RC __i SF-shousetsu-o kai-ta] gakusha$_i$-ga saikin shoten-de
Sci-Fi novel-ACC write-PST professor-NOM recently bookstore-at
tokushuu-sa-re-ta.
feature-do-PASS-PST
'The professor$_i$ [RC who __i wrote a sci-fi novel] was recently featured in a bookstore.'

b. [RC gakusha-ga __j kai-ta] SF-shoosetsu$_j$-ga saikin shoten-de
professor-NOM write-PST Sci-Fi novel-NOM recently bookstore-at
tokushuu-sa-re-ta.
feature-do-PASS-PST
'The sci-fi novel$_j$ [RC that the professor wrote __j] was recently featured in a bookstore.'

Unlike the (a) and (b) sentences in (22) and (23), Experiment 3b revealed that the acceptability of (24a) and (24b) are virtually identical (see Figure 2.5).

The (a) and (b) examples in (22) through (24) should be structurally the same, since both involve extraction (of the subject and object, respectively) out of a particular environment. In terms of processing these examples in linear order, however, (a) and (b) appear very different. In the examples in (a) with subject-extraction, the verb and its object (*kai-ta* 'wrote' and *SF-shousetsu* 'sci-fi novel', respectively) are adjacent and processed first, before the more distant subject filler (*gakusha* 'professor'). In the examples in (b), however, with object-extraction, the verb and the subject are adjacent and processed first, while it is the object filler that is more distant.

Additionally, the sentences from (22) through (24) differ from each other in terms of the dependency length between the verb and its extracted arguments; while dependencies in (24) cross one relative clause boundary, the ones in (22) cross a *koto*-clause and a relative clause boundary, and the ones in (23) cross two relative clause boundaries. Moreover, while Conditions 2 and 3 each contain only one long-distance dependency, Condition 4 contains two.

These distinctions are important; long-distance dependencies are known to lead to a significant drop in acceptability, and the longer the dependency (i.e., the more clause boundaries it has to cross), the larger the drop (Frazier & Clifton, 1989; Kluender & Kutas, 1993; Alexopoulou & Keller, 2007; Fanselow, 2021; Goodall, 2021), It is generally assumed that this drop results from an increased burden on the processor. This could explain why the acceptability of Condition 3 (=(24)) is significantly higher than the acceptability of both Condition 2 and 4 (=(22) and (23)), as there is no long-distance dependency in the former. Similarly, as the only

condition that involves two long-distance dependencies, Condition 4 was judged to be lower in acceptability than both Conditions 2 and 3.

Similar to the existence of long-distance dependencies, there are also consequences of extracting the subject (as in (a) sentences in (22) through (24)) as opposed to extracting the object (as in (a) sentences in (22) through (24)) in terms of processing. Most current approaches to clause structure assume that objects are arguments of the lexical verb, but that subjects are arguments of a higher functional head, such as *v*. This reflects the fact that the semantic role of the subject depends on the nature of the verb and its object (Marantz, 1984).[8] Thus, in case of long-distance extraction of the subject, the object remains adjacent to the verb and it receives its semantic role from the verb. When the subject is then encountered, it receives its semantic role from the verb-object combination straight away. In case of the long-distance extraction of the object, semantic role assignment proceeds in a more roundabout way; the subject remains adjacent to the verb, but it cannot receive a thematic role from just the verb. It is only after encountering the object, and the verb assigning the role to the object, can the subject finally receive its role. In terms of processing, then, one would expect structures where the verb and its object are processed first to be easier than ones where the verb and the subject are processed first, before the object has been encountered, because in the latter case, the subject must wait for the object for its semantic role assignment.[9] This idea has been pursued in Nakamura and Miyamoto (2013) (see also Miyamoto & Nakamura, 2011), where they investigate what causes

[8] To illustrate, given an incomplete phrase *John threw*, the exact semantic role of *John* changes depending on the object; it is agent if the phrase continues as in *John threw a ball*, whereas it is experiencer if *John threw a fit*.
[9] Another way of looking at this pattern is from a processing perspective: Kwon et al. (2013) argue that in head-final subject relative clauses, the parser can compile the VP straightaway, and then all that needs to happen at the head noun is predication of this VP of the subject (head noun). In contrast, in object relative clauses, at the head noun position, they need to slot the object head noun into the missing argument slot of the VP before predicating the VP of the relative clause subject. Put simply, while object relative clauses involve a two-step process (compile VP, then predicate), whereas subject relative clauses only involve the latter step.

the asymmetry in the acceptability of subject-extraction (=(23a)) and object-extraction (23b) double relatives. With a self-paced reading experiment, they observed significant slowdowns in object-extraction double relatives at the position of the subject (*gakusha* 'professor' in (23b)) and the position of the object (*SF-shoseetsu* 'sci-fi novel'). They claim that the slowdowns reflect the difficulty due to the fact that the semantic role of the subject cannot be determined by the verb alone, and thus it is delayed until the object appears. Such slowdowns would not happen in subject-extraction double relatives, given that the object can receive its semantic role from the verb alone, and the subject that appears afterwards can receive its semantic role from the verb and the object without delay. They thus propose an *object before subject bias* or ObS, which is a processing preference to "have the role of the object assigned before the role of the subject" (Nakamura & Miyamoto 2013; p.305) and propose that ObS is what underlies the asymmetry between subject- and object-extraction double relatives.

Given the patterns of processing preferences presented above, as well as Nakamura & Miyamoto's (2013) ObS, we expect long-distance dependencies to lower acceptability in general, especially if there are more than one of them in a sentence. But if those long-distance dependencies involve processing the verb and its object (i.e., assigning a semantic role to the object) before processing the extracted subject, such dependencies should be easier than the ones involving processing the verb and its subject before processing the extracted object. Given all of this, and borrowing the term from Nakamura and Miyamoto, we can now formulate ObS as follows:

(25) *Object before subject bias* (ObS)

When long-distance dependencies between arguments and predicates are being resolved, the dependency associated with the object argument must be fully resolved before the

dependency associated with the subject argument.

We use the verb *resolve* to indicate that an argument and a predicate have been identified, to be differentiated from *fully resolve* which indicates that the semantic role of the argument has been assigned. Let us now examine how ObS plays out in sentences in (22) through (24). First, some of the sentences in Condition 2 (e.g., (22a)) involve the object (*SF-shoseetsu* 'sci-fi novel') and the verb (*kai-ta* 'wrote') that are adjacent and their dependency is fully resolved first, while the subject (*gakusha* 'professor') comes much later across a *koto*-clause. In others (e.g., (22b)), however, it is the subject and the verb that are adjacent and resolved first, while the object appears much later, which means that the subject cannot be fully resolved until the object is because its semantic role depends on the object. Given ObS, one clearly expects (22b) to be more difficult to process and presumably less acceptable than (22a), given the degrading effect that processing difficulty often has on acceptability (Fanselow, 2021). As we have seen, this expectation is borne out.

Second, sentences in Condition 4 (=(23)) begin with the verb, with both of its arguments missing, so there are two dependencies to process. In (23a), though, the object dependency gets fully resolved first, followed by the subject. One can thus fully resolve each dependency one at a time. In contrast, the verb and its subject are resolved first in (23b), but the dependency cannot be fully resolved since the object is undetermined. It is only after the object appears, across two relative clause boundaries, that the subject dependency is fully resolved. Again, this difference in processing procedures was reflected in acceptability judgments in our experiment; whereas there was a clear relative clause island effect in both extraction types, there was further degradation with object-extraction (=(23b)), in accord with the prediction of ObS.

Lastly, for Condition 3, while the dependency between the object and the verb gets fully resolved immediately in (24a), the semantic role assignment of the subject is delayed until the object in the relative clause head position is identified in (24b). As we have seen, however, there is no difference in acceptability between (24a) and (24b). This would not be surprising, though, given that neither sentence involves a dependency that is of comparable length to Condition 2 and 4. Thus, while there is most likely some processing cost associated with (24b) compared with (24a) (which aligns with the previous findings that the Japanese parser finds it easier to process subject relative clauses than object relative clauses; Miyamoto & Nakamura, 2003; Ishizuka et al., 2003; Ueno & Garnsey, 2008), the cost differential is apparently not captured by acceptability experiments such as this.

In sum, ObS offers a straightforward and intuitively appealing account of the difference between subject-extraction and object-extraction across different structures.[10]

2.6 General discussion

Through a series of three acceptability experiments, we have seen that double relatives in Japanese show signs of an island effect, even when they are relatively acceptable, that this pattern of island effects is seen with an A'-movement dependency (relativization), but not with a non-A'-movement dependency (anaphora), and that there seems to be no significant difference in the size of an additional decrement in acceptability between high-acceptability (subject-extraction) and low-acceptability (object-extraction) double relatives.

[10] A reviewer has asked if the preference for subject extraction out of an embedded clause goes against the cross-linguistic pattern where object extraction is preferred. We note, however, that the relevance of this pattern to Japanese double relatives is complicated by the fact that double relatives involve both short-distance and long-distance extractions, which have opposite preferences (McDaniel et al., 2015). Furthermore, a large part of the dispreference for subject extraction out of embedded clauses seems to stem from the COMP-trace effect (e.g., Pesetsky, 2017), which does not appear to exist in Japanese.

These findings lead us to two major conclusions. First, they suggest that relative clauses are islands in Japanese. This is in accord with what has been claimed based on other types of dependencies in the language (scrambling and *wh*-in-situ), as well as with what has traditionally been claimed in many other languages. Second, they suggest that Japanese does not have a way to evade the island status of relative clauses, contrary to what has often been claimed in the literature. That is, relativization out of a relative clause does not occur by means of a non-movement dependency, as in analyses where the "gap" within the relative clause is a null pronoun (e.g., Perlmutter, 1972; Kuno, 1973; Murasugi, 2000; Fukui & Takano, 2000), and it does not occur by means of an A'-movement that does not reach into the relative clause itself, as in analyses where the gap is in the major subject position outside of the relative clause (Sakai, 1994; Han & Kim, 2004; Ishizuka, 2009; Han, 2013). In these latter analyses, we would not expect to observe an island effect with double relatives, but in fact we do. Instead, our findings suggest that relativization out of a relative clause is exactly what it appears to be: A'-extraction out of the relative clause itself. This conclusion is bolstered by the fact that both high-acceptability (subject-extraction) and low-acceptability (object-extraction) double relatives show island effects of approximately the same size. The difference in acceptability between the two is plausibly due to the fact that in the high-acceptability cases, the verb and its object are able to be processed first and the subject is delayed, while in the low-acceptability cases, the verb and the subject are processed first and the object is delayed. Since the semantic role of the subject depends on the nature of the verb and its object, it makes sense that sentences in which the verb and the subject are processed first, with a substantial delay for the object, result in higher processing costs (as predicted by ObS) and lower acceptability.

At least superficially, our conclusions are not surprising. The idea that relative clauses are islands and that relativization creates an A'-dependency that is sensitive to islands has been part of mainstream thinking in syntax for many decades (Ross, 1967; Chomsky, 1977). What is perhaps surprising is that our results suggest that extraction out of a relative clause produces only a mild island effect. That is, when processing of the sentence is able to unfold without any special difficulty (i.e., when ObS is satisfied), acceptability is still relatively high for double relatives, despite the presence of an island effect. This is what occurs when a subject is extracted out of a relative clause, as in (23a). In this case, the verb and its object can be processed first, before the subject, so processing proceeds relatively smoothly. Nonetheless, extraction out of an island occurs and a detectable island effect results (as signaled by the superadditive interaction between extraction and structure). The island effect is small, though, suggesting that extraction out of a relative clause island itself causes only a mild degradation of acceptability. In a case where an object is extracted out of relative clause, as in (23b), processing is much more difficult, so this plus the island effect result in a larger degradation of acceptability. Both subject-extraction (23a) and object-extraction (23b) involve an island violation, then, but subject-extraction allows us to see a "purer" form of this violation, without the addition of a major processing difficulty.

Subject-extraction out of relative clauses, as in (23a), is thus an example of a "subliminal island effect," i.e. a case where there is a clear island effect but where the island-violating condition is relatively acceptable (Almeida, 2014; Keshev & Meltzer-Asscher, 2019; Kush et al., 2018; Stigliano & Xiang, 2021). As we saw for subject-extraction in Experiment 1 (Section 2.3) and Experiment 3 (Section 2.5), there was a significant superadditive interaction between extraction and the type of embedded clause, which is indicative of an island effect. In addition,

the condition in which there was extraction out of a relative clause (Condition 4) had a mean z-score of 0.06 in Experiment 1 and 0.17 in Experiment 3, which is higher than the mean for fillers that were designed to be of intermediate acceptability (-0.45 in Experiment 1 and -0.46 in Experiment 3), confirming the observation of many researchers that double relatives can be relatively acceptable in Japanese. In our analysis, the "subliminal" nature of the island effect here is simply due to the fact that relative clauses yield very mild island effects. When this effect is paired with significant processing difficulty, as occurs with object-extraction out of relative clauses, acceptability drops to the point that the effect is no longer "subliminal." We saw in Experiment 3 (Section 2.5) that object-extraction yielded a superadditive interaction of approximately the same size as with subject-extraction, but the mean z-score of the island-violating condition (Condition 4) was -0.39, which is comparable to the mean acceptability of intermediate fillers and higher than bad fillers (-0.77). These results for object-extraction are more typical of what is ordinarily expected with island phenomena, but a comparison with the subject-extraction case suggests that these are due both to the island effect itself and to processing difficulties (specifically the ObS factor discussed in Section 2.5).

In fact, earlier work has shown that other types of extraction out of relative clauses in Japanese also lead to a relatively mild island effect when additional factors are not present. Specifically, Tanaka & Schwartz (2018) (discussed in Section 2.2.1) found an island effect when *wh*-dependencies span a relative clause boundary, but there is no overt movement in this case and no filler-gap dependency to be resolved, so there is no special difficulty (and no ObS violation) in processing the subject and object in relation to the verb. As expected, then, the mean acceptability of such sentences in z-scores was positive, suggesting relatively high acceptability, despite the island effect. With scrambling out of a relative clause, in contrast,

Fukuda et al. (2022) found a clear island effect and relatively low acceptability of the island-violating structure. ObS proposed in the previous section does not readily explain why this is the case; the dependency between the scrambled object and the verb should be fully resolved as soon as the verb appears, and the subject (which is the relative clause head) follows. Unlike what we saw in object-extraction double relatives, then, there is no delay in assigning the subject role while the object is pending, so scrambling out of a relative clause should be an ObS-satisfying structure. It may be that factors aside from ObS are causing the low acceptability, however. For instance, presupposed domains are well-known to be islands for extraction (Bianchi & Chesi, 2014; Diesing, 1992; Fiengo & Higginbotham, 1981), and it is conceivable that scrambling makes its extraction domain particularly presupposed and thus opaque for extraction. We will return to the case of scrambling out of relative clauses in Chapter 5 to investigate this possibility further, by manipulating the presuppositionality of the extraction domain.

Note that unlike accounts of subliminal island effects in the literature where the effect is entirely due to processing costs (e.g., Keshev & Meltzer-Asscher, 2019, where they found a small amount of superadditivity even with binding relations across a *wh*-island in Hebrew), our account attributes the subliminal effect entirely to the syntax. Specifically, we have seen that the island effect arises only with A'-movement dependencies (and not with other dependencies, as in Experiment 2) and it occurs whether or not there are additional factors that cause difficulties for processing (as in Experiments 1 and 3). When processing is harder, as in ObS-violating sentences, acceptability declines, but the size of the island effect remains the same, as we saw in Experiment 3.

Could we extend the proposal that relative clause island effects are mild by themselves to languages like English, where the effect seems to be much more noticeable? The English equivalents of (22a) and (22b), for instance, do not appear to be very acceptable:

(26) a. *The professor$_i$ [$_{RC2}$ who the sci-fi novel$_j$ [$_{RC1}$ that __$_i$ wrote __$_j$] was recently featured in a bookstore] looks proud.
b. *The sci-fi novel$_j$ [$_{RC2}$ that the professor$_i$ [$_{RC1}$ who __$_i$ wrote __$_j$] was recently featured in a bookstore] is popular among college students.

There may be a difference in acceptability between these two, but neither seems to approach the level observed with sentences like (23a) in Japanese. This is perhaps not surprising, given that both sentences in (26) involve extraction out of a subject (i.e., the DP containing the relative clause is a subject), which generally leads to unacceptability in English (but not in Japanese; see Omaki et al., 2020), but even if we correct for this, as in (27), the sentences still do not seem very acceptable.[11]

(27) a. *The professor$_i$ [$_{RC2}$ that we read the sci-fi novel$_j$ [$_{RC1}$ that __$_i$ wrote __$_j$]] looks proud.
b. *The sci-fi novel$_j$ [$_{RC2}$ that we know the professor$_i$[$_{RC1}$ who __$_i$ wrote __$_j$]] is popular among college students.

Initial appearances suggest, then, that extracting an argument out of a relative clause generally results in a very perceptible decline in acceptability in English regardless of whether it involves subject-extraction, as in (27a), or object-extraction, as in (27b). There may be many

[11] In (27a) (and also (26a)), there is long-distance extraction out of the embedded subject position in the presence of overt material in CP, so these are likely instances of COMP-trace effect. Even if we correct for this, though, the sentence does not appear to be very acceptable:
(i) *The professor$_i$ [$_{RC2}$ that we read the sci-fi novel$_j$ [$_{RC1}$ that somebody gave __$_j$ to __$_i$]] looks proud.

contributing factors at play here, but it is worth noting that in neither (27a) or (27b) do we have a situation as in Japanese (23a), where the dependencies between the verb and its argument are resolved one at a time. As we have seen, the linear order in (23a) is in accord with ObS and allows for relatively easy processing of the verb and its arguments. This order does not obtain in either sentence in (27), though, and indeed, such a situation is impossible to create in English, given the word order in the language, so we do not expect to find English equivalents of (23a) that are of similarly high acceptability. Chapter 4 investigates the potential factors that make English double relatives particularly low in acceptability in a greater detail, by considering other ways to increase the acceptability of double relatives in English, as discussed in Vincent et al. (2022).

2.7 Chapter summary

Through a series of acceptability judgment experiments, this chapter provided supporting evidence for the proposal that double relatives in Japanese do exhibit a small yet significant relative clause island effect, despite the fact that some of them have been judged to be relatively well-formed. The chapter then argued that the penalty of a relative clause island violation is in fact mild, and that the sentences that appear to exhibit a strong relative clause island effect in fact involve not only the relative clause island violation, but also processing-oriented factors including (and possibly not limited to) the Object before Subject Bias. This chapter has thus provided a unifying analysis of the two types of double relatives (subject- and object-extraction, with the former judged to be more acceptable than the latter), and proposed that a penalty of violating an island can be mild, the phenomena previously called a subliminal island.

Acknowledgements

Chapter 2, in part, has been submitted for publication of the material as it may appear in Natural Language and Linguistic Theory, Takahashi, M., and Goodall, G. The dissertation author was the primary researcher and author of this paper.

Chapter 3

Using language models to probe the nature of "mild" relative clause island effects

3.1 Introduction

The last chapter made a proposal, which will be a point of focus for the rest of this dissertation, that the relative clause island effect itself is milder than previously thought. What a "mild" island effect means is that a sentence in violation of an island can still be judged relatively acceptable, despite a significant superadditive drop in acceptability. Results of the experiments from the last chapter led to the conclusion that Japanese double relatives exemplify mild island effects. What I call mild island effects should thus be distinguished from "weak" islands, as the latter typically refers to islands that permit extraction (i.e., no superadditive drop in acceptability) under certain circumstances (Szabolcsi & den Dikken, 2003; Szabolcsi 2006; Szabolcsi & Lohndal, 2017).

The last chapter featured subject or object relativization out of another relative clause (double relatives) in Japanese, exemplified in (1) below.

(1) a. *subject-extraction double relative*

[$_{RC2}$ [$_{RC1}$ __$_i$ __$_j$	kai-ta]	SF-shoosetsu$_j$-ga	saikin
	write-PST	Sci-Fi novel-NOM	recently
shoten-de	tokushuu-sa-re-ta	gakusha$_i$]-wa	hokorashige-da.
bookstore-at	feature-do-PASS-PST	professor-TOP	looks.proud-COP

'The professor$_i$ [$_{RC2}$ who the sci-fi novel$_j$ [$_{RC1}$ that __$_i$ wrote __$_j$] was recently featured in a bookstore] looks proud.'

b. *object-extraction double relative*

[$_{RC2}$ [$_{RC1}$ __$_i$ __$_j$	kai-ta]	gakusha$_i$-ga	saikin
	write-PST	professor-NOM	recently
shoten-de	tokushuu-sa-re-ta	SF-shoosetsu$_j$]-wa	
bookstore-at	feature-do-PASS-PST	Sci-Fi novel-TOP	
daigakusee-ni		ninki-da.	
college.students-DAT		popular-COP	

'The sci-fi novel$_j$ [$_{RC2}$ that the professor$_i$[$_{RC1}$ who __$_i$ wrote __$_j$] was recently featured in a bookstore] is popular among college students.'

Acceptability judgment experiments from the last chapter challenged the analyses that some of the instances of double relatives (the subject-extraction type, as in (1a)) actually involve no A'-movement from inside the relative clause, and thus no island violation (Sakai, 1994; Han & Kim, 2004; Ishizuka, 2009). In particular, the experiments demonstrated that even subject-extraction double relatives exhibit a superadditive drop in acceptability, indicative of an island effect, despite relatively high acceptability. Results of the experiments are instead in support of the alternative analysis, which states that double relatives uniformly involve A'-movement from inside the relative clause. And while the penalty for violating the relative clause island itself is mild, some cases of double relatives (object-extraction double relatives, as in (1b)) have been judged to be ill-formed because of additional factors, exemplified by not observing the object-before-subject (ObS) preference.

Subject-extraction double relatives such as (1a) thus exemplify a "subliminal island effect," i.e., an island effect where a sentence is still judged relatively acceptable even when it is violated (Almeida, 2014; Keshev & Meltzer-Asscher, 2019). The nature of such island effects is

still being debated; instead of positing an island that triggers only a small degradation in acceptability when violated, subliminal island effects have been attributed to factors such as processing costs and variability in judgments across participants and/or items. Hence, before concluding that relative clauses trigger only a mild effect when extraction happens out of them, the nature of such an island effect needs to be evaluated more closely.

The rest of this chapter is organized as follows: Section 3.2 presents the literature that explains what may underlie subliminal island effects. The section also proposes one way to help us understand what they are, which involves collecting the equivalent of acceptability judgments from neural language models. The section then details how one can go about collecting "judgments" from those models and explains how doing so can provide us with insights that we may not be able to obtain easily from human participants. Section 3.3 presents the pattern of judgments regarding sentences with double relatives showcased by several language models trained with Japanese text, which overlaps with the judgments given by human participants from Chapter 2. Section 3.4 discusses what our results from language models inform us about the nature of relative clause island effects.

3.2 Background

3.2.1 What may underlie mild island effects

One of the studies that has attempted to identify the nature of mild island effects is Keshev and Meltzer-Asscher (2019), which featured *wh*-island effects in Hebrew, as exemplified in (2). The sentences below differ in the length of dependency formed by relativization (matrix versus embedded resolution), and whether the embedded clause is a *that*-clause (non-island) or a *wh*-clause (potential island). Similar to subject-extraction double relatives in Japanese, (2d)

displays a superadditive drop in acceptability in relation to the other conditions even though sentences involving dependency across a *wh*-clause such as (2d) have been judged to be well-formed.

(2) a. *Matrix resolution, embedded that-clause*
ha-safranit mekira et ha-profesor ha-kašuax še-hisik
the-librarian knows ACC the-professor the-strict that-gathered
še-ha-metargelet telamed et ha-student ha-mitkaše.
that-the-assistant will+teach ACC the-student the-weak
'The librarian knows the strict professor$_i$ that __$_i$ gathered that the assistant will teach the weak student.'

b. *Embedded resolution, embedded that-clause*
ha-safranit mekira et ha-student ha-mitkaše še-ha-profesor
the-librarian knows ACC the-student the-weak that-the-professor
ha-kašuax hisik še-ha-metargelet telamed.
the-strict gathered that-the-assistant will+teach
'The librarian knows the weak student$_i$ that the strict professor gathered that the assistant will teach __$_i$.'

c. *Matrix resolution, embedded wh-question*
ha-safranit mekira et ha-profesor ha-kašuax še-hisik matai
the-librarian knows ACC the-professor the-strict that-gathered when
še-ha-metargelet telamed et ha-student ha-mitkaše.
that-the-assistant will+teach ACC the-student the-weak
'The librarian knows the strict professor$_i$ that __$_i$ gathered when the assistant will teach the weak student.'

d. *Embedded resolution, embedded wh-question*
ha-safranit mekira et ha-student ha-mitkaše še-ha-profesor
the-librarian knows ACC the-student the-weak that-the-professor
ha-kašuax hisik matai še-ha-metargelet telamed.
the-strict gathered when that-the-assistant will+teach
'The librarian knows the weak student$_i$ that the strict professor gathered when the assistant will teach __$_i$.'

Instead of concluding that Hebrew *wh*-islands are an instance of subliminal islands, Keshev and Meltzer-Asscher proposed that the superadditive drop in acceptability in (2d) is due to the need to maintain two long-distance dependencies (*wh*-dependency and dependency created

by relativization) at once, instead of the *wh*-island violation. They thus conducted an acceptability judgment experiment where they replaced the gap associated with the extracted noun (*ha-student* in (2b,d), located at the end of the sentences) with a resumptive pronoun *oto* 'him.' As the dependency between a noun and its resumptive pronoun is not derived by A'-movement, (2d) with the resumptive pronoun should no longer exhibit the superadditive effect if such an effect is truly due to the *wh*-island violation. Against this prediction, the superadditive effect persisted even with the resumptive pronoun, which led Keshev and Meltzer-Asscher to conclude that the processing cost of maintaining multiple dependencies has given rise to what looks like a small island effect triggered by the *wh*-clause.

Another study by Kush et al. (2018, 2019) observed what appears to be another case of subliminal island in the extraction out of *whether*-islands in Norwegian. In their series of formal acceptability experiments featuring the extraction (*wh*-movement) out of various islands (*whether*, complex NP, relative clause, subject, and adjunct), they assessed the acceptability of sentences with *wh*-movement out of a *whether*-clause as in (3d), relative to sentences with *wh*-movement out of a *that*-clause (=3b) as well as sentences with short *wh*-movement (=3a,c). They found a significant interaction between the manipulated factors, suggestive of an island effect exhibited by (3d).

(3) a. {Hvem / Hvilken gjest} __ tror [at Hanne bakte kaken?]
who / which guest thinks that Hanne baked cake.DEF
'Who/Which guest thinks that Hanne baked the cake?'

b. {Hva / Hvilken kake} tror gjesten [at Hanne bakte __?]
what / which cake thinks guest.DEF that Hanne baked
'What/Which cake does the guest think that Hanne baked?'

c. {Hvem / Hvilken gjest} __ lurer på [om Hanne bakte kaken?]
who / which guest wonders on if/whether Hanne baked cake.DEF

'Who wonders whether Hanne baked the cake?'

d. {Hva / Hvilken kake} lurer gjesten på [om Hanne
what / which cake wonders guest.DEF on if/whether Hanne
bakte __?]
baked
'What does the guest wonder whether Hanne baked?'

At the same time, the difference-in-differences (DD) scores measuring the size of island effects (Sprouse et al., 2012) were smaller for *whether*-island than the scores for the other types of island, and the mean z-score acceptability of sentences like (3d) was positive.

Given these results, Kush et al. investigated how individual participants' judgments of sentences like (3d) may have differed, and how the judgments may have changed throughout the experiment. Kush et al. made two observations: First, some participants seem to have consistently judged sentences with a *whether*-island violation as well-formed (i.e., a positive z-score). Second, there were several participants who judged the first token to be ill-formed (i.e., a negative z-score) but the second one to be well-formed. Put differently, some participants in Kush et al.'s experiment seem to have gone through adaptation[1] after only a single exposure to *whether*-island violation. Consequently, Kush et al. argued that what looks like a mild effect of violating *whether*-island in Norwegian is due to a subset of participants who consistently accepted sentences like (3d), as well as those who went through adaptation.

Summarizing so far, there have been two cases in the literature where what looked like a mild island effect turned out to be the product of something else. In Keshev and Meltzer-Asscher's (2019) study, what appears to be a *wh*-island effect turned out not to be real, as the same patten of drop in acceptability was observed in a resumptive dependency as well. In

[1] Kush et al. chose the term *adaptation* instead of *satiation*, as the latter refers to the increase in acceptability for sentences that are considered uniformly unacceptable, at least initially. In the case of *whether*-island violation in Norwegian, some participants seem to have consistently accepted sentences with the violation.

contrast, in Kush et al. (2018, 2019), while some people experience a typical *whether*-island effect, others do not, and some may undergo adaptation after a single exposure. We are thus left wondering whether there is truly such a thing as a mild or subliminal island effect, or if it would disappear once we control for processing costs or underlying variability in judgments across participants and/or items. This would challenge the proposal about relative clause island effect made in the last chapter, which exists independently from processing factors (e.g., not following ObS) to cause a small degradation in acceptability.

The possibility that a superadditive effect emerges as a result of maintaining two dependencies, which was proposed by Keshev and Meltzer-Asscher (2019), has already been tested in our experiments with Japanese double relatives. The design of Experiment 2 of Chapter 2 closely resembles Keshev and Meltzer-Asscher's experiment; our experiment also replaced one of the gaps created by relativization with the anaphor *jibun*, which should not be sensitive to island boundaries on a par with resumptive pronouns. Contrary to Keshev and Meltzer-Asscher's experiment, however, double relatives with *jibun* no longer exhibited the superadditive drop in acceptability. From this finding, we concluded that the superadditive effect observed in sentences with double relatives is due to the A'-movement out of a relative clause, rather than processing factors such as maintaining two long-distance dependencies (created by relativization) at once. In contrast, we have yet to explore the possibility that the subliminal island effect is a result of adaptation, which was put forth by Kush et al. (2018, 2019). We propose that collecting acceptability "judgments" from large language models is one way to do this, and it could give us unique insights into the discussion of adaptation effects, for their characteristics that they do not experience adaptation in the way humans do.

3.2.2 Large language models as subjects of psycholinguistic experiments

Recent advancements in large language models (LLMs) have led researchers to evaluate whether LLMs are knowledgeable about various syntactic constraints (Linzen et al., 2016; Lau et al., 2017; Bernardy & Lappin, 2017; Kuncoro et al., 2018; Gulordava et al., 2018; Futrell et al., 2018, 2019; Marvin & Linzen, 2018; Wilcox et al., 2018). Most contemporary LLMs undergo unsupervised training (or "self-learning"), meaning that they are capable of learning linguistic regularities and making predictions about upcoming words by simply getting exposed to vast amounts of text data over an extended period of time. As the learning process shares some similarity with the process of child language acquisition, the linguistic knowledge exhibited by LLMs has been applied to discussing the learnability of syntactic rules in the absence of explicit or domain-specific input (Warstadt & Bowman, 2022; Wilcox et al., 2023). Furthermore, language modeling has been successfully used to identify the type of input required to learn various syntactic rules (Pearl & Sprouse, 2013), thanks to the ability to precisely control the amount and type of input fed to the model. In working with young children, in contrast, it would be nearly impossible to know exactly how many words and what type of sentences they have been exposed to.

Syntactic islands are one of the most-studied phenomena in the line of research introduced above. To exemplify, Wilcox et al. (2018) examined whether language models are capable of learning long-distance filler-gap dependencies in English, as well as to avoid forming such dependency when it crosses an island. Similar to experiments with human participants, they created stimuli with a factorial design that look as follows (grayed regions are the critical regions, as explained below):

(4) a. *no wh-licensor, no gap*
I know that my brother said that our aunt devoured the cake at the party.
b. *wh-licensor, no gap*
*I know what my brother said that our aunt devoured the cake at the party.
c. *no wh-licensor, gap*
*I know that my brother said that our aunt devoured __ at the party.
d. *wh-licensor, gap*
I know what my brother said that our aunt devoured __ at the party.

Of the sentences above, (4b) and (4c) are ungrammatical as they have either a filled gap or an unlicensed gap. In addition, Wilcox et al. prepared a set of sentences involving a *whether*-island, as follows:

(5) a. *no wh-licensor, no gap*
I know that my brother said whether our aunt devoured the cake at the party.
b. *wh-licensor, no gap*
*I know what my brother said whether our aunt devoured the cake at the party.
c. *no wh-licensor, gap*
*I know that my brother said whether our aunt devoured __ at the party.
d. *wh-licensor, gap*
*I know what my brother said whether our aunt devoured __ at the party.

In addition to the two conditions (5b) and (5c), the *wh*-licensor, gap condition (=5d) is also ungrammatical because of the *whether*-island violation.

In a typical acceptability experiment with human subjects, researchers ask them to provide judgments about sentences such as (4) and (5) as a whole. Alternatively, Wilcox et al. measured *surprisal* values that the models assign to individual words (Hale, 2001; Levy, 2008). Surprisal is a metric of how much the model is "surprised" to see new data, and it is the log inverse probability as formulated below.

(6) $-\log_2 \mathbb{P}(w_k|h_{k-1})$

According to the formula, defining the new data as the current word or w_k, surprisal of that word $S(w_k)$ is calculated given the h_{k-1}, the hidden state (internal representation of the model's knowledge or memory) after processing all the previous words in a sentence. Surprisal has been shown to correlate with measures of sentence processing difficulty experienced by humans, such as reading times and brain responses (Smith & Levy, 2013, Frank et al., 2015; Goodkind & Bicknell, 2018; Heilbron et al., 2022). Going back to the example with a filler-gap dependency in (4), if the model has learned that the dependency must involve an appropriate licensor and its gap, it should be surprised to see the filled gap *the cake* in (4b), resulting in a high surprisal value for the word. Likewise, it should be surprised to see the unlicensed gap in (4c), manifested as a high surprisal value on the PP *at the party* (since the gap itself is silent). Wilcox et al. designed a metric called *licensing interaction* to assess the model's overall knowledge about filler-gap dependencies, based on the surprisal values of the critical region (gaps or filled gaps and all the subsequent words, grayed in (4) and (5)), calculated as follows:[2]

(7) *Licensing interaction*

a. If a model has learned that a licensor needs a gap, surprisal for the critical region (*S*) of (4b) – *S* of (4a) is expected to be a *positive* number.
b. If the model has learned that a gap needs a licensor, *S* of (4d) – *S* of (4c) is expected to be a *negative* number.
c. Therefore, if the model has learned filler-gap (or gap-filler) dependencies, [*S* of (4b) – *S* of (4a)] – [*S* of (4d) – *S* of (4c)] is expected to be a large *positive* number.

[2] The formula in (7) is similar to the one for the DD score, which is typically used to measure the size of island effect. The DD score utilized in the last chapter was calculated as the difference in mean acceptability between the two -EXTRACTION conditions subtracted from the one between the two +EXTRACTION conditions, as follows: [-EXTRACTION, *koto*-clause – -EXTRACTION, relative clause] – [+EXTRACTION, *koto*-clause – +EXTRACTION, relative clause].

Importantly, if the model has also learned island constraints, the value of licensing interaction for sentences with a *whether*-island violation such as (5) is predicted to change in two ways: First, the model would be *less* surprised to see the filled gap *the cake* in (5b), because the model would stop expecting a gap corresponding to the licensor (*what*) once it realizes that the two would be across the *whether*-island (see Stowe, 1986; Traxler & Pickering, 1996; Phillips, 2006 for equivalent behaviors among human subjects regarding the loss of filled-gap effects when a sentence contains an island). In other words, the model would consider the appearance of the filled gap inside the island as entirely unrelated to the licensor outside the island. Similarly, the model would be *less* surprised to see the unlicensed gap in (5c), or at least any surprises that occur would be solely due to the lack of direct object of an obligatorily transitive verb *devour*, rather than the lack of licensor outside the *whether*-island. With the smaller surprisal values of (5b) and (5c), the value of the licensing interaction is also expected to become smaller when sentences involve an island, indicating that the model has learned that dependency cannot be formed across the island. These predictions made by Wilcox et al. were borne out; the models they tested showed a significant drop in licensing interaction values for sentences with an island such as (5), compared with sentences without an island such as (4). Wilcox et al. thus concluded that language models can learn to represent not only long-distance filler-gap dependencies, but also the constraints on them such as islands.

In summary, with factorial-design stimuli and metrics such as surprisal, which has been shown to correlate with human sentence processing difficulty, it is possible to probe syntactic knowledge of language models. Crucially, investigating the nature of relative clause island effects can benefit from assessing language models' knowledge of such effects. Recall that what looks like a mild island effect could be due to adaptation, where the second instance of island

violation receives a much higher rating than the first. While it would be difficult to keep humans from experiencing adaptation after even a single exposure, a language model never adapts to an ill-formed sentence unless such sentences are added to the training data as a part of training or fine-tuning. Hence, if a mild island effect emerges because of adaptation, the effect is predicted to be persistently strong when it is experienced by a language model.

In the next section, I will present results from evaluating several language models trained with Japanese text and whether they are knowledgeable about dependencies formed by relativization, as well as the relative clause island.

3.3 Experiment

3.3.1 Models tested

The language models included in this experiment were all trained with Japanese text. All of them are Transformer models (Vaswani et al., 2017), which are a type of neural network that underlies the vast majority of state-of-the-art AI technologies. Transformer models are powered by the mechanism of attention, which enables the models to take all of their previous hidden states (instead of a fixed number of previous hidden states) into account when predicting the next state (Bahdanau et al., 2014). Attention should thus make it easier for the model to learn long-distance dependencies, making Transformer models suitable subjects for testing sentences with gap-filler dependencies and their relevant constraints in Japanese.

I tested four versions of Japanese GPT-2 developed by rinna Co., available through Huggingface platform.[3] The size and content of the training data were identical across the models; they were all trained on a combination of the Japanese portion of CC-100 (a multilingual

[3] https://huggingface.co/rinna

corpus of various genres of internet texts)[4] and Japanese Wikipedia.[5] The size of the two data sources was ≈70GB for CC-100, and ≈5GB for Japanese Wikipedia according to the documentation (Sawada et al., 2024). The training data was estimated to consist of 553 million sentences.[6] Four versions of the model differed in their configurations, including the number of parameters (variables internal to the model whose values get updated through training), and I will refer to them as follows: Large (1.3 billion parameters), Medium (336 million), Small (110 million), and Xsmall (37 million).

3.3.2 Materials

The set of sentences evaluated by the models had a 2x2x2x2 design, differing in four factors: Whether a sentence has a valid relative clause licensor (i.e., relativized noun phrase in the head position) (RC-LICENSOR), whether it has a gap corresponding to the licensor (GAP), whether the embedded clause of a sentence is an island (relative clause) or a non-island (*koto*-clause) (ISLAND), and whether the relative clause licensor is associated with the embedded subject or object position (LICENSOR-POSITION). Sample pairs of stimuli involving the relative clause licensor corresponding to the embedded subject position are exhibited below; in each pair, an innermost clause is a non-island *koto*-clause in sentence (a), while it is a relative clause in sentence (b).

(8a) below exemplifies the condition with no relative clause licensor, no gap, and an embedded *koto*-clause, and (8b) is its equivalent with an embedded relative clause (headed by

[4] https://data.statmt.org/cc-100/
[5] https://dumps.wikimedia.org/other/cirrussearch
[6] The estimate is based on the size of the two data sources, as well as the data from another Japanese language model (https://huggingface.co/ku-nlp/deberta-v2-large-japanese) trained on CC-100 (85GB, 619M sentences) and Japanese Wikipedia (54GB, 326M sentences).

suiri shoosetsu 'the mystery novel'). These sentences should thus be judged as well-formed by humans.

(8) *[-RC-licensor] [-gap]*

a. [[***koto*** Gakusha-ga suiri syousetsu-o kai-ta] koto-ga saikin
professor-NOM mystery novel- ACC write-PST fact-NOM recently
syoten-de tokusyuu-sa-re-ta] koto-wa hokorashii.
book.store-at feature-do-PASS-PST fact-TOP proud
'(I'm) proud of the fact [that the fact [that the professor wrote the mystery novel] recently got featured at a bookstore].'

b. [[**RC** Gakusha-ga __$_j$ kai-ta] suiri syousetsu$_j$-ga saikin syoten-de
professor-NOM write-PST mystery novel- NOM recently book.store-at
tokusyuu-sa-re-ta] koto-wa hokorashii.
feature-do-PASS-PST fact-TOP proud
'(I'm) proud of the fact [that the mystery novel$_j$ [that the professor wrote__$_j$] recently got featured at a bookstore].'

In (9), there is a noun phrase *gakusha* 'professor' in place of *koto* in (8), acting as a relative clause licensor. As this condition involves a filled gap, it is supposed to be ungrammatical as in the English example (=(4b)).

(9) *[+RC-licensor] [-gap]*

a. [[***koto*** Gakusha-ga suiri syousetsu-o kai-ta] koto-ga saikin
professor-NOM mystery novel- ACC write-PST fact-NOM recently
syoten-de tokusyuu-sa-re-ta] gakusha-wa hokorashige-da.
book.store-at feature-do-PASS-PST professor-TOP looks.proud-COP
'The professor [who the fact that [the professor wrote a mystery novel] was recently featured in a bookstore] looks proud.'

b. [[**RC** Gakusha-ga __$_j$ kai-ta] suiri syousetsu$_j$-ga saikin syoten-de
professor-NOM write-PST mystery novel-NOM recently book.store-at
tokusyuu-sa-re-ta] gakusha-wa hokorashige-da.
feature-do-PASS-PST professor-TOP looks.proud-COP
'The professor [who the mystery novel$_j$ [that the professor wrote __$_j$] was recently featured in a bookstore] looks proud.'

It is important to note, however, that Japanese is known to allow gapless relative clauses, or relative clauses with a filled gap. To exemplify, the following sentences are almost identical to (9):

(10) a. [[*koto* **Gakusei-ga** suiri syousetsu-o kai-ta] koto-ga saikin
student-NOM mystery novel- ACC write-PST fact-NOM recently
syoten-de tokusyuu-sa-re-ta] gakusha-wa hokorashige-da.
book.store-at feature-do-PASS-PST professor-TOP looks.proud-COP
'The professor [who the fact that [the student wrote a mystery novel] was recently featured in a bookstore] looks proud.'

b. [[RC **Gakusei-ga** __j kai-ta] suiri syousetsu$_j$-ga saikin
student-NOM write-PST mystery novel- NOM recently
syoten-de tokusyuu-sa-re-ta] gakusha-wa hokorashige-da.
book.store-at feature-do-PASS-PST professor-TOP looks.proud-COP
'The professor [who the mystery novel$_j$ [that the student wrote __$_j$] was recently featured in a bookstore] looks proud.'

The only difference between (9) and (10) is that, while the filled gap in (9) is identical to the licensor *gakusha*, the one in (10) is a different noun *gakusei* 'student'. By having a filled gap not identical to the licensor, (10) is likely to be accepted as a gapless relative clause, which can be interpreted as: "The professor whose student was recently featured in a bookstore for having written a mystery novel looks proud." In contrast, (9) is most likely to be parsed as involving a filled gap, because the filled gap and the licensor are identical, and thus judged to be ill-formed on par with (4b) in English.

Both sentences in (11) are also supposed to be ill-formed, because of the lack of appropriate licensor for the gap inside the embedded clause.

(11) *[-RC-licensor] [+gap]*

a. [[*koto* __ suiri syousetsu-o kai-ta] koto-ga saikin
mystery novel- ACC write-PST fact-NOM recently
syoten-de tokusyuu-sa-re-ta] hokorashige-da.
book.store-at feature-do-PASS-PST looks.proud-COP
'__ [who the fact that [__wrote a mystery novel] was recently featured in a bookstore] looks proud.'

b. [[RC __ __$_j$ kai-ta] suiri syousetsu$_j$-ga saikin syoten-de
write-PST mystery novel- NOM recently book.store-at
tokusyuu-sa-re-ta] hokorashige-da.
feature-do-PASS-PST looks.proud-COP
'__ [who the mystery novel$_j$ [that __ wrote __$_j$] was recently featured in a bookstore] looks proud.'

Finally, (12) exemplifies the condition that has both a relative clause licensor and its corresponding gap, which is supposed to make the sentences well-formed. Recall, however, that (12b) involves the dependency between the licensor and gap across a relative clause island, and thus it is an instance of subject-extraction double relative, which was shown to exhibit an island effect in Chapter 2.

(12) *[+RC-licensor] [+gap]*

a. [[*koto* __$_i$ suiri shoosetsu-o kai-ta-koto]-ga saikin
mystery novel-ACC write-PST-fact-NOM recently
shoten-de tokushuu-sa-re-ta] gakusha$_i$-wa hokorashige-da.
bookstore-at feature-do-PASS-PST professor-TOP looks.proud-COP
'The professor$_i$ [who the fact that [__$_i$ wrote a mystery novel] was recently featured in a bookstore] looks proud.'

b. [[RC __$_i$ __$_j$ kai-ta] suiri shoosetsu$_j$-ga saikin
write-PST mystery novel-NOM recently
shoten-de tokushuu-sa-re-ta] gakusha$_i$-wa hokorashige-da.
bookstore-at feature-do-PASS-PST professor-TOP looks.proud-COP
'The professor$_i$ [who the mystery novel$_j$ [that __$_i$ wrote __$_j$] was recently featured in a bookstore] looks proud.'

As mentioned, the stimuli exemplified in (8) to (12) involve the relative clause licensor corresponding to the subject position of the embedded verb *kai-ta* 'wrote'. Another version of stimuli was created where the relative clause licensor is in the embedded object position (*suiri syousetsu* 'mystery novel' in the sentences above), and changing the main clause predicate accordingly (e.g., replacing *hokorashige-da* 'looks proud' with *ninki-da* 'is popular').[7] 8 sets of sentences were generated, including the one in (8) to (12), in order to control for any lexical effects. The full set of stimuli and code for this experiment is available on the following GitHub repository: https://github.com/matakahas/gap_filler_SCiL_2024.

3.3.3 Procedure

I loaded all four GPT-2 language models from Huggingface platform onto Google Colab (Python 3 CPU runtime). Each test sentence was tokenized with SentencePiece,[8] and a model output a surprisal value for each token (refer to (6) for the formula). I used minicons,[9] a wrapper function for Transformer library, for tokenization and surprisal calculation. I exported the results to csv files where each row contains a token and its assigned surprisal value.

3.3.4 Data analysis

The surprisal values of the critical region (grayed in (8) to (12)) were analyzed, starting with where a relative clause licensor occurs in +RC-LICENSOR conditions (=(9), (12)), the word immediately following the missing licensor in (11), or *node* 'because' occupying the licensor

[7] Stimuli were made to be as similar to the ones used in the experiments in Chapter 2, at least with regard to the conditions that existed in those experiments (=(8) and (12)). Some minor changes were necessary, however, because a couple of words were not recognized by the models; those words were replaced with the <UNKNOWN> token in the output, which could unintentionally affect surprisal values. Replacing *SF-syousetsu* 'Sci-Fi novel' with *suiri syousetsu* 'mystery novel,' the former of which was used for the stimuli in Chapter 2, is one such case.
[8] https://github.com/google/sentencepiece
[9] https://github.com/kanishkamisra/minicons

position in (8). The critical region also included any subsequent tokens until the end of the sentence, as the effect of the presence or absence of the relative clause licensor could spill over to those tokens. Because each condition and item have different starting points for the critical region, the regions needed to be indicated manually.

In order to measure the model's knowledge about the gap-filler dependency formed by relativization, I first subtracted the mean surprisal of -RC-LICENSOR, -GAP condition from that of the +RC-LICENSOR, -GAP condition (e.g., *S* of (9a) – *S* of (8a)). And I did the same for +GAP conditions (e.g., *S* of (12a) – *S* of (11a)). Then, I constructed a linear mixed-effects model predicting the values of [-RC-LICENSOR] – [+RC-LICENSOR] based on GAP, as well as the random intercepts and slopes of item.

In addition, I measured the value of licensing interaction for each lexical set of stimuli, in accord with the formula in (7), in order to evaluate whether the model has also learned the constraint that the gap-filler dependency cannot be formed across another relative clause. To test for significance, I conducted pairwise *t*-tests between the licensing interactions of -ISLAND sentences (e.g., [*S* of (9a) – *S* of (8a)] – [*S* of (12a) – *S* of (11a)]) and those of +ISLAND sentences (e.g., [*S* of (9b) – *S* of (8b)] – [*S* of (12b) – *S* of (11b)]).

3.3.5 Predictions

If language models trained with Japanese text are capable of learning gap-filler dependencies formed by long-distance relativization, they should display the following behaviors with regard to -ISLAND sentences (i.e., sentence (a) in (8) to (12)): First, according to (7a), if a model has learned that a relative clause licensor needs a gap (unless the filled gap is not identical to the licensor, which enables the gapless relative clause reading), surprisal for the critical region

in +RC-LICENSOR, -GAP sentences like (9a) should be higher than the one in -RC-LICENSOR, -GAP sentences like (8a). In other words, *S* of (9a) – *S* of (8a) will be a positive number. Second, according to (7b), if a model has also learned that a gap needs a licensor, surprisal for the critical region in -RC-LICENSOR, +GAP sentences like (11a) should be higher than the one in +RC-LICENSOR, +GAP sentences like (12a); thus, *S* of (12a) – *S* of (11a) is expected to be a negative number. Putting the two together as in (7c), the knowledge about gap-filler dependency should be manifested as a large positive value of the licensing interaction: [*S* of (9a) – *S* of (8a)] – [*S* of (12a) – *S* of (11a)].

Furthermore, if the Japanese language models have also learned the constraint that a gap-filler dependency cannot be formed across a relative clause, surprisal values are predicted to change for +ISLAND sentences (i.e., sentence (b) in (8) to (12)) as follows: First, (9b) is expected to be less surprising than (9a), given the model should treat the occurrence of the filled gap as unrelated to the relative clause licensor in the main clause. In other words, *S* of (9b) – *S* of (8b) should be smaller than *S* of (9a) – *S* of (8a). Second, (11b) should also be less surprising than (11a), or the surprisal associated with (11b) should be simply due to the lack of argument for the predicate *hokorashige-da* 'looks proud'. Accordingly, the licensing interaction for +ISLAND sentences [*S* of (9b) – *S* of (8b)] – [*S* of (12b) – *S* of (11b)] is predicted to be smaller than [*S* of (9a) – *S* of (8a)] – [*S* of (12a) – *S* of (11a)] for -ISLAND sentences.

Last but not least, given the property that language models do not undergo adaptation, they should consistently judge subject-extraction double relatives as low in acceptability, on a par with object-extraction double relatives, if the mild island effect seen among subject-extraction double relatives is due to adaptation (i.e., aggregating over items changing from low to high acceptability). In the current experimental setting, the pattern of surprisal values when

exposed to subject-extraction double relatives is predicted to be similar to when the models are exposed to object-extraction double relatives. In particular, the expected drop in licensing interaction values from -ISLAND to +ISLAND sentences should be similar between LICENSOR-POSITION: subject and LICENSOR-POSITION: object sentences.

3.3.6 Results

Results are summarized in the figure below.

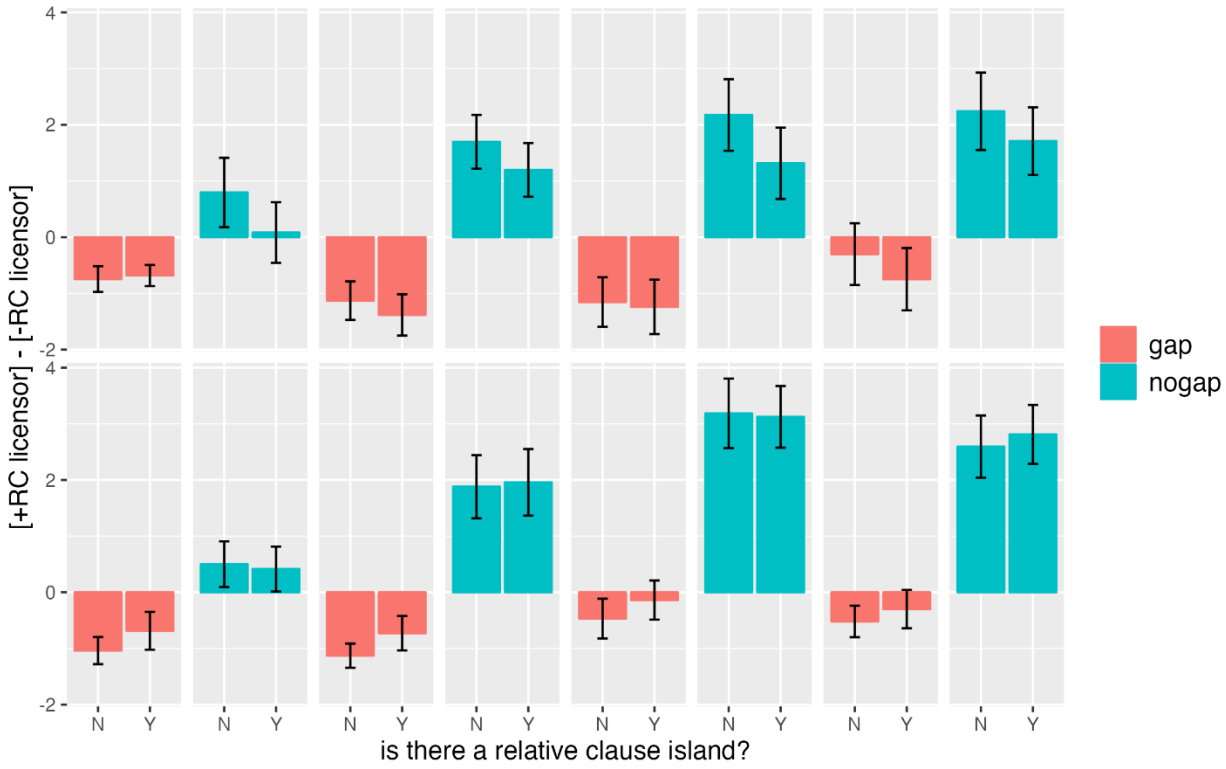

Figure 3.1: Values of subtracting mean surprisal of the -RC-LICENSOR condition from that of the +RC-LICENSOR condition by island status, extraction type (subject on top; object on bottom), and model size (from left: Large, Medium, Small, Xsmall). Error bars represent 95% confidence intervals

To reiterate, subtracting the mean surprisal of the -RC-LICENSOR condition from the +RC-LICENSOR condition should result in a positive value when there is no gap or the gap is filled

(e.g., *S* of (9a) – *S* of (8a)), and a negative value if there is a gap (e.g., *S* of (12a) – *S* of (11a)). Overall, that prediction seems to be borne out in Figure 3.1, indicating that the language models featured in this experiment have learned that a relative clause licensor needs a gap, and vice versa. In other words, the models have learned the gap-filler dependency created by long-distance relativization. The observation is confirmed by the linear mixed-effects model, which revealed a significant main effect of GAP such that the value of [-RC-LICENSOR] - [+RC-LICENSOR] was higher among -GAP than +GAP ($\beta = 2.32$, SE = 0.53, $p < 0.01$).

The previous section also predicted that the value of the licensing interaction among +ISLAND sentences would be smaller than that of -ISLAND sentences. Table 3.1 below lists the licensing interaction values between -ISLAND and +ISLAND sentences by model and licensor position (subject versus object).

Table 3.1: Mean licensing interaction values broken down by island status, extraction type, and model

	Large		Medium		Small		Xsmall	
	subject	object	subject	object	subject	object	subject	object
Non-island	1.54	1.54	3.01	2.82	3.65	3.33	3.11	2.54
Island	1.10	0.77	2.69	2.58	3.26	2.55	3.11	2.46
difference *p*-value	0.08	<0.01*	0.08	<0.05*	0.06	<0.001*	0.99	0.62

The value of licensing interaction is numerically smaller among +ISLAND sentences than -ISLAND sentences in most cases. Pairwise *t*-tests revealed that some of the differences are statistically significant. Importantly, however, cases where the difference is significant are limited to when sentences involve object-extraction double relatives ("object" columns in Table 3.1); in contrast, when sentences involve subject-extraction double relatives ("subject" columns

in Table 3.1), all the models except Xsmall had *p*-values that are slightly above the significance threshold ($\alpha = 0.05$).

3.3.7 Discussion

Results of the current experiment are in line with previous studies (Linzen et al., 2016; Bernardy & Lappin, 2017; Kuncoro et al., 2018; Gulordava et al., 2018; Futrell et al., 2018, 2019; Marvin & Linzen, 2018; Wilcox et al., 2018, 2023) that language models trained with a large volume of text can demonstrate their knowledge of long-distance dependencies. As far as I am aware, this is the first study to test such knowledge in Japanese, whereas almost all of the previous studies have focused on English. Also note that the dependency featured in this experiment is relativization in Japanese, which forms a gap-filler order, unlike the dependency featured in the previous studies, which has a filler-gap order (e.g., *wh*-movement in English). This experiment has therefore demonstrated that language models can learn long-distance dependencies regardless of language and the linear order of filler and gap.

In addition to the gap-filler dependency formed by relativization, the language models appear to know that the dependency cannot be formed across another relative cause, ending up in a double relative. The models demonstrated this knowledge with a drop in the value of licensing interaction when sentences involved a relative clause island, compared with the sentences that did not. Critically, while the drop was statistically significant when the sentences involved object extraction out of a relative clause island, it was only marginally significant when the sentences involved subject extraction out of the island.

Before discussing the implications of the pattern of licensing interaction values, it is necessary to address the clear effect of model on the output. As stated in Section 3.3.1, this

experiment used four versions of GPT-2 differing in the number of parameters. The effect of model manifested in two ways: First, the model with the smallest parameter size (XSmall) failed to exhibit knowledge of relative clause islands, unlike the other models. Second, Figure 3.1 shows a pattern that the larger a model is, the smaller the value of [-RC-LICENSOR] - [+RC-LICENSOR] is. To confirm the significance of this pattern, I added the size of model as another factor to the linear mixed-effects model presented in Section 3.3.4. The model revealed that the values of [-RC-LICENSOR] - [+RC-LICENSOR] were indeed significantly different depending on the model size, such that the Xsmall ($\beta = 1.35$, SE = 0.60, $p < 0.05$), Small ($\beta = 1.79$, SE = 0.60, $p < 0.01$), and Medium ($\beta = 1.355$, SE = 0.60, $p < 0.05$) models had higher values than the Large model.

In the field of large language model research, an increase in the number of parameters largely correlates with better performance, and it has been pointed out that the advancement of AI in general has been simply driven by the wider availability of computational resources, such as GPU, that enable the training of models with billions of parameters (Sutton, 2019). The results we are seeing here are in line with this pattern; it was only the model with the smallest number of parameters (37M) that failed to exhibit knowledge of relative clause islands. Bear in mind that the number of parameters can vary independently from the size of the dataset; all four models used in this study were exposed to an identical set of data during the training. Hence, it should be the internal representations that models picked up from the training data and stored on the parameters that made a difference, and my results suggest that a model requires at least ≈100M parameters (given that the Small model had 110M parameters) in order to successfully capture

knowledge of the relative clause island. The reason why the threshold lies at this number, however, is a topic that must be left for future research.[10]

While a model with a sophisticated architecture and a large number of parameters is considered generally advantageous, some studies have pointed out that such a model may deviate from behaviors observed among humans. Oh and Schuler (2023) found that Transformer models with a larger number of parameters produced surprisal values that are less parallel to human reading time data compared with smaller models. They attributed this pattern to the larger models' tendency to assign low surprisal values to open-class words, because of their extensive domain knowledge as a result of getting trained with a massive volume of Internet text, and storing the knowledge in the parameters. Such a pattern can be confirmed in our data; taking an open-class word (noun) *suiri shoosetsu* 'the mystery novel' as an example, the surprisal value assigned to the word in sentences like (8a) was 18.42 (Xsmall), 16.52 (Small), 15.74 (Medium), and 14.44 (Large). Thus, the effect of model size on the values of [-RC-LICENSOR] - [+RC-LICENSOR] and licensing interaction is not unexpected, especially given different parameter values. Nevertheless, all the models except XSmall exhibited parallel results, with a significant drop in licensing interaction values for the object-extraction cases, as well as a marginally significant drop in licensing interaction values for the subject-extraction cases.

[10] On the naïve assumption that the number of model parameters equates to the number of neurons in the brain, even Large model (with 1B parameters) would have a fraction of the number of neurons in the human brain, which is approximately 86B (Herculano-Houzel, 2009). Hence, the proposed threshold of ≈100M is unlikely to have anything to do with human biology. Keep in mind, though, that the actual number of neurons in charge of language processing in the cerebrum should be much smaller than the total number of neurons in the brain.

3.4 General discussion

The current study probed the grammatical knowledge of large language models trained with Japanese text, testing them on the gap-filler dependency formed by relativization and the constraint that the dependency cannot be formed across another relative clause. Chapter 2 demonstrated that the constraint does exist among Japanese speakers, regardless of whether extracting a subject or an object out of the relative clause island. The findings led to the proposal that violating the island itself gives rise to only a mild degradation in acceptability, as seen in the fact that subject-extraction double relatives have a significant yet small island effect, and their z-score acceptability is positive. While such an island effect has been observed in prior studies (and given the term "subliminal island"), the nature of those islands is still undetermined; some studies claim that what looks like a mild island effect is in fact not real, as it is not limited to filler-gap (or gap-filler) dependency, and it arises due to some processing costs (Keshev & Meltzer-Asscher, 2019), or aggregating over inter-item and/or inter-participant variability in judgments (Kush et al., 2018, 2019). The present chapter has proposed that collecting "judgments" from language models can contribute novel insights to this discussion, thanks to their property that they do not experience the adaptation effect in the way human subjects do. Using word-by-word surprisal and licensing interaction (with the latter derived from the former) as metrics to measure the models' grammatical knowledge, the experiment in this chapter has revealed that Japanese language models are in fact knowledgeable about gap-filler dependencies formed by relativization. Additionally, the decreased value of licensing interaction among sentences with a relative clause island is indicative of the models' sensitivity to the island (except for Xsmall model), but the decrease was statistically significant only when the sentences involved object extraction out of the island.

Let us now compare the results obtained from the language models with those obtained from human subjects in Chapter 2. Human subjects judged sentences with object-extraction double relatives as clearly ill-formed, while they judged those with subject-extraction double relatives as intermediate, even though both types of double relative showed a superadditive drop in acceptability suggestive of an island effect. On the other hand, language models showed a significant drop in licensing interaction values for sentences involving object-extraction double relatives. But for sentences involving subject extraction, although there was a drop in licensing interaction values, it was only marginally significant (see Table 3.1). The difference in metrics notwithstanding, there appear to be similarities in the pattern of results among humans and language models; the significant drop in licensing interaction values when sentences involved object-extraction double relatives corresponds to the structure triggering a clear island effect in terms of acceptability. How about the sentences with subject-extraction double relatives? It certainly requires caution to interpret *p*-values that are just above the < 0.05 threshold; taking only *p*-values into consideration, $p = 0.06$ is essentially the same as $p = 0.99$, which ought to be interpreted as "not significant". At the same time, Table 3.1 does show us a clear trend among all but Xsmall model with a drop in licensing interaction values when sentences involved subject-extraction double relative. Thus, a non-significant drop in the licensing interaction values, accompanied with clear numerical trends, could reflect not only knowledge of the relative clause island effect, but also knowledge that the effect is relatively mild. Recall that a language model's representation of linguistic knowledge depends primarily on what is in the training data, especially if the training is unsupervised. The non-significant drop in licensing interaction values therefore means that the training data of models in this experiment may have included sentences

with subject-extraction double relatives. That is not surprising, considering that such sentences have been traditionally judged to be well-formed, and thus they may be produced occasionally.

Assuming that the non-significant drop in licensing interaction values reflects the mild effect of violating a relative clause island, the fact that such a pattern of drop was observed among language models suggests that the mild island effect observed among humans is not due to some inter-item and/or inter-participant variability in judgments; instead, relative clauses may be an island that triggers only a mild drop in acceptability when violated. At the same time, the fact that there was a significant drop in licensing interaction values for object-extraction double relatives could suggest that the models have a grammatical constraint against extracting out of a relative clause island, whose penalty when violated is severe. Alternatively, in line with our proposal that the penalty itself is relatively mild, the models may have exhibited the knowledge of not only the constraint, but also processing biases such as ObS, considering that the latter have been known to affect surprisal values among language models in the same way that they affect various metrics among humans, such as acceptability and reading time (Smith & Levy 2013; Futrell et al., 2019; Wilcox et al., 2020, 2023; Shain et al., 2024).

3.5 Chapter summary

This chapter has investigated the nature of what appears to be a mild island effect; in particular, whether there is such an island type that triggers only a small degradation in acceptability when violated, or whether what looks like one is due to processing costs related to long-distance dependencies, or variability in judgments across participants and/or items. I proposed leveraging large language models to address this question because of their properties, as well as metrics such as surprisal values that we can obtain from those models, in order to

probe their knowledge of dependencies formed by relativization, and the constraint that the dependency may not cross another relative clause. All of the language models I tested demonstrated knowledge of the dependency, and almost all of them (except the model with the smallest number of parameters) appear to have learned the relative clause island constraint as well. Importantly, the knowledge of the constraint was clearer when the relevant sentences involved object extraction out of a relative clause, than when they involved subject extraction out of a relative clause. Such an asymmetry resembles the pattern of acceptability judgments made by human subjects in Chapter 2. And most importantly, the fact that models exhibited somewhat weak signals about the relative clause island violated by subject extraction, despite their immunity to variability across participants and items, can be interpreted as a sign that relative clauses indeed trigger only a mild effect.

Chapter 4

Double relatives of English

4.1 Introduction

Towards the end of Chapter 2, I discussed the status of relative clause islands in English. As mentioned there, the relativization out of another relative clause in English appears to be not very acceptable. The following sentences exemplify double relatives in English:

(1) a. *The professor$_i$ [$_{RC2}$ who the sci-fi novel$_j$ [$_{RC1}$ that __$_i$ wrote __$_j$] was recently featured in a bookstore] looks proud.

b. *The sci-fi novel$_j$ [$_{RC2}$ that the professor$_i$ [$_{RC1}$ who __$_i$ wrote __$_j$] was recently featured in a bookstore] is popular among college students.

(1a) is an instance of a subject-extraction double relative, and (1b) exemplifies an object-extraction double relative. Chapter 2 demonstrated that the former in Japanese is judged to be more acceptable than the latter, although both do exhibit a relative clause island effect of a comparable size. The difference in acceptability was attributed to a processing factor called Object before Subject Bias (ObS), which was formulated as follows:

(2) *Object before Subject Bias* (ObS)

> When long-distance dependencies between arguments and predicates are being resolved, the dependency associated with the object argument must be fully resolved before the dependency associated with the subject argument can be.

According to ObS, sentences with subject-extraction double relatives in Japanese are relatively easy to process, as the dependency between the object and the verb is adjacent and fully resolved first, even before encountering the subject. In contrast, sentences with object-extraction double relatives involve the subject and the verb encountered first, but as the semantic role of the subject is contingent on the object, there is a significant delay in fully resolving the subject dependency until the object is identified. Put differently, only subject-extraction double relatives satisfy ObS, and the resulting effect on acceptability is as documented in Chapter 2.

Applying ObS to the English cases, in sentences with subject-extraction double relatives (e.g., (1a)), the parser encounters the subject and then the object, followed by the verb. Likewise, in sentences with object-extraction double relatives (e.g., (1b)), the object and then the subject are encountered before the verb. Thus, in both cases, the subject and object dependencies are resolved simultaneously. Thus, ObS does not predict any asymmetry in acceptability between subject- and object-extraction double relatives. In line with this, the two sentences in (1) do not immediately give the appearance of having drastically different acceptability. The processing pattern of English double relatives is different from the Japanese ones, where dependencies are encountered and fully resolved one at a time without delays (subject-extraction double relative), or where the verb and the subject are encountered but the subject dependency remains not fully resolved until the object appears (object-extraction double relative).

Note, however, that there are several factors that could be making the acceptability of English double relatives as in (1) particularly low. First of all, both sentences in (1) involve

extraction out of the (passive) subject, which generally leads to unacceptability in English (Ross, 1967; Chomsky, 1973, 1986; Huang, 1982), whereas it does not in Japanese (Omaki et al., 2020). We can avoid the violation of subject island and construct double relatives like (3) below. Given that the subject island effect in English has been found and replicated across experiments (Sprouse, 2007; Sprouse et al., 2011; Sprouse et al., 2012), double relatives like the one in (3) could be judged as more acceptable than the ones in (1).

(3) *?The professor$_i$ [$_{RC2}$ who I liked the sci-fi novel$_j$ [$_{RC1}$ that __$_i$ wrote __$_j$]] looks proud.

And yet, (3) remains unacceptable, possibly because of another factor: There is a long-distance extraction out of the embedded subject position in the presence of overt material in within a CP layer, which is likely an instance of the COMP-trace effect (e.g., Pesetsky, 2017). Once we correct for the COMP-trace effect by targeting the embedded object position for the long-distance extraction, a sentence with a double relative looks as follows:

(4) The sci-fi novel$_j$ [$_{RC2}$ that I liked the professor$_i$ [$_{RC1}$ who __$_i$ wrote __$_j$]] sold very well.

Importantly, there is an additional difference between (3) and (4) besides the existence of COMP-trace effect; while the double relative in (3) involves two dependencies (formed by relativization) that are crossing, (4) involves nested dependencies, as illustrated in (5) below.[1]

[1] It is worth noting that the crossing versus nested dependencies in (5) are in terms of syntactic representation; they represent the relativized NPs and their original positions. In terms of parsing, the two dependencies are similar in that they are resolved simultaneously (when *wrote* is encountered). In contrast, in sentences such as (6) below, there is a crossing versus nested distinction in both the syntactic representation and in parsing.

(5) a. The professor$_i$ [$_{RC2}$ who I liked the sci-fi novel$_j$ [$_{RC1}$ that __$_i$ wrote __$_j$]] looks proud.

b. The sci-fi novel$_j$ [$_{RC2}$ that I liked the professor$_i$ [$_{RC1}$ who __$_i$ wrote __$_j$]] sold very well.

Cross-linguistically, crossing dependencies are known to be harder to process than nested dependencies (Fodor, 1978; Frazier & Fodor, 1978; Rochemont & Culicover, 1990; Pickering & Barry, 1991), which possibly explains why the former is rarer than the latter (Steedman, 1985). In (6) and (7) below, for instance, the sentence with a crossing dependency is reported to be more difficult to process than the sentence with a nested dependency.

(6) a. Which sonatas$_i$ are these violins$_j$ easy to play t_i on t_j? *crossing*
b. Which violins$_j$ are these sonatas$_i$ easy to play t_i on t_j? *nested*

(7) a. Who$_i$ do you know what books$_j$ to persuade t_i to read t_j? *crossing*
b. What books$_j$ do you know who$_i$ to persuade t_i to read t_j? *nested*

In sum, the acceptability of English double relatives could be negatively affected by the subject island effect, and in addition, the acceptability of double relatives like (1a) could be particularly low because of the COMP-trace effect and the fact that there is a crossing dependency. Notice that the last two factors co-occur in (1); double relatives without the COMP-trace effect (=1b) form nested dependencies, while those with a COMP-trace effect (=1a) form crossing dependencies. Through two acceptability judgment experiments presented in the subsequent sections, we will attempt to tease apart the two confounding factors and measure the extent to which they impact the acceptability of double relatives in English. Section 4.2 presents Experiment 1, which examines the acceptability of sentences differing in whether there is a relative clause island violation (i.e., whether a sentence has a double relative), and the presence

of a COMP-trace effect. The condition with both a relative clause island violation and a COMP-trace effect is also the one with crossing dependencies. As such, if the crossing dependency type impacts acceptability on top of the other two factors, the combination of a relative clause island violation and a COMP-trace effect will be superadditive, statistically manifesting as an interaction effect. Section 4.3 follows up on the results of Experiment 1 with Experiment 2, which features sentences with an embedded ditransitive verb. By relativizing its arguments (direct and indirect objects), it becomes possible to compare double relatives with crossing versus nested dependencies without the former also invoking a COMP-trace effect. The two experiments demonstrate that, while the COMP-trace effect negatively affects the acceptability of English double relatives, whether double relatives involve crossing or nested dependencies does not. Crucially, even after controlling for the COMP-trace effect, the acceptability of English double relatives remains low. Section 4.4 discusses potential causes for this low acceptability, where we propose that English double relatives are particularly challenging for the parser as they involve resolving multiple long-distance dependencies at once.

4.2 Experiment 1

4.2.1 Participants

36 participants, who self-reported that they spoke English as (one of) their first language(s), were recruited for this experiment. They all passed the attention check procedure described below, and thus all of their data were included for analysis (age range = 19-30, mean = 21.2) Participants received course credits for completing the experiment through UC San Diego's SONA system, an online human subject pool management platform.

4.2.2 Materials

All critical stimuli contain a relative clause modifying the main clause object (headed by *children* in the sample provided in (8) below), but this relative clause varies according to a 2 x 2 factorial design: Long-distance extraction (relativization) of the head noun is either from a complementizer (*that*) clause, which is -ISLAND, or from another relative clause, which is +ISLAND. In addition, the extraction is either from the embedded object position (-COMP-TRACE; indicated with the gap indexed *j* in (8)), or the embedded subject position (+COMP-TRACE; indicated with the gap indexed *i*). Since the embedded clause is introduced by a complementizer, the subject extraction leads to a COMP-trace effect. Note that +ISLAND sentences (i.e., sentences with double relatives) involve two dependencies, while -ISLAND sentences involve one. As such, while Condition 2 involves only the COMP-trace effect, Condition 4 involves the relative clause island violation in addition to the COMP-trace effect. And between the two +ISLAND conditions, Condition 3 involves the long-distance relativization of the object, which does not trigger a COMP-trace effect. Importantly, the dependencies formed by the object and the one formed by the subject in Condition 3 are nested, as the former is longer. In contrast, Condition 4 involves the long-distance relativization of the subject, which is an instance of the COMP-trace effect. And by relativizing the subject for a longer distance than the object, Condition 4 also ends up with crossing dependencies.

(8) Condition 1 ISLAND: -; COMP-TRACE: -
I waved at the children$_j$ [$_{RC}$ who I believe [that the teacher is running with __$_j$]].

Condition 2 ISLAND: -; COMP-TRACE: +
I waved at the children$_i$ [$_{RC}$ who I believe [that __$_i$ are running with the teacher]].

Condition 3 ISLAND: +; COMP-TRACE: - (nested dependency)
I waved at the children$_j$ [$_{RC}$ who I know the teacher$_i$ [$_{RC}$ who __$_i$ **is** running with __$_j$]].

Condition 4 ISLAND: +; COMP-TRACE: + (crossing dependency)
I waved at the children$_i$ [$_{RC}$ who I know the teacher$_j$ [$_{RC}$ that __$_i$ **are** running with __$_j$]].

For +ISLAND conditions, the inner relative clause (headed by *the teacher*) was placed in the object position in order to prevent the additional violation of the subject island. As (8) exemplifies, stimuli in Condition 3 and 4 were identical except for the number feature of the copula *be* inside the inner relative clause. The copula was used in order to avoid the ambiguity as to which of the relativized noun phrases (e.g., *children* or *teacher*) is the subject of the inner relative clause verb. All the nouns except for the ones that are relativized in Condition 3 and 4 were the first-person singular pronoun *I*, in order to avoid any processing cost of introducing multiple full noun phrases in a sentence.

20 lexically-matched sets as in (8) were created. Stimuli were counterbalanced through a Latin-square procedure, resulting in 4 lists (5 items per condition; 20 items per list). 40 fillers were also created, consisting of sentences with varying degrees of acceptability: 10 fillers of expected high acceptability, 10 of intermediate acceptability (e.g., sentences with center-embedding), and 20 of low acceptability (e.g., sentences violating the Coordinate Structure Constraint). Fillers were identical across lists, and each of the lists consisted of 60 items. The order of stimuli was pseudo-randomized such that two critical items never appeared in a row. The full sets of stimuli for experiments in this chapter are available at the following Open Science Framework page: https://osf.io/m9z6p/.

4.2.3 Procedure

The experiment was hosted on PCIbex (Zehr & Schwarz, 2018). Participants were instructed to rate how natural each sentence sounded by clicking on a number on a scale from 1

(very unnatural) to 7 (very natural). Unlike the experiments in Chapter 3, participants did not have to complete the language background questionnaire as they had been pre-screened with questions like "Are you a native English speaker?" offered by the SONA system.

To screen out participants who were not attending to the task, responses to the 10 filler items with the highest mean acceptability scores across all participants and the 10 with the lowest scores were identified. Participants whose ratings were more than 2 standard deviations away from the mean for 5 or more of these 20 items were excluded from further analysis. No one was filtered out in this way for this experiment, and thus everyone in the initial group of 36 participants was included in the final dataset (with 9 participants in each of the 4 lists).

4.2.4 Data analysis

Raw acceptability scores were converted to z-scores prior to analysis. A linear mixed-effects regression model was created using the lmerTest package (Kuznetsova et al., 2017) in R (R Core Team, 2024). The model predicts the acceptability (in z-scores) of sentences as a function of relativization across an embedded clause and the position from which the relativization took place. Since the maximal model (as recommended in Barr et al., 2013) did not converge, I used random intercepts for participant and item, and random slopes of ISLAND for participant and item.

4.2.5 Predictions

First of all, given the well-documented effect of violating an island on sentence acceptability, there should be a main effect of ISLAND, such that the +ISLAND conditions (Condition 3, 4) are rated lower in acceptability than -ISLAND conditions (Condition 1, 2), as the

former involve a relative clause island violation. Likewise, as the negative effect of the COMP-trace effect on acceptability has also been found and replicated (e.g., Cowart, 1997, 2003), there should be a main effect of COMP-TRACE such that +COMP-TRACE conditions (Condition 2, 4) are judged to be less acceptable than -COMP-TRACE conditions (Condition 1, 3). In addition to the main effects of the two manipulated factors, recall that the two +ISLAND conditions also differ in that Condition 3 (-COMP-TRACE) involves nested dependencies, while Condition 4 (+COMP-TRACE) involves crossing dependencies. How would the further contrast of crossing and nested dependencies affect acceptability? Two possible outcomes are illustrated in Figure 4.1. If crossing dependencies do not lower acceptability independently from the COMP-trace effect, the drop in acceptability from -ISLAND conditions to +ISLAND conditions should look parallel between Condition 1 and 3 (-COMP-TRACE) and Condition 2 and 4 (+COMP-TRACE), similar to the left panel in Figure 4.1. Statistically, there should only be main effects of COMP-TRACE and ISLAND. In contrast, if crossing dependencies formed by double relatives lower acceptability on top of the COMP-trace effect, the drop in acceptability in Condition 4 is expected to exhibit a superadditive effect, where the drop is larger going from Condition 2 (+COMP-TRACE, -ISLAND) to 4 (+COMP-TRACE, +ISLAND) than going from Condition 1 (-COMP-TRACE, -ISLAND) to 3 (-COMP-TRACE, +ISLAND), similar to the right panel in Figure 4.1. This superadditive pattern of acceptability should be manifested as an interaction effect between ISLAND and COMP-TRACE, suggesting that the drop in acceptability cannot be simply attributed to the combination of relative clause island effect and the COMP-trace effect.

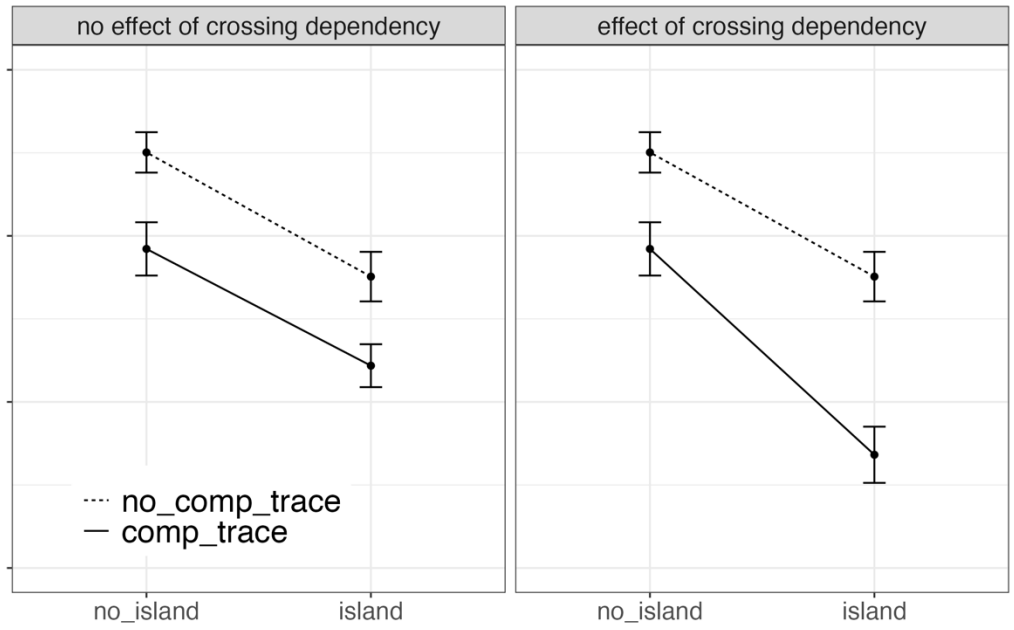

Figure 4.1: Predicted outcome if there is no effect of crossing dependency (left) vs. if there is an effect of crossed dependency (right)

4.2.6 Results

Mean z-scores for the four conditions of Experiment 1 are presented in Figure 4.2.

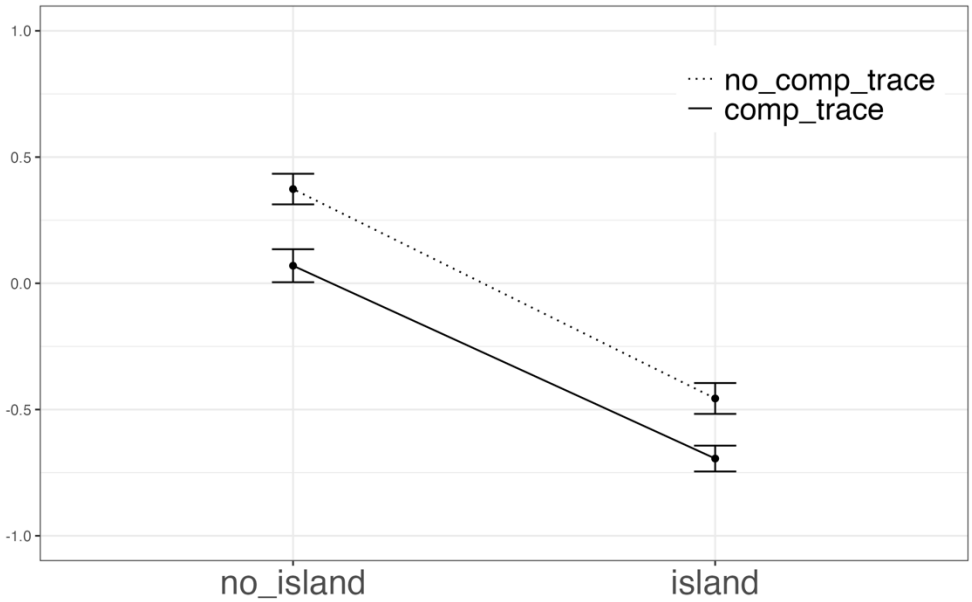

Figure 4.2: Aggregated responses to critical items in Experiment 1

The model revealed a significant main effect of ISLAND in the expected direction; +ISLAND sentences were rated lower than their -ISLAND counterparts (β = 0.83, SE = 0.11, p < 0.001). There was also the main effect of COMP-TRACE (β = -0.23, SE = 0.08, p < 0.01), such that +COMP-TRACE sentences were rated lower than -COMP-TRACE sentences, also as predicted. But crucially, the interaction between these two factors was not significant (β = -0.08, SE = 0.11, p = 0.44). That is, the decline in acceptability for double relatives with a crossing dependency was no greater than the decline predicted by the additive effect of the relative clause island violation and COMP-trace effect. As the previous section laid out, this pattern does not support an independent effect of dependency type (crossing versus nested) on acceptability.

4.2.7 Discussion

The outcome of Experiment 1 demonstrates that, first of all, the acceptability of double relatives in English is low across the board; the mean acceptability of +ISLAND conditions Condition 4 was -0.69 (+COMP-TRACE) and -0.46 (-COMP-TRACE), respectively, which is closer to the mean acceptability of "bad" fillers (-0.70) than the mean acceptability of "intermediate" fillers (0.45). Furthermore, the COMP-trace effect was also shown to negatively affect acceptability. And the COMP-trace effect had parallel effects between non-island and island sentences, closely resembling the left panel of Figure 4.1; the drop in acceptability from Condition 1 to 2 seems identical to the drop in acceptability from Condition 3 to 4. This pattern was statistically confirmed by the main effects of both ISLAND and COMP-TRACE, but the lack of an interaction between them.

Crucially, the parallel COMP-trace effect regardless of island status goes against what I predicted if crossed dependencies play an additional role in modulating acceptability; under the

prediction, the drop in acceptability due to the COMP-trace effect from Condition 3 to 4 was expected to be larger than the drop in acceptability from Condition 1 to 2, since Condition 4 also involves crossed dependencies.

A possibility remains, however, that a floor effect may be masking a true interaction. As the results show, sentences with double relatives that also involve a COMP-trace effect had rather low acceptability. If such sentences were indeed rated near the lowest possible acceptability, there may not have been sufficient room left on the scale to reflect the further drop in acceptability due to crossing dependencies. The possibility of floor effect is challenged, however, by looking at the distribution of mean acceptability of individual stimuli on Figure 4.3.

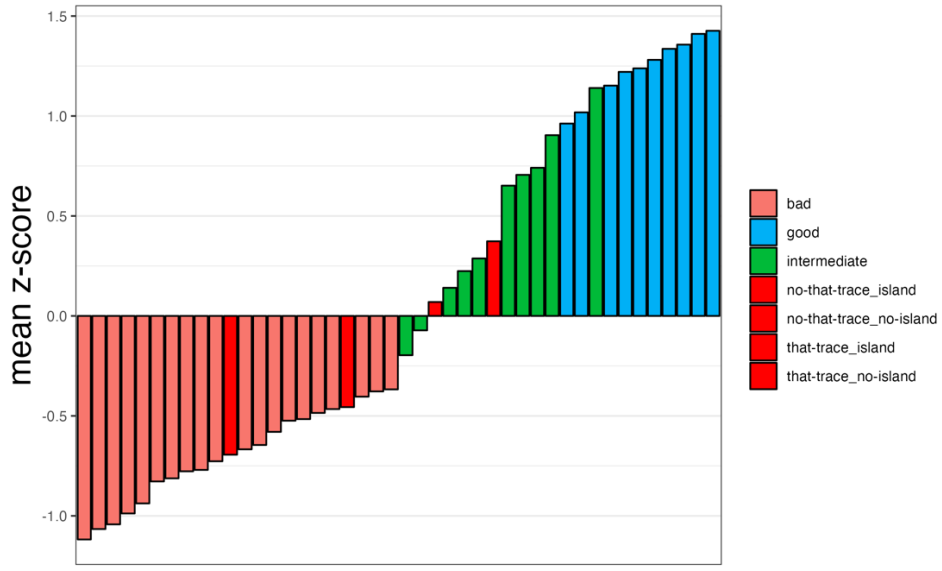

Figure 4.3: Mean z-scores of individual stimuli from Experiment 1

The mean acceptability of critical items is indicated as red bars, and the leftmost red bar indicates the mean acceptability of sentences with double relatives and the COMP-trace effect/crossing dependencies. Notice that there are several fillers whose mean acceptability was even lower than such double relative sentences. Hence, the sentences were not given the lowest

possible acceptability, and if the crossing dependencies had had any negative effect, that could have been reflected in the acceptability ratings. The fact that it did not suggests that this type of dependency is unlikely to significantly lower acceptability.

Despite the fact that a floor effect is an unlikely explanation for the results, it would still be ideal to test the effect of crossing versus nested dependencies in isolation, rather than in conjunction with the COMP-trace effect, if that is possible. Stimuli in this experiment featured double relatives derived by subject and object extraction, in order to make them similar in structure to the Japanese double relatives from the previous chapters. As we have seen, though, long-distance relativization of the subject inevitably leads to a COMP-trace effect in the case of English. Alternatively, we can test the effect of crossing versus nested dependencies more directly with double relatives derived by relativizing the arguments of a ditransitive verb (direct and indirect objects), which does not trigger a COMP-trace effect. The next experiment thus features such double relatives.

4.3 Experiment 2

4.3.1 Participants

A new group of 38 participants, who self-reported that they spoke English as (one of) their first language(s), were recruited for this experiment on the SONA system. The same attention check procedure based on the fillers as in Experiment 1 was used and two participants were excluded in this way, resulting in 36 participants (age range = 19-30, mean = 21.3) whose responses were analyzed.

4.3.2 Materials

The stimuli had the same structure as in Experiment 1, except that the main clause object originated as a direct object or an indirect object of the embedded ditransitive verb, as in the sample stimuli in (9) below. Stimuli varied according to a 2 x 2 factorial design: Relativization was either from a complementizer (*that*) clause/non-island or from another relative clause/island (±ISLAND), and either from the direct object position or the indirect object position (POSITION). As the long-distance relativization is no longer taking place out of the subject position in any of the conditions, the COMP-trace effect is irrelevant; instead, different extraction positions lead to forming either a nested dependency (if out of the indirect object position) or a crossing dependency (if out of the direct object position).

(9) Condition 1 ISLAND: -; POSITION: indirect object
I praised the teacher$_j$ [$_{RC}$ that I believe [that I gave the book to __$_j$ last year]].

Condition 2 ISLAND: -; POSITION: direct object
I praised the book$_i$ [$_{RC}$ that I believe [that I gave __$_i$ to the teacher last year]].

Condition 3 ISLAND: +; POSITION: indirect object (nested dependency)
I praised the teacher$_j$ [$_{RC}$ that I remember the book$_i$ [$_{RC}$ that I gave __$_i$ to __$_j$ last year]].

Condition 4 ISLAND: +; POSITION: direct object (crossing dependency)
I praised the book$_i$ [$_{RC}$ that I remember the teacher$_j$ [$_{RC}$ that I gave __$_i$ to __$_j$ last year]].

Since most of the common ditransitive verbs involve an inanimate noun as the direct object and an animate noun as the indirect object, the two arguments that underwent relativization in Experiment 2 differed in animacy (e.g., *teacher* and *book* in (9)), unlike in Experiment 1 (e.g., *children* and *teacher* in (8)). Hence, care was taken to select main clause verbs that are equally compatible with both animate and inanimate objects (e.g., *praise the teacher* and *praise the book*).

The current experiment used the same set of fillers, as well as the pseudo-randomization method, as Experiment 1.

4.3.3 Procedure

The same experimental procedures were used as in Experiment 1.

4.3.4 Data analysis

As in Experiment 1, a linear mixed-effect model with random effects of subject and item was used to fit the data and test for significance. The model predicts the acceptability (in z-scores) of sentences as a function of relativization across an embedded clause and the position from which the relativization took place. The model that converged used random intercepts for participant and item.

4.3.5 Predictions

Given the results of Experiment 1, as well as the recurrent findings in the literature, the main effect of ISLAND is expected to be significant, such that the conditions with a relative clause island violation (Condition 3, 4) are rated lower in acceptability than -ISLAND conditions (Condition 1, 2). On top of the expected effect of relative clause island violation, if a crossing dependency further lowers acceptability, I predict there to be a significant interaction effect of ISLAND and POSITION; while there should not be a major difference in acceptability between Condition 1 and 2 (as neither involves multiple filler-gap dependencies), Condition 4 should be rated even lower than Condition 3.

4.3.6 Results

Mean z-scores for the four conditions of Experiment 2 are presented in Figure 4.4.

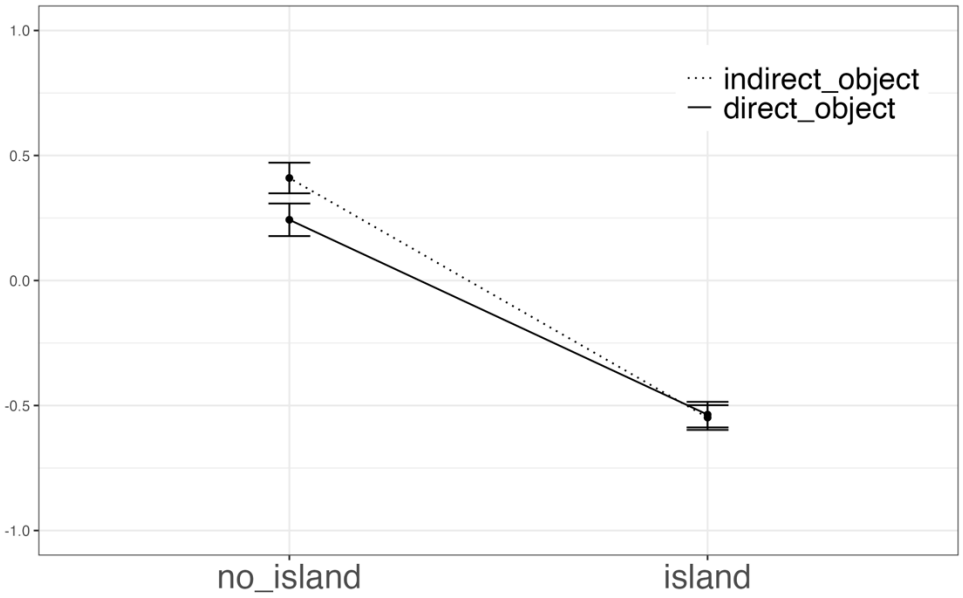

Figure 4.4: Aggregated responses to critical items in Experiment 2

The model revealed a significant main effect of ISLAND in the expected direction; +ISLAND sentences were rated lower than their -ISLAND counterparts (β = 0.79, SE = 0.10, p < 0.001). There was no main effect of POSITION (β = -0.01, SE = 0.09, p = 0.94), and the interaction between ISLAND and POSITION was not significant either (β = 0.17, SE = 0.11, p = 0.12). Hence, as in Experiment 1, this experiment failed to find an effect of crossing versus nested dependencies on acceptability.

4.3.7 Discussion

Results of the present experiment resemble those of Experiment 1, in that it replicated the effect of the relative clause island violation, as well as the lack of the effect of crossing versus nested dependencies. As in Experiment 1, as the sentences here with a crossing dependency are

already in violation of an island, any additional negative effect of a crossing dependency may not have been captured. Figure 4.5 below shows the distribution of mean acceptability of individual stimuli.

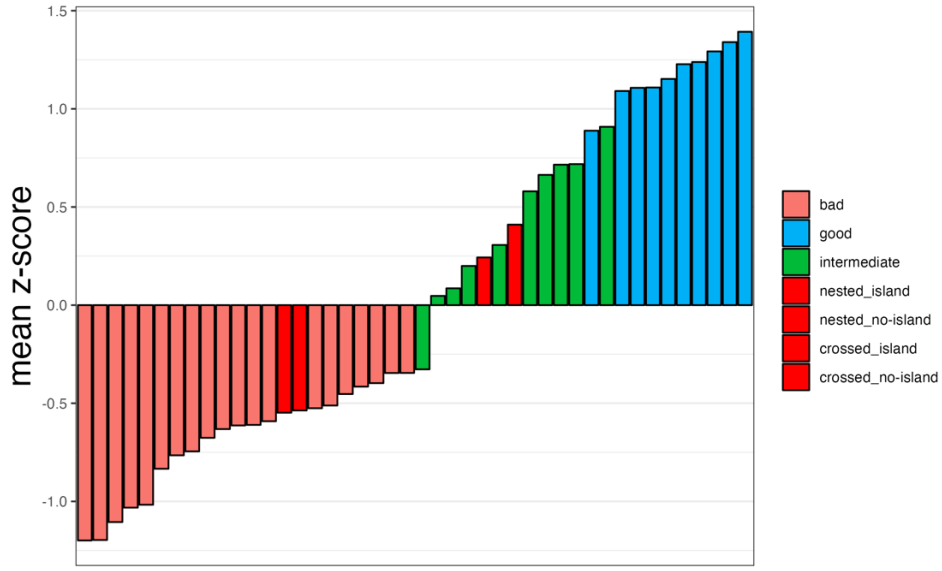

Figure 4.5: Mean z-scores of individual stimuli from Experiment 2

As there are several fillers that had even lower acceptability than sentences with double relatives containing a crossing dependency, it should have been possible for participants to express a preference for a nested over a crossing dependency. The fact that they instead gave highly similar ratings to the two types of double relatives suggests that crossing dependencies do not further degrade the acceptability of English double relatives.

Through the two experiments, it has become evident that crossing dependencies do not lower acceptability relative to a nested dependency baseline in English; double relatives are judged as low in acceptability regardless of dependency type. The next section discusses potential reasons why the effect of crossing versus nested dependencies was not detected in our experiments. It also reviews the possible factors that may contribute to the low acceptability of

English double relatives, referencing the object before subject (ObS) bias that was proposed for Japanese double relatives.

4.4 General discussion

4.4.1 Absence of the effect of crossing versus nested dependencies

Both of our experiments have demonstrated that whether dependencies are crossing or nested does not affect the acceptability of English double relatives. This finding is perhaps surprising, given the evidence that nested dependencies are processed more easily than crossing dependencies, as well as the evidence that the type of dependencies affects acceptability (Kuno & Robinson, 1972; Pesetsky, 1982), as in the pairs of sentences below (repeated from (6) and (7) in Section 4.1).

(10) a. Which sonatas$_i$ are these violins$_j$ easy to play __$_i$ on __$_j$? *crossing*
b. Which violins$_j$ are these sonatas$_i$ easy to play __$_i$ on __$_j$? *nested*

(11) a. Who$_i$ do you know what books$_j$ to persuade __$_i$ to read __$_j$? *crossing*
b. What books$_j$ do you know who$_i$ to persuade __$_i$ to read __$_j$? *nested*

One difference between the sentences above and the sentences with double relatives is possibly the ease of projecting the gaps (indicated with underscores). In both (10) and (11), the extraction takes place out of an infinitival clause, which is known to cause a smaller island effect than a finite *wh*-clause (Chomsky, 1986).[2] Hence, even though the *wh*-island is violated in both (10) and (11), the parser may still project the gaps inside the *wh*-island with confidence, which in

[2] (10) is an instance of the so-called *tough*-construction, which is thought to involve a null *wh*-operator (Chomsky, 1977). Under this analysis, the *tough*-movement in (10) is indeed an instance of *wh*-island violation on par with (11).

turn allows the preference for a nested dependency over a crossing one to play a role. Conversely, double relatives in our experiments involve not only the extraction out of an island, but also out of a finite island. Considering this, the parser may not be able to posit the gaps with enough confidence, which masks the crossing versus nested distinction. As evidence of this claim, as Chapter 3 showed, humans as well as large language models become less likely to expect a gap upon finding out that doing so would lead to forming a dependency across an island (Stowe, 1986; Traxler and Pickering, 1996; Phillips, 2006; Wilcox et al., 2018). Furthermore, the pair of sentences with double relatives presented below involve relativization out of an infinitive/non-finite clause.

(12) a. ? I just bought a violin$_j$ [that I composed a sonata$_i$ [to play __$_i$ on __$_j$]].
b. ?* I just composed a sonata$_i$ [that I bought a violin$_j$ [to play __$_i$ on __$_j$]].

The difference in acceptability between the sentences above may now be easier to observe, such that the sentence with nested dependencies is more acceptable than the one with crossed dependencies.

In summary, although the distinction between nested and crossing dependencies may have affected acceptability in our experiments, it was perhaps too subtle to be detected (the absence of a floor effect notwithstanding).[3]

[3] Nakamura and Miyamoto (2013) reach a similar conclusion about Japanese double relatives, namely that the crossing versus nested distinction is irrelevant to why some double relatives are judged to be less acceptable than others. As presented in Section 2, they claim that processing difficulty (and presumably a drop in acceptability) arises when there is a delay in semantic role assignment, rather than when dependencies are crossing. They refer to Bach et al.'s (1986) experiment on crossing versus nested dependencies in Dutch and German, where a sentence with multiple embeddings results in a crossing dependency in Dutch and a nested dependency in German, exemplified as (ia) and (ib) below.

(i)	a.	Arnim	heeft	Wolfgang	de lerares	de knikkers	laten
		Arnim	had	Wolfgang	the teacher	the marbles	let

4.4.2 Sources of low acceptability of English double relatives

Initially, we presented examples of English double relatives such as (13), which are structurally similar to the ones in Japanese featured in the previous chapters.

(13) a. *The professor$_i$ [$_{RC2}$ who the sci-fi novel$_j$ [$_{RC1}$ that __$_i$ wrote __$_j$] was recently featured in a bookstore] looks proud.

b. *The sci-fi novel$_j$ [$_{RC2}$ that the professor$_i$ [$_{RC1}$ who __$_i$ wrote __$_j$] was recently featured in a bookstore] is popular among college students.

However, Section 4.1 pointed out that sentences such as (13) involve several potential factors that contribute to their low acceptability: A subject island effect, a COMP-trace effect, and crossing dependencies. The subject island effect, which is known to be robust in English (Ross, 1967; Chomsky, 1973, 1986; Huang, 1982) but is either very small or non-existent in Japanese (Ross, 1967; Kuno, 1973; Saito, 1985, 1992; Lasnik & Saito, 1992; Omaki et al., 2020), was avoided in both experiments in this chapter by placing the relative clause island in object

---

	helpen	opruimen.				
	help	clean.up				
b.	Arnim	hat	Wolfgang	der Lehrerin	die Murmeln	aufräumen
	Arnim	had	Wolfgang	the teacher	the marbles	clean.up
	helfen	lassen.				
	help	let				

'Arnim let Wolfgang help the teacher collect up the marbles.'

They report that participants judged Dutch sentences with a crossing dependency as more acceptable than the parallel German sentences with a nested dependency. Bach et al. attributed the preference for a crossing dependency to the fact that, even though the German parser can build the clause consisting of the innermost noun (*teacher*) and its verb (*clean up*) right away, readers need to wait until the end of the sentence, when they see the outermost verb *let*, to figure out where the embedded clauses should be attached to in a sentence. In contrast, the Dutch parser builds the clause consisting of the outermost noun (*Armin*) and its verb (*let*) first, which also informs them where the other embedded clauses belong in a sentence.

Applying this analysis to the stimuli with double relatives in Japanese, Nakamura and Miyamoto conclude that the crossing versus nested distinction is irrelevant to Japanese double relatives, since the main clause predicate (e.g., *hokorashige-da* 'looks proud' in (14b)) is located at the end of a sentence, regardless of whether the double relatives involve a crossing or a nested dependency.

position. Moreover, Experiment 1 evaluated the effect of the COMP-trace effect, which is triggered by relativizing the subject out of an embedded clause introduced by an overt complementizer. While the COMP-trace effect is attested in various languages with overt complementizers, extraction of the embedded subject in Japanese does not result in a violation of the COMP-trace, given the lack of an overt relative clause pronoun. Thus, in Japanese, there is no difference between the equivalents of Conditions 3 and 4 in Experiment 1 in terms of acceptability. This is why it is not problematic that an instance of double relatives that has high acceptability involves an embedded subject extraction, as in (14) below. The English counterpart of the sentence (=13a), however, is clearly unacceptable.

(14) [$_{RC2}$ [$_{RC1}$ __$_i$ __$_j$ kai-ta] SF-shoosetsu$_j$-ga saikin shoten-de
write-PST Sci-Fi novel-NOM recently bookstore-at
tokushuu-sa-re-ta] gakusha$_i$-wa hokorashige-da.
feature-do-PASS-PST professor-TOP looks.proud-COP
'The professor$_i$ [$_{RC2}$ who the sci-fi novel$_j$ [$_{RC1}$ that __$_i$ wrote __$_j$] was recently featured in a bookstore] looks proud.'

Whereas the COMP-trace effect turned out to play a significant role in the acceptability of English double relatives, whether the dependencies are crossing or nested did not, as discussed in the previous section; English double relatives with both types of dependencies were rated equally low (with mean z-scores around -0.5), as exemplified below.

Given that the acceptability of English double relatives remains low even after controlling for multiple factors, one possible conclusion is that the penalty of violating the relative clause island is simply more severe in English than in languages such as Japanese, where some double relatives are judged relatively well-formed. Alternatively, there may be other factors that make English double relatives particularly ill-formed, which enables us to maintain

that the penalty of violating the relative clause island itself is no more severe than in Japanese. In accounting for the reason why some Japanese double relatives are more acceptable than others, we proposed ObS, whose definition is repeated from (2) below.

(15) *Object before Subject Bias* (ObS)

> When long-distance dependencies between arguments and predicates are being resolved, the dependency associated with the object argument must be fully resolved before the dependency associated with the subject argument can be.

As Chapter 2 discussed, ObS is derived from two principles: Firstly, resolving long-distance dependencies leads to a processing cost, and it becomes more costly the longer they are or the more dependencies the parser must resolve. Secondly, assigning a semantic role to an object must precede assigning a semantic role to subject, since the latter depends on the former (and the verb). Combining the two, we proposed ObS, which explains why sentences with long-distance subject extraction Japanese are easier to process (and thus higher in acceptability) than long-distance object extraction. Section 4.1 mentioned that English double relatives are unlike both types of Japanese double relatives; they involve encountering the relativized subject and object (in either order) first, followed by the verb in the innermost clause, at which point the parser needs to resolve two dependencies simultaneously. This, however, means that the parser has the necessary information about the object upon encountering the verb, and thus there is no necessary delay in assigning the subject role, unlike in object-extraction double relatives in Japanese. Therefore, while ObS does not lead us to expect an asymmetry in the acceptability of subject- and object-extraction double relatives in English, it does not capture their low acceptability.

We propose that the source of low acceptability of English double relatives is the need to resolve multiple argument-predicate dependencies at once. Resolving a long-distance dependency involves integrating a dislocated element back into the phrase structure, which has been shown to incur processing cost (Gibson, 1991, 1998). And such a cost may be compounded when the integration must proceed in parallel for multiple dislocated elements. Let us call this principle the One Dependency at a Time Bias, as formulated below.

(16) *One Dependency at a Time Bias*

When long-distance dependencies between arguments and predicates are being resolved, the dependencies must be resolved one at a time, instead of being resolved simultaneously.

In Japanese double relatives, subject and object dependencies are resolved one at a time, although the subject dependency is not *fully* resolved until the object appears in the case of object-extraction double relatives. This account based on the simultaneous resolution of multiple dependencies also explains why the acceptability of double relatives remained low in Experiment 2, as in the sentences below (repeated from (9)).

(17) a. I praised the teacher$_j$ [$_{RC}$ that I remember the book$_i$ [$_{RC}$ that I gave __$_i$ to __$_j$ last year]].
b. I praised the book$_i$ [$_{RC}$ that I remember the teacher$_j$ [$_{RC}$ that I gave __$_i$ to __$_j$ last year]].

Recall that sentences such as (17) do not involve long-distance subject extraction, which enabled avoiding a COMP-trace effect in Experiment 2. Hence, ObS would not have a prediction about the acceptability of these sentences. Instead, the low acceptability of (17) could be attributed to the One Dependency at a Time Bias; the parser encounters the direct and indirect

objects prior to the ditransitive verb *gave*. As ditransitive verbs can be followed by either the direct or the indirect object (e.g., *I gave the book to the teacher* or *I gave the teacher the book*), the parser must wait until the following *to* in order to confirm that the direct object follows the verb. This makes the timing of resolving direct and indirect object dependencies simultaneous, leading to a processing cost.

4.4.3 Potential counterexample: Japanese Scrambling

We have argued so far that ObS correctly predicts the contrast in acceptability between subject- and object-extraction double relatives in Japanese. At the same time, ObS does not predict one type of double relative to be easier than the other for English. In the last section, we attributed the uniformly low acceptability of English double relatives to the fact that they always involve having to resolve multiple dependencies at once, as formulated in (16).

Given the word order properties of English, it is impossible to derive double relatives in English in which dependencies are resolved one at a time, and as a result, it should be impossible to find instances of relative clause island violations in English that are judged to be relatively well-formed. That is not the case, however – there have been multiple reported instances of double relatives in English that seem to be relatively well-formed, and what they seem to have in common is that they involve extraction from a relative clause in a semantically non-presupposed domain (Kuno, 1976; McCawley, 1981; Chung & McCloskey, 1983; Vincent et al., 2022).

In addition, Chapter 2 touched upon the fact that ObS makes an incorrect prediction regarding the acceptability of scrambling out of a relative clause in Japanese, which is exemplified below.

(18) ?*Ano hon$_j$-o John-ga [$_{RC}$ __$_i$ __$_j$ kai-ta] hito$_i$-o
that book-ACC John-NOM write-PST person-ACC
sagashi-tei-ru rasii.
look.for-PROG-PRS seem
'That book, John seems to be looking for the person who wrote (it).'

(18) is in line with ObS, as it involves resolving the dependency between the scrambled object and its verb before encountering the subject, on a par with subject-extraction double relatives. Furthermore, dependencies are resolved one at a time (the object and then subject) in (18), which is in line with the One Dependency at a Time Bias. Nevertheless, sentences such as (18) have been shown to exhibit a clear island effect, similar to object-extraction double relatives, in previous studies (Saito, 1985; Fukuda et al., 2022). Therefore, just as we have seen with factors such as the COMP-trace effect and crossing versus nested dependencies, satisfying one factor or even two does not necessarily result in a well-formed outcome, due to the other factors that also play a role. With this background, the next chapter turns to discussing the effect of presuppositionality of extraction domain on the acceptability of double relatives in English, and examines whether this effect might be at work in Japanese, particularly in the case of scrambling out of a relative clause in Japanese, as well.

4.5 Chapter summary

This chapter focused on English, which has been claimed to exhibit a robust relative clause island effect, and investigated various factors that could be contributing to the low acceptability of English double relatives. Such factors include the subject island effect, the COMP-trace effect, and double relatives forming crossing dependencies. In the two experiments carried out in this chapter, Experiment 1 compared double relatives with and without the COMP-

trace effect, where the ones with a COMP-trace effect also form crossing dependencies. While the results revealed the effects on acceptability of a COMP-trace effect, there was no additional negative impact of a crossing dependency on acceptability. Then, in order to compare crossing versus nested dependencies more directly, Experiment 2 featured double relatives derived by relativizing the arguments of a ditransitive verb (direct and indirect objects), avoiding the COMP-trace effect. Again, results showed that the two types of double relatives – containing crossing dependencies versus nested dependencies – had equally low acceptability. We proposed that the reason why the effect of dependency type (crossing versus nested) was absent is because the participants could not project gaps inside of a finite relative clause island with enough confidence. Additionally, as for the reason why the acceptability of English double relatives remained low, even after controlling for factors such as the COMP-trace effect, it could be due to the fact that English double relatives involve resolving two long-distance dependencies at once, unlike the Japanese double relatives where dependencies are resolved one at a time.

Chapter 5

Effects of extraction domain on scrambling out of relative clauses in Japanese

5.1 Introduction

Chapter 4 examined the case of relative clause island violations in English, particularly the violation by further relativization (double relative), in comparison to the same type of island violation in Japanese. Double relatives in the two languages are exemplified below.

(1) a. *The professor$_i$ [$_{RC2}$ who the sci-fi novel$_j$ [$_{RC1}$ that __$_i$ wrote __$_j$] was recently featured in a bookstore] looks proud.

b. [$_{RC2}$ [$_{RC1}$ __$_i$ __$_j$ kai-ta] SF-shoosetsu$_j$-ga saikin shoten-de
write-PST Sci-Fi novel-NOM recently bookstore-at
tokushuu-sa-re-ta] gakusha$_i$-wa hokorashige-da.
feature-do-PASS-PST professor-TOP looks.proud-COP
'The professor$_i$ [$_{RC2}$ who the sci-fi novel$_j$ [$_{RC1}$ that __$_i$ wrote __$_j$] was recently featured in a bookstore] looks proud.'

By simply looking at examples like (1), the penalty for violating the relative clause island may appear to be more severe in English than in Japanese. However, it was proposed in Chapter 4 that the penalty is equally mild in both languages, but there are additional factors negatively impacting the acceptability of sentences such as English (1a). One such factor is the cost of not

following the One Dependency at a Time Bias (ODT); as (1a) involves the verb (*wrote*) appearing after both the subject and the object, the two argument dependencies must be resolved at the same time. This is why the acceptability remained low even when those dependencies were formed by the direct and indirect objects (as in Experiment 2 of Chapter 4), instead of the subject and the object. Given all the factors possibly involved in the low acceptability of (1a), as well as the word order properties of English, it may seem difficult for English to have double relatives that are as acceptable as (1b). And as long as there are only instances of double relatives of low acceptability as in (1a), we cannot rule out the possibility that the penalty of a relative clause island violation can be heavier in one language than another, after all.

In addition to English double relatives, there is another case of relative clause island violation that potentially challenges the proposal that the penalty of the violation is mild. The last chapter also pointed out that both the Object before Subject Bias (ObS; which was proposed to explain the acceptability of double relatives in Japanese) and ODT would predict scrambling out of a relative clause in Japanese to be relatively acceptable, as it involves resolving the object dependency, followed by resolving the subject dependency. previous observations (Saito, 1985; Fukuda et al., 2022), however, agree that it is clearly ill-formed, as exemplified below.

(2) ?*Ano hon$_j$-o John-ga [$_{RC}$ __$_i$ __$_j$ kai-ta] hito$_i$-o
that book-ACC John-NOM write-PST person-ACC
sagashi-tei-ru rasii.
look.for-PROG-PRS seem
'That book, John seems to be looking for the person who wrote (it).'

These observations about English double relatives and scrambling out of a relative clause in Japanese need to be examined more closely before concluding that the penalty of violating a

relative clause island is mild across languages and extraction types. The remainder of the chapter is therefore organized as follows: Section 5.2 introduces previous studies on the presuppositionality of extraction domains, which has been proposed to modulate the acceptability of sentences with a relative clause island violation, including English double relatives. The section then shows why the factor may also be relevant to the acceptability of scrambling out of a relative clause in Japanese. Section 5.3 presents an acceptability judgment experiment featuring Japanese scrambling taking place out of extraction domains varying in their presuppositionality, and reveals that manipulating the presuppositionality of extraction domain affects its acceptability, similar to the effect it has on English. Section 5.4 discusses the findings in further detail.

5.2 Background

5.2.1 Presuppositionality of extraction domain

The previous section suggested that English double relatives can never have a configuration where long-distance dependencies are fully resolved one at a time, unlike in Japanese double relatives. While it should then be impossible to find instances of relatively acceptable English double relatives, such instances have been documented in the literature, including the following:

(3) a. This is the child who there is nobody who is willing to accept.
(Kuno, 1976)

b. This is the one that Bob Wall was the only person who hadn't read.
(McCawley, 1981)

c. That's one trick that I've known a lot of people who've been taken in by.
(Chung & McCloskey, 1983)

The sentences above do not involve simultaneous resolution of subject and object dependencies; in (3a), for instance, the subject dependency is resolved at *is*, and the object dependency at *accept*). In addition to satisfying ODT, relativization in these sentences takes place out of a relative clause that is the pivot of an existential construction (=3a), the predicate nominal (=3b), and the object of the verb *know* (=3c). Vincent et al. (2022) suggest that what these environments have in common is that they are all non-presuppositional; there is a consensus in the literature that extraction out of a presuppositional DP is more difficult than extracting out of a non-presuppositional DP (Erteschik-Shir & Lappin, 1979; Fiengo & Higginbotham, 1981; Diesing, 1992; Bianchi & Chesi, 2014). The pivot of an existential and predicate nominals, as seen in (3a,b), are canonical examples of non-presuppositional DPs. As for (3c), while a transitive object is typically considered a presuppositional DP, Vincent et al. point out that there is a set of "transitive verbs that can be used in an existential way to introduce a referent — and therefore do not presuppose their direct object (p.11)", which they call *evidential existentials*. In one of their experiments, they investigated whether extraction out of a relative clause results in a smaller island effect if the relative clause is the direct object of such verbs. They first identified some of the evidential existential (EE) verbs (e.g., *meet*, *find*, *talk to*) through a norming study, which consisted of question-response pairs as follows:

(4) Question: Is there anyone who can decode this script?

a. Yeah, I'm sure **there's** someone who can decode it. *Existential*

b. Yeah, I **talked to** someone who can decode it. *Evidential existential*

c. Yeah, I **criticized** someone who can decode it. *Ordinary transitive*

A response to questions like the one above with EE verbs (=(4b)) was rated as more natural than the one (=(4c)) with ordinary transitive verbs that cannot be used in an existential (non-presuppositional) way. Such transitive verbs, including *criticize* and *slap*, were abbreviated as VT.

Vincent et al. then conducted an acceptability judgment experiment where stimuli varied according to two factors: Whether relativization is out of a main clause or an embedded clause (LENGTH; short versus long), and the type of verb that takes the embedded clause as its complement (VERB TYPE). The types of verbs were as follows: CP (e.g., *realize*, whose complement was a non-island *that*-clause), EE, and ordinary transitive verbs that cannot be used in an existential (non-presuppositional) way, which was abbreviated as VT (e.g., *slap*). The following shows a sample set of stimuli in Vincent et al.'s Experiment 2:

(5) Condition 1 LENGTH: short; VERB TYPE: CP
This is the woman$_i$ that __$_i$ just realized [$_{that}$ that an expert could prove this claim].

Condition 2 LENGTH: long; VERB TYPE: CP
This is a claim$_i$ that I just realized [$_{that}$ that an expert could prove __$_i$].

Condition 3 LENGTH: short; VERB TYPE: EE
This is the woman$_i$ that __$_i$ just **found** [$_{RC}$ an expert who could prove this claim].

Condition 4 LENGTH: long; VERB TYPE: EE
This is a claim$_i$ that I just **found** [$_{RC}$ an expert who could prove __$_i$].

Condition 5 LENGTH: short; VERB TYPE: VT
This is the woman$_i$ that __$_i$ just **slapped** [$_{RC}$ an expert who could prove this claim].

Condition 6 LENGTH: long; VERB TYPE: VT
This is a claim$_i$ that I just **slapped** [$_{RC}$ an expert who could prove __$_i$].

Results of their experiment are shown in Figure 5.1 below.

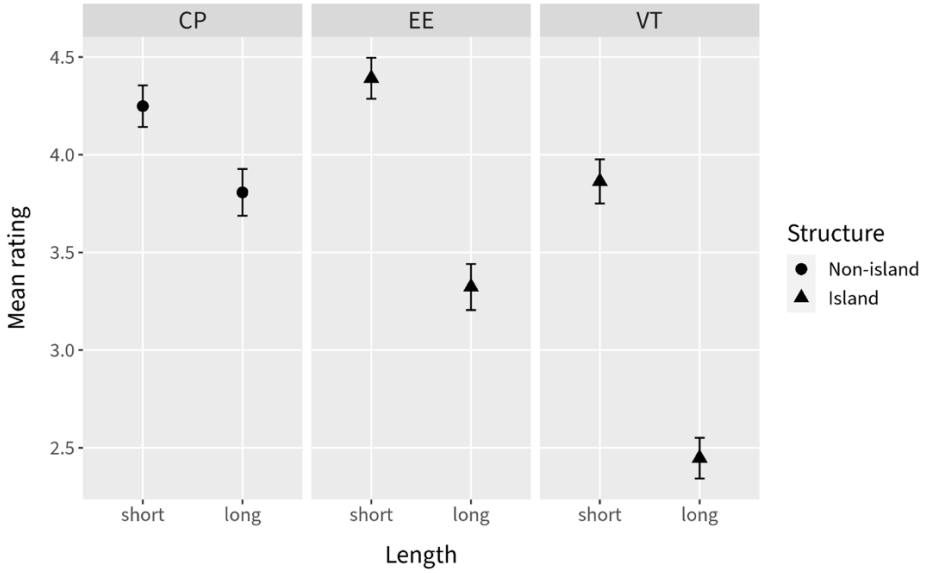

Figure 5.1: Aggregated responses to critical items in Experiment 2 by Vincent et al. (2022)

Vincent et al. found significant interactions between LENGTH and VERB TYPE in both CP vs. EE and CP vs. VT comparisons, suggestive of a relative clause island effect for both verb types. Critically, they also showed that the size of island effect was smaller when the verb type was EE than when it was VT (see Section 5.3.4 for the statistical method they used to find this out). They thus concluded that relative clauses that are the direct object of a verb that can be used in a non-presuppositional way (EE) are more transparent for extraction than the relative clauses with a typical transitive verb (VT).

If the presuppositionality of extraction domain impacts the acceptability of English double relatives, could it have also played a role in the experiments presented in Chapter 4? Among the stimuli created for the two acceptability experiments in the last chapter, the relative clause from which further relativization took place was always the direct object of one of four verbs (*remember*, *know*, *believe*, or *recognize*). Vincent et al. did test the presuppositionality of *know* (which is also featured in (3c)), and they classified it as an EE verb, which should facilitate

an extraction out of its direct object. For *remember*, *believe*, and *recognize*, however, informal judgments from a native English speaker rated them to be less natural as a part of the response to the question in (4), compared with *know*. In order to see whether the presuppositionality of these verbs affected the acceptability of double relatives, I divided a subset of critical items (the ones with double relatives) of Experiment 1 in Chapter 4 into the ones with *know* and with one of the other three verbs. Sentences in each of the groups is exemplified below. To review, a sentence triggers the COMP-trace effect (and crossing dependency) if it has the plural copula *are* in the inner relative clause (RC1) associated with the matrix object.

(6) a. I waved at the children$_j$ [$_{RC2}$ who I *know* the teacher$_i$ [$_{RC1}$ who __$_i$ is/are running with __$_j$]].
b. I messaged the teenagers$_j$ [$_{RC2}$ who I *recognize* the boy$_i$ [$_{RC1}$ who __$_i$ is/are playing with __$_j$]].

I then performed Welch's *t*-tests to compare the mean z-score acceptability of the two groups of sentences (the assumption of equal variances was not assumed due to unequal sample sizes). For sentences without the COMP-trace effect (i.e., (6a,b) with *is*), the mean acceptability of the *know* group (-0.35) was significantly higher than that of the others (-0.49) ($t(44.58) = -2.75$, $p < 0.01$). Likewise, for sentences with the COMP-trace effect (i.e., (6a,b) with *are*), the mean acceptability of the *know* group (-0.52) was also significantly higher than that of the others (-0.75) ($t(48.12) = -2.09$, $p < 0.05$). Keep in mind that different acceptability could be due to a difference in the other lexical items, and thus it may have nothing to do with the choice of verbs. Nevertheless, the fact that there seems to be a lexical effect due to *know* is intriguing, all the more so because the head of the relative clause island is definite (e.g., *the teacher* in (6a)), which is known to be presuppositional (Strawson, 1950).

To sum up, Vincent et al.'s experiment and the ones from the last chapter show some evidence that the acceptability of English double relatives can be improved, particularly by making the inner relative clause the object of a non-presuppositional verb. Let us now turn to the discussion of scrambling out of relative clause in Japanese, and whether its acceptability could also be influenced by the presuppositionality of the extraction domain.

5.2.2 Scrambling out of relative clauses in Japanese

Scrambling is an optional change in word order available to languages with a robust case marking system like Japanese. The most common instance of scrambling involves moving the object NP to the sentence-initial position, deriving OSV word order instead of the canonical SOV order.[1] Unlike relativization in Japanese, scrambling is a leftward movement and it thus leads to filler-gap dependency, the same as relativization in English. Scrambling can be long-distance, where an NP inside of an embedded clause is moved to a position in the main clause. As long-distance scrambling is considered A'-movement, scrambling out of a relative clause is known to trigger an island effect, exemplified by the following example (repeated from (2)):

[1] Scrambling a nominative-marked NP (i.e., subject) is known to be disallowed in Japanese (Saito, 1985; Nemoto, 1993), as the following example involving scrambling out of a *to*-clause (not an island) shows:

(i)	*Sono hon$_i$-ga	John-ga	[t_i	yoku	ure-tei-ru	to]	omot-tei-ru.
	that book-NOM	John-NOM		well	sell-PROG-PRS	that	think-PROG-PRS

According to Saito (1985), the ban on subject scrambling is due to different case assignment mechanisms between subject and object; in Japanese, the trace of the scrambled object receives (abstract) Case from the verb, satisfying the condition that variables must have Case (Chomsky, 1981). The trace of subject scrambling, on the other hand, cannot receive any abstract Case, violating the condition. So in theory, subject scrambling of any length should be disallowed in Japanese, including the case of "quantifier float," in which a numeral quantifier and its associated subject NP are separated by scrambling the former, as follows:

(ii)	*Gakusei-ga$_i$ sake-o	t_i	sannin	nonde-i-ru.
	students-NOM	sake-ACC	three.CL	drink-PROG-PRS

'Three students are drinking sake.'

See Saito's (1985) Chapter 3.2 for further discussion.

(7) ?*Ano hon$_i$-o John-ga [$_{RC}$ ___$_i$ ___$_j$ kai-ta] hito-o
that book-ACC John-NOM write-PST person-ACC
sagashi-tei-ru rasii.
look.for-PROG-PRS seem
'That book, John seems to be looking for the person who wrote (it).'

Chapter 2 introduced an acceptability experiment by Fukuda et al. (2022), where sentences such as (7) were evaluated in a factorial-design experiment. They compared long-distance scrambling out of a non-island clause headed by *to* 'that,' as in (8b), to long-distance scrambling out of a relative clause, as in (8d), in addition to the baseline without scrambling ((8a) with a non-island and (8c) with a relative clause).

(8) a. *non-island / no scrambling*
Roodookumiai-no riidaa-wa [$_{CP}$ kaisha-no juuyaku$_i$-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 sasete-i-nai-to] hihan-shi-ta.
reflect make-PROG-NEG-COMP criticize-do-PST
'The union leader criticized that the company's executives were not making the drastic sales growth reflected in the employees' salaries.'

b. *non-island / scrambling*
Oohabana uriage-no nobi-o$_j$ roodookumiai-no riidaa-wa
drastic sales-GEN growth-ACC union-GEN leader-TOP
[$_{CP}$ kaisha-no juuyaku$_i$-ga ___$_j$ juugyooin-no kyuuyo-ni
company-GEN executives-NOM employee-GEN salary-to
han'ee sasete-i-nai-to] hihan-shi-ta.
reflect make-PROG-NEG-COMP criticize-do-PST
'The drastic sales growth, the union leader criticized that the company's executives were not making (them) reflected in the employees' salaries.'

c. *island / no scrambling*
Roodookumiai-no riidaa-wa [$_{RC}$ ___$_i$ oohabana uriage-no
union-GEN leader-TOP drastic sales-GEN
nobi-o juugyooin-no kyuuyo-ni han'ee sasete-i-nai]
growth-ACC employee-GEN salary-to reflect make-PROG-NEG
kaisha-no juuyaku$_i$-o hihan-shi-ta.
company-GEN executives-ACC criticize-do-PST

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

d. *island / scrambling*

Oohabana	uriage-no	nobi-o$_j$	roodookumiai-no	riidaa-wa
drastic	sales-GEN	growth-ACC	union-GEN	leader-TOP
[$_{RC}$ ___$_i$	___$_j$ juugyooin-no	kyuuyo-ni	han'ee sasete-i-nai]	
	employee-GEN	salary-to	reflect make-PROG-NEG	
kaisha-no	juuyaku$_i$-o	hihan-shi-ta.		
company-GEN	executives-ACC	criticize-do-PST		

'The drastic sales growth, the union leader criticized the company's executives who were not making (them) reflected in the employees' salaries.'

Fukuda et al. observed a superadditive drop in acceptability in the island / scrambling condition (=8d), indicating that scrambling out of a relative clause leads to an island effect.

Importantly, unlike the other types of dependency across a relative clause (some relativization or a *wh*-dependency, the latter of which does not involve overt movement), scrambling out of a relative clause is generally judged to be ill-formed without exception. As mentioned in Section 5.1, this judgment goes against the predictions of ObS and ODT; in sentences like (8d), the scrambled object and the verb are fully resolved first, before encountering the subject in the relative clause head position. All other things being equal, this sentence should thus be judged just as good as subject-extraction double relatives like (1b), contrary to the previous observations.

Given that ObS and ODT do not capture the acceptability of scrambling out of a relative clause, let us consider the possibility that while these principles apply regardless of dependency type, there is an additional factor at play in (8d) that does not affect (1b), or at least not to the same extent. In particular, given the effect of the presuppositionality of extraction domain in English double relatives seen in the last section, one may wonder whether it is the scrambling out of a presuppositional extraction domain that amplifies the relative clause island effect. As far as I

am aware, none of the studies on Japanese scrambling took this factor into consideration; in Fukuda et al.'s (2022) experiments, for instance, all of their stimuli seem to have involved a relative clause as the object of a typical transitive verb (e.g., *hihansuru* 'to criticize'), which is presumably a presuppositional position. If the presuppositionality of extraction domain is relevant to scrambling in Japanese, scrambling out of a relative clause should be facilitated when the relative clause is in a less presupposed domain, similar to relativization out of a relative clause in Japanese when it observes ObS. The acceptability experiment presented in the next section puts this prediction to the test.

5.3 Experiment

5.3.1 Participants

50 participants, who were self-identified native Japanese speakers, were recruited on CrowdWorks. Two of them were excluded for their performance in the selected filler items, on par with the previous experiments on Japanese presented in Chapter 2 (details provided below), leaving 48 participants (age range = 25-58, mean = 41.1) whose data were analyzed. All participants reported that Japanese was their first language, and their parents primarily used Japanese to communicate with them. Participants received approximately $3 (in Japanese yen) for participation.

5.3.2 Materials

Stimuli in this experiment vary according to the following two factors: Whether there is extraction (scrambling) out of an embedded clause (EXTRACTION), and the type of embedded clause from which the extraction takes place (ENVIRONMENT). There are three levels for the

ENVIRONMENT factor: The baseline involves a declarative CP headed by the complementizer *-to* 'that' (*to*-clause), which behaves as a non-island. A non-presuppositional RC involves a relative clause inside the *to*-clause, and the relative clause is the pivot of an existential construction (i.e., X in *there is X*), which is a canonical example of a non-presuppositional domain. A presuppositional RC also involves a relative clause inside the *to*-clause, but the relative clause is in the direct object position of the transitive verb *see*, which is presuppositional.

The +EXTRACTION conditions all involve long-distance scrambling. In order to maximize acceptability of those conditions, the stimuli followed the processing preference discussed in Fukuda et al. (2022), which is named the *long-before-short* preference (Dryer, 1980; Hawkins, 1994; Yamashita & Chang, 2001). As the name suggests, Japanese parsers are known to prefer processing longer elements before shorter ones in a sentence, and this preference has influenced whether a superadditive effect on acceptability emerges in cases such as scrambling out of the subject (Jurka, 2010; Jurka et al., 2011; Omaki et al. 2020) and noun complements (Yano, 2019; Fukuda et al., 2022). In order to satisfy this preference, +EXTRACTION stimuli were created such that the scrambled noun phrase was longer than any other noun phrases that it was scrambled over.

A sample set of stimuli is provided below. The noun phrase that undergoes scrambling is grayed. Given the nature of scrambling as an optional operation with no major impact on the propositional meaning (unlike how the meaning changes drastically from *the dog chased the cat* to *the cat chased the dog*, for instance), the English translation is not repeated for +EXTRACTION conditions.

(9) Condition 1 ENVIRONMENT: Baseline; EXTRACTION: -

Zemi-no	seito-wa	[*to*	wakate	hyouronka-ga	amari
seminar-GEN	student-TOP		junior	critic-NOM	not.well
shira-re-tei-nai	SF-shoosetsu-o		asa-no	jyouhoubangumi-de	
know-PASS-PRS-NEG	Sci-Fi novel-ACC		morning-GEN	talk.show-at	
zessan-shi-ta-to]	it-ta.				
praise-do-PST-that	say-PST				

'A student from the seminar said [*to* that a junior critic praised the Sci-Fi novel that was not well-known on a morning talk show].'

Condition 2 ENVIRONMENT: Baseline; EXTRACTION: +

Amari	shira-re-tei-nai	SF-shoosetsu-o$_j$	zemi-no	seito-wa
not.well	know-PASS-PRS-NEG	Sci-Fi novel-ACC	seminar-GEN	student-TOP
[*to* wakate	hyouronka-ga __$_j$	asa-no	jyouhoubangumi-de	
junior	critic-NOM	morning-GEN	talk.show-at	
zessan-shi-ta-to]	it-ta.			
praise-do-PST-that	say-PST			

(same as Condition 1)

Condition 3 ENVIRONMENT: Non-presuppositional RC; EXTRACTION: -

Zemi-no	seito-wa	[*to* [RC	amari	shira-re-tei-nai
seminar-GEN	student-TOP		not.well	know-PASS-PRS-NEG
SF-shoosetsu-o	asa-no	jyouhoubangumi-de	zessan-shi-ta]	
Sci-Fi novel-ACC	morning-GEN	talk.show-at	praise-do-PST	
wakate	hyouronka-ga	iru-to]	it-ta.	
junior	critic-NOM	there.is-that	say-PST	

'A student from the seminar said [*to* that there is a junior critic [RC who praised the Sci-Fi novel that was not well-known on a morning talk show]].'

Condition 4 ENVIRONMENT: Non-presuppositional RC; EXTRACTION: +

Amari	shira-re-tei-nai	SF-shoosetsu-o$_j$	zemi-no	seito-wa
not.well	know-PASS-PRS-NEG	Sci-Fi novel-ACC	seminar-GEN	student-TOP
[*to* [RC __$_i$	__$_j$	asa-no	jyouhoubangumi-de	zessan-shi-ta]
		morning-GEN	talk.show-at	praise-do-PST
wakate	hyouronka$_i$-ga	iru-to]	it-ta.	
junior	critic-NOM	there.is-that	say-PST	

(same as Condition 3)

Condition 5 ENVIRONMENT: Presuppositional RC; EXTRACTION: -

Zemi-no	seito-wa	[*to* [RC	amari	shira-re-tei-nai
seminar-GEN	student-TOP		not.well	know-PASS-PRS-NEG
SF-shoosetsu-o	asa-no	jyouhoubangumi-de	zessan-shi-ta]	
Sci-Fi novel-ACC	morning-GEN	talk.show-at	praise-do-PST	
wakate	hyouronka-o	mi-ta-to]	it-ta.	
junior	critic-ACC	see-PST-that	say-PST	

'A student from the seminar said [*to* that they saw a junior critic [RC who praised the Sci-Fi novel that was not well-known on a morning talk show]].'

Condition 6 ENVIRONMENT: Presuppositional RC; EXTRACTION: +

Amari	shira-re-tei-nai	SF-shoosetsu-o$_j$	zemi-no	seito-wa
not.well	know-PASS-PRS-NEG	Sci-Fi novel-ACC	seminar-GEN	student-TOP
[*to* [RC __$_i$	__$_j$	asa-no	jyouhoubangumi-de	zessan-shi-ta]
		morning-GEN	talk.show-at	praise-do-PST
wakate	hyouronka$_i$-o	mi-ta-to]	it-ta.	
junior	critic-ACC	see-PST-that	say-PST	

(same as Condition 5)

24 lexically-matched sets as in (9) were created, and they were divided into 6 lists using a Latin square design (with 4 items from each combination). I also modified the set of 40 fillers from the previous experiments to make them longer in order to match the length of the critical items, without changing their acceptability. Fillers consisted of 10 sentences of expected high acceptability, 10 of intermediate acceptability, and 20 of low acceptability. The identical set of fillers was mixed with the critical items across lists. The resulting list of stimuli consisted of 24 critical items + 40 fillers = 64 in total. The full set of stimuli for this experiment is available at the following Open Science Framework page: https://osf.io/9fdsy/.

5.3.3 Procedure

This experiment was hosted on PCIbex (Zehr & Schwarz, 2018). Participants were instructed to rate how natural each sentence sounded by clicking on a number on a scale from 1 (*goku fushizen* 'very unnatural') to 7 (*goku shizen* 'very natural'). Participants also completed a brief language background questionnaire after the rating task.

To screen out participants who were not attending to the task, responses to the 10 filler items with the highest mean acceptability scores across all participants and the 10 with the

lowest scores were identified. Participants whose ratings were more than 2 standard deviations away from the mean for 5 or more of these 20 items were excluded from further analysis. Two participants were filtered out in this way for this experiment, and after recruiting additional participants to make up for the lost data, I obtained the final dataset of 48 participants (with 8 participants in each of the 6 lists).

5.3.4 Data analysis

As in the previous experiments, linear mixed-effects regression models were created with the sentence acceptability (in z-scores) as the dependent variable. Similar to what Vincent et al. (2022) did for their analysis, I created one model based on data that only included Baseline and Non-presuppositional RC (Condition 1 through 4), and another model based on data that only included Baseline and Presuppositional RC (Condition 1, 2, 5, 6). Each of the models predicted acceptability based on the type of extraction environment (ENVIRONMENT), whether there is an extraction out of an embedded clause (EXTRACTION), and their interaction. The models that converged included random intercepts of participant and item, random slopes of ENVIRONMENT for participant, and random slopes of EXTRACTION for item.

To address the question of whether the presuppositionality of extraction environment affects the acceptability of scrambling out of a relative clause, two approaches were taken: First, another model based on data that excluded Baseline (Condition 3 through 6) was made, which predicted acceptability with ENVIRONMENT, EXTRACTION, and their interaction. Second, I created a model with data from all 6 conditions, but applied the so-called *contrast coding* to the ENVIRONMENT variable. When a categorical variable has multiple levels as in this experiment (Baseline | Non-presuppositional RC | Presuppositional RC), a regression model defines one of

the levels (Baseline, in this case) as the reference level or the intercept, and computes coefficients from comparing that reference level and each of the other levels (Baseline vs. Non-presuppositional RC, Baseline vs. Presuppositional RC). This, however, prevents us from comparing the two non-reference levels. Contrast coding such as Helmert coding solves this problem, because it lets the model compare a level against the mean of all the subsequent levels. For example, suppose a variable has 4 levels; A, B, C, and D. With Helmert coding, the model compares A vs. *mean*(B+C+D), B vs. *mean*(C+D), and C vs. D. In the case of the ENVIRONMENT variable, the model would compare Baseline vs. *mean*(Non-presuppositional RC+Presuppositional RC), and Non-presuppositional RC vs. Presuppositional RC, the latter of which informs us whether the two levels (and their interaction with EXTRACTION) have significantly different impacts on acceptability.[2]

5.3.5 Predictions

Considering the previous observations and experimental findings, scrambling out of a relative clause should lead to a superadditive effect on acceptability (manifested by a significant interaction effect between ENVIRONMENT and EXTRACTION), regardless of the presuppositionality of extraction domain.

At the same time, scrambling out of a relative clause is predicted to be facilitated when it takes place out of a non-presuppositional environment. This prediction could be borne out in two ways: A non-presuppositional environment can improve the acceptability, but it may not make

[2] Helmert coding was applied manually instead of R's built-in `contr.helmert` function, which would result in reverse Helmert coding (e.g., with the variable with 4 levels A, B, C, and D, the comparisons would start with D vs. *mean*(A+B+C)). I created a 2x3 matrix with the values specified below, and assigned the matrix as the contrast matrix of the ENVIRONMENT variable.

```
my_helmert_env = matrix(c(2, -1, -1, 0, 1, -1), ncol = 2)
contrasts(data$Extraction_env) = my_helmert_env
```

the relative clause island effect smaller. Under this scenario, the Baseline-less model and the full model with Helmert coding (in the Non-presuppositional RC vs. Presuppositional RC comparison) would show a significant main effect of ENVIRONMENT but no interaction effect between ENVIRONMENT and EXTRACTION. Alternatively, a non-presuppositional environment can both improve the acceptability and make the island effect smaller. Such a scenario (the same as what Vincent et al. 2022 observed) should be manifested as a significant main effect of ENVIRONMENT and an interaction between ENVIRONMENT and EXTRACTION by the two models.

5.3.6 Results

Mean z-scores for the six conditions of the experiment are presented in Figure 5.2.

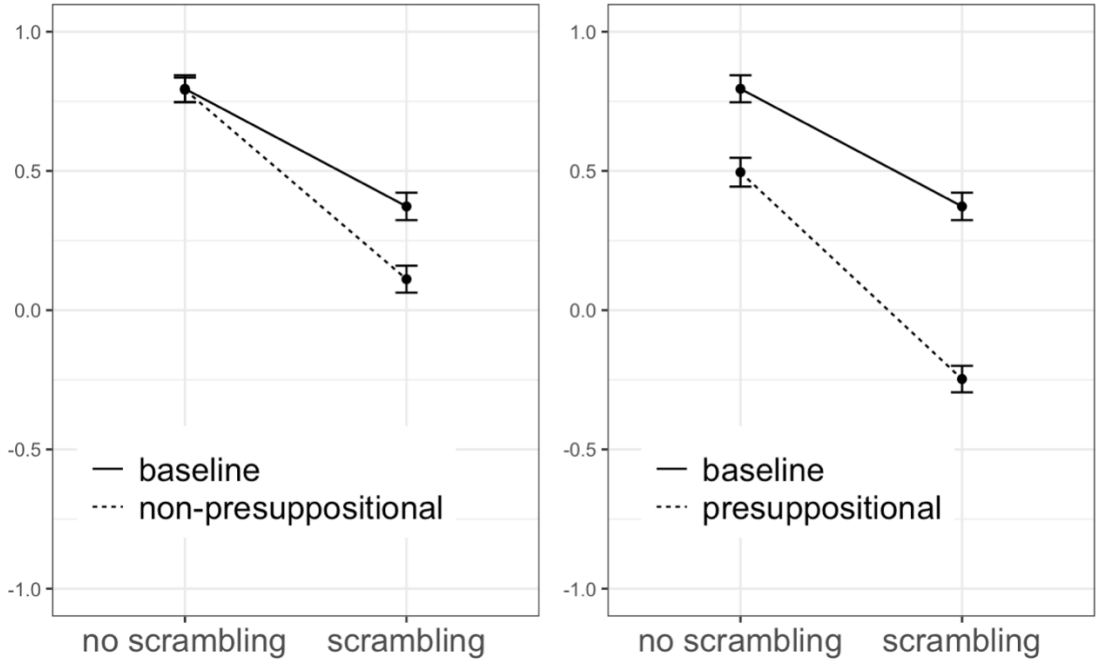

Figure 5.2: Aggregated responses to critical items in the present experiment (scores of Baseline conditions are repeated in the two plots)

The model including Baseline and Non-presuppositional RC conditions (left panel of Figure 5.2) revealed a significant main effect of EXTRACTION in the expected direction;

+EXTRACTION sentences were rated lower than their -EXTRACTION counterparts (β = -0.42, SE = 0.08, $p < 0.001$). There was no main effect of ENVIRONMENT (β = -0.01, SE = 0.07, $p = 0.96$). Importantly, the interaction effect between EXTRACTION and ENVIRONMENT was significant (β = -0.26, SE = 0.09, $p < 0.01$), indicative of an island effect associated with scrambling out of a relative clause in a non-presuppositional environment. The other model including Baseline and Presuppositional RC conditions (right panel of Figure 5.2) revealed a significant main effect of EXTRACTION (β = -0.42, SE = 0.08, $p < 0.001$), as well as a significant main effect of ENVIRONMENT (β = -0.30, SE = 0.07, $p < 0.001$). Similar to the first model, this model also revealed a significant interaction effect between EXTRACTION and ENVIRONMENT (β = -0.32, SE = 0.09, $p < 0.001$), suggestive of an island effect associated with scrambling out of a relative clause in a presuppositional environment.

Turning to the effect of extraction domain on the acceptability of scrambling out of a relative clause, the Baseline-less model revealed a significant main effect of EXTRACTION (β = -0.68, SE = 0.10, $p < 0.001$) as well as ENVIRONMENT (β = -0.30, SE = 0.08, $p < 0.001$), the latter of which indicates that Non-presuppositional RC sentences were scored higher in acceptability than Presuppositional RC sentences (regardless of the presence of scrambling). The model failed to find an interaction between EXTRACTION and ENVIRONMENT (β = -0.06, SE = 0.10, $p = 0.55$), suggesting that the size of the relative clause island effect between the two extraction domains is not significantly different. An equivalent pattern of results was observed for the full model with Helmert coding; comparing Non-presuppositional RC and Presuppositional RC levels, the model revealed a significant main effect of ENVIRONMENT (β = 0.16, SE = 0.02, $p < 0.001$), but failed to do so for the interaction between EXTRACTION and ENVIRONMENT (β = 0.03, SE = 0.04, $p = 0.47$).

5.3.7 Discussion

The present experiment demonstrated that scrambling out of a relative clause leads to an island effect, in accord with the previous observations, whether the relative clause is in a semantically non-presuppositional or presuppositional domain. More importantly, the experiment aimed to investigate whether scrambling can be facilitated by the semantically non-presuppositional extraction domain. Results suggest that such a domain improves the acceptability of sentences (with and without scrambling) to the point that scrambling out of a relative clause ends up with a positive z-score, indicative of intermediate acceptability, which was also observed for subject-extraction double relatives in Chapter 2. At the same time, however, the non-presuppositional extraction domain does not seem to make the relative clause island effect smaller.

To facilitate the comparison between the current experiment and Vincent et al.'s, Figure 5.3 below shows Vincent et al.'s results, and the results from Figure 5.2 replotted to make them look more similar to theirs (with z-scores converted to raw acceptability).

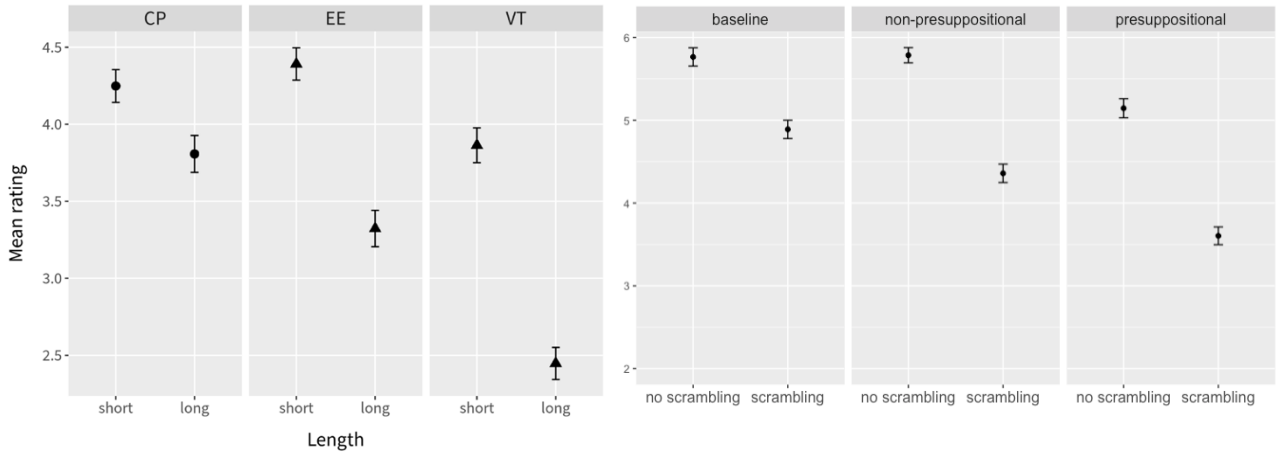

Figure 5.3: Results of Vincent et al.'s (2022) Experiment 2 (left) and the present experiment (right)

One major difference in the pattern of results between the two experiments is that the condition with an extraction out of the presuppositional relative clause has a steeper drop in acceptability in Vincent et al.'s experiment ("VT" column in the left figure) than in my experiment ("presuppositional" column in the right figure). This contrast matches what we have observed, namely the absence of an interaction between EXTRACTION and ENVIRONMENT in both the Baseline-less model and the full model with Helmert coding.

5.4 General discussion

This chapter has examined the existence and size of the island effect triggered by scrambling, as opposed to relativization, out of a relative clause in Japanese, and whether manipulating the presuppositionality of the extraction domain would enable us to observe relatively well-formed instances of scrambling out of a relative clause. Scrambling was featured in this experiment because it has been observed to be clearly ill-formed, contrary to the prediction made by ObS and ODT that it should be as acceptable as subject-extraction double relatives. And considering the effect of the presuppositionality of its extraction domain on English double relatives demonstrated by Vincent et al. (2022), it was hypothesized that making the extraction domain non-presuppositional may facilitate the scrambling.

The acceptability experiment presented in this chapter manipulated whether the relative clause is in a position that is semantically non-presuppositional (pivot of the existential construction) or presuppositional (object of transitive verb *to see*), with a prediction that extraction is facilitated from the non-presuppositional domain. The following table summarizes the extraction domains of the present experiment and one of the experiments conducted by Vincent et al. (2022), the former of which was modeled after the latter:

Table 5.1: Extraction environments of the present experiment and Vincent et al.'s (2022) experiment

	Japanese scrambling out of relative clause	English double relative (Vincent et al., 2022)
Baseline	*to*-clause	*that*-clause
Non-presuppositional RC	pivot of existential construction	direct object of evidential existential verbs (e.g., *find*)
Presuppositional RC	direct object of transitive verb *miru* 'see'	direct object of ordinary transitive verbs (e.g., *slap*)

A similar pattern of results was observed in the two experiments, as illustrated in Figure 5.3. In particular, having a non-presuppositional relative clause seems to improve acceptability compared with a presuppositional relative clause in both cases, even when there is no extraction out of it. At the same time, for English double relatives, a non-presuppositional relative clause not only improves the acceptability of conditions with such a relative clause, but also reduces the size of the relative clause island effect triggered by extraction. For scrambling out of a relative clause, on the other hand, a non-presuppositional relative clause seems to have only the former effect.

Two questions arise with regard to the results just presented: Firstly, why was there a difference in acceptability between non-presuppositional and presuppositional relative clauses even when there was no extraction out of them? Secondly, why does the presuppositionality of extraction domain have differing effects in the two instances of relative clause island violation? The subsequent sections will discuss these questions, as well as how the presuppositionality of the extraction domain may impact the size of the relative clause island effect across different types of extraction.

5.4.1 Low acceptability of presuppositional relative clauses

Regarding the first question about why the sentences with a presuppositional relative clause had lower acceptability than the ones with a non-presuppositional relative clause, even the ones without scrambling out of it, one possibility is that participants perceived a sentence with a presuppositional relative clause as unnatural in the absence of any appropriate context. In my experiment, each sentence was presented without any context. Thus, when a sentence included a presuppositional relative clause, participants may have found it odd as the referent denoted by the DP containing the relative clause (e.g., *wakate hyouronka* 'junior critic' in (9)) had not been established in the discourse. Also recall Lin and Bever (2010)'s claim from Chapter 2 that not only is the content of relative clauses typically something that is already familiar to the hearer, but it is also supposed to aid the parser in identifying the referent (i.e., head noun of a relative clause) among all the referents previously mentioned. Experiment 3a of Chapter 2 showed that the acceptability of a double relative improves when it is preceded by context that introduces multiple referents of the same entities, which builds an expectation that they are to be disambiguated by relative clauses (*contrastive context*), compared with when the context does not do so. Therefore, with an appropriate context, the acceptability of conditions with a presuppositional relative clause in the current experiment could have been different.

In Vincent et al.'s experiment, on the other hand, each sentence was accompanied by a "preceding question," as exemplified below.

(10) a. *Preceding question*: Is there anyone who could prove this claim?
b. This is the woman$_i$ that __$_i$ just **found** [$_{RC}$ an expert who could prove this claim].
c. This is the woman$_i$ that __$_i$ just **slapped** [$_{RC}$ an expert who could prove this claim].

Preceding questions like (10a) were included in order to support the existential interpretation of the relative clause in the sentence. While such an existential interpretation is compatible with the evidential existential (EE) verbs such as *find* in (10b), ordinary transitive (VT) verbs like *slap* were found to be incompatible with the interpretation. Put differently, the contexts that Vincent et al. provided were appropriate for sentences with EE verbs (=(10b)), but inappropriate for sentences with VT verbs (=(10c)).

Thus, to answer the first question, sentences with a presuppositional relative clause (even those without extraction) may have reduced acceptability either because there was no context (as in my experiment), or because they were preceded by an incompatible context (as in Vincent et al.'s experiment).[3] And the fact that participants saw such sentences in an experimental setting may have made them particularly sensitive to the existence and type of context, in a way that they would not be in ordinary conversations.

5.4.2 Effect of presuppositionality on the size of relative clause island effect

As for the second question, a non-presuppositional extraction domain seems to make a relative clause more transparent for extraction in the case of English double relatives. For Japanese scrambling out of a relative clause, in contrast, while a non-presuppositional relative clause did improve the acceptability of sentences with (and without) scrambling, we did not find the evidence that it also makes the relative clause island effect smaller; in other words, there was no interaction effect between the type of extraction domain and scrambling. Hence, on the

[3] Another possibility, which applies to the stimuli of my experiment but not Vincent et al.'s, is that the presuppositional conditions may have been rated lower than the non-presuppositional conditions because the former involves an additional argument, which is the subject of *miru* 'to see'. It is occupied by a null pronoun *pro*, and it refers to *zemi-no seito* 'a student of the seminar,' which is the subject of the main clause verb *iu* 'to say.'

surface, presuppositionality of extraction domain appears to have differing effects on English double relatives and Japanese scrambling out of a relative clause.

There is, however, a reason to believe that the presuppositionality of extraction domain does affect the size of the island effect in Japanese scrambling, on a par with English double relatives, but such an effect may have been masked in my experiment due to the choice of lexical items for the presuppositional relative clause conditions (Condition 5 and 6 in (9)). In particular, Vincent et al. used a variety of transitive verbs (*slap*, *describe*, *criticize*, *imitate*) whose direct object is the relative clause island, whereas my experiment used a single verb *miru* 'to see'. Using this particular verb could have led to the following consequences: First, as participants saw 4 instances of extraction out of the relative clause predicated by the same transitive verb, there could have been a slight adaptation effect. Second, because *to see* is a more common verb than ones like *slap*, seeing the former in a sentence was perhaps less jarring than seeing the latter, despite the lack of context. Third, *to see* may have an EE property to some extent given that "Yeah, I saw someone who can decode it." is not very strange of an answer to the question "Is there anyone who can decode this script?", which was the test (repeated from (4)) to determine whether a verb is EE or VT by Vincent et al. And lastly, Kush et al. (2013) pointed out that in English, the effect of an extraction out of a relative clause is ameliorated when the relative clause is the object of verbs like *to see*, as the following exemplifies:

(11) That was the bill$_j$ that [$_{RC}$ he saw many senators [$_{RC}$ who supported __$_j$ in the congress]].

They attributed this amelioration effect to grammatical illusion; because verbs like *to see* can take a small clause as its complement (e.g., *June saw Mary eat cheese*), sentences such as

(11) are temporarily parsed as an instance of extraction out of a small clause, which is not an island. Although *miru* 'to see' in Japanese does not take small clause complements (or clausal complements for that matter),[4] we may obtain a different pattern of results if we used a verb that does not take a small clause complement in other languages.

A future study on scrambling out of a relative clause should, therefore, create stimuli with a wider range of transitive verbs taking the relative clause island as their direct object. If the presuppositionality of extraction domain does alter the size of relative clause island effect, parallel to English double relatives, the acceptability of scrambling out of a presuppositional relative clause (Condition 6 in (9)) should be lower than what we observed in the present experiment, resembling Vincent et al.'s pattern of results more closely. And with the lower acceptability, we may obtain a significant difference in the size of island effect between non-presuppositional and presuppositional relative clauses.

5.4.3 Relative clause island effect across extraction types

This chapter started with the observation that scrambling out of a relative clause has been judged ill-formed, while some instances of relativization out of another relative clause (double relative) have been judged relatively acceptable. Previous chapters have argued that Japanese double relatives can be relatively acceptable if they follow ObS and ODT. The current chapter, however, has shown that satisfying those does not always lead to a small island effect, due to additional factors such as the presuppositionality of extraction domain. Would this factor play a

[4] A canonical Japanese small clause involves an accusative-marked subject and an adjective in the stem form, predicated by cognitive verbs such as *omou* 'to think' (e.g., Takahashi, 2002).

(i) John-wa [SC Mary-o kawaiku] omot-ta.
John-TOP Mary-ACC cute think-PST
'John thought that Mary was cute. / John considered Mary cute.'

role in double relatives as well? To facilitate the comparison between two types of extraction out of a relative clause, the following show scrambling out of a relative clause repeated from (9), and double relatives with the same set of lexical items as in (9):

(12) a. Scrambling out of a non-presuppositional RC

Amari shira-re-tei-nai SF-shoosetsu-o$_j$ zemi-no
not.well know-PASS-PRS-NEG Sci-Fi novel-ACC seminar-GEN
seito-wa [*to* [RC __$_i$ __$_j$ asa-no jyouhoubangumi-de
student-TOP morning-GEN talk.show-at
zessan-shi-ta] wakate hyouronka$_i$-ga iru-to] it-ta.
praise-do-PST junior critic-NOM there.is-that say-PST
'A student from the seminar said [*to* that there is a junior critic [RC who praised the Sci-Fi novel that was not well-known at a morning talk show]].'

b. Scrambling out of a presuppositional RC

Amari shira-re-tei-nai SF-shoosetsu-o$_j$ zemi-no
not.well know-PASS-PRES-NEG Sci-Fi novel-ACC seminar-GEN
seito-wa [*to* [RC __$_i$ __$_j$ asa-no jyouhoubangumi-de
student-TOP morning-GEN talk.show-at
zessan-shi-ta] wakate hyouronka$_i$-ga iru-to] it-ta.
praise-do-PST junior critic-NOM there.is-that say-PST
'A student from the seminar said [*to* that they saw a junior critic [RC who praised the Sci-Fi novel that was not well-known at a morning talk show]].'

(13) a. Relativization out of a non-presuppositional RC

Kore-wa [RC [RC __$_i$ __$_j$ asa-no jyouhoubangumi-de
this-TOP morning-GEN talk.show-at
zessan-shi-ta] wakate hyouronka$_i$-ga iru-to zemi-no
praise-do-PST junior critic-NOM there.is-that seminar-GEN
seito-ga it-ta] amari shira-re-tei-nai SF-shoosetsu$_j$-da.
student-NOM say-PST not.well know-PASS-PRES-NEG Sci-Fi novel-COP
'This is the not-well-known Sci-Fi novel [RC that a student from the seminar said that there is a junior critic [RC who praised at a morning talk show]].'

b. Relativization out of a non-presuppositional RC

Kore-wa [RC [RC __$_i$ __$_j$ asa-no jyouhoubangumi-de
this-TOP morning-GEN talk.show-at
zessan-shi-ta] wakate hyouronka$_i$-o mi-ta-to zemi-no
praise-do-PST junior critic-ACC see-PST-that seminar-GEN
seito-ga it-ta] amari shira-re-tei-nai SF-shoosetsu$_j$-da.
student-NOM say-PST not.well know-PASS-PRES-NEG Sci-Fi novel-COP

'This is the not-well-known Sci-Fi novel [$_{RC}$ that a student from the seminar said that they saw a junior critic [$_{RC}$ who praised at a morning talk show]].'

Sentences with double relatives as in (13) do not seem very acceptable, presumably because they violate ObS; as relativization in Japanese is rightward movement, by relativizing *not-well-known Sci-Fi novel* out of the relative clause headed by *junior critic*, the semantic role assignment of *junior critic* is delayed until the end of sentence when *not-well-known Sci-Fi novel* appears. (13) could also be low in acceptability because the heavy NP *not-well-known Sci-Fi novel* is processed last, contrary to the long-before-short preference (see Section 5.3.2 for details). But crucially, on top of these factors, the sentence where the relative clause island is in a non-presuppositional environment (=(13a)) seems to be better than the one with a presuppositional relative clause island (=(13b)), in accord with the scrambling pair in (12).

If the presuppositionality of extraction domain affects the acceptability of Japanese double relatives as in (13), how about the ones that were tested in the experiments from Chapter 2? The double relatives from the chapter always involved relativization out of a relative clause in the passive subject position, as exemplified below.

(14) a. [$_{RC2}$ [$_{RC1}$ __$_i$ __$_j$ kai-ta] SF-shoosetsu$_j$-ga saikin
write-PST Sci-Fi novel-NOM recently
shoten-de tokushuu-sa-re-ta gakusha$_i$]-wa hokorashige-da.
bookstore-at feature-do-PASS-PST professor-TOP looks.proud-COP
'The professor$_i$ [$_{RC2}$ who the sci-fi novel$_j$ [$_{RC1}$ that __$_i$ wrote __$_j$] was recently featured in a bookstore] looks proud.'

b. [$_{RC2}$ [$_{RC1}$ __$_i$ __$_j$ kai-ta] gakusha$_i$-ga saikin
write-PST professor-NOM recently
shoten-de tokushuu-sa-re-ta SF-shoosetsu$_j$]-wa
bookstore-at feature-do-PASS-PST Sci-Fi novel-TOP
daigakusee-ni ninki-da.
college.students-DAT popular-COP

'The sci-fi novel$_j$ [$_{RC2}$ that the professor$_i$[$_{RC1}$ who __$_i$ wrote __$_j$] was recently featured in a bookstore] is popular among college students.'

Examining the precise presuppositionality of the extraction environment as they appeared in the experiments' stimuli has to be left for future research, but I speculate that the environment may permit an existential (non-presuppositional) interpretation for two reasons. First, the inner relative clause was always marked with the nominative case marker -*ga*, which contrasts with what is often referred to as the topic marker -*wa*, which can also attach to the subject noun phrase (e.g., Kuroda, 1965, 1992, 2005). Only the former can attach to the subject in an existential context without triggering any implied meanings; for instance, while *otokonohito-ga iru* [man-NOM there.is] simply means 'there is a man', *otokonohito-wa iru* [man-TOP there.is] is accompanied by a contrastive reading along the lines of 'there is the man (but there is no woman)'. And as indicated by the English translation, *wa*-marking may make the attached noun phrase definite, which is presuppositional, unlike *ga*-marking. Second, Section 2.2.3 of Chapter 2 presented Ishizuka's (2009) claim that double relatives are more acceptable with one group of predicates (unaccusatives, passives, and adjectival/nominal predicates) than with the other (unergatives, transitives), the former of which was referred to as *be*-type predicates. If *be*-type predicates correspond to the predicates that permit the existential reading of its argument, as seems at least plausible, it is not surprising that double relatives in (14), which involves a passive, have relatively high acceptability. It may be the case, then, that the passive subject position in examples like (14) is non-presuppositional, which would mean that extraction out of this position would be facilitated, and ODT would be satisfied as well. As Chapter 2 demonstrated, though, (14a) involving long-distance subject relativization is more acceptable than (14b) involving long-distance object relativization, for only the former also satisfies ObS.

Overall, these examples point to the conclusion that the set of factors affecting the acceptability of sentences with relative clause island violations stays the same across extraction types. And the cases where those factors are maximally controlled for (such as subject-extraction double relative and scrambling out of a non-presuppositional relative clause) enable us to see the relatively mild penalty of violating the relative clause island itself.

5.5 Chapter summary

This chapter has introduced a factor aside from the Object before Subject Bias and One Dependency at a Time Bias that could be modulating the acceptability of sentences with a relative clause island violation. Namely, it is known that extraction out of a relative clause is facilitated when the relative clause is in a non-presuppositional environment, such as the pivot of existential construction and the direct object of "evidential existential" verbs (Vincent et al., 2022). Featuring Japanese sentences with scrambling out of a relative clause, the experiment presented in this chapter manipulated the presuppositionality of the domain from which scrambling takes place. Specifically, stimuli in the experiment contrasted relative clauses in two extraction domains (the pivot of existential construction and the transitive object), and results revealed a significant island effect for both environments. Crucially, it was revealed that a non-presuppositional environment improves the acceptability of relative clauses with (and without) scrambling out of them. At the same time, though, the non-presuppositional environment does not make a relative clause island effect smaller, unlike what has been found in English double relatives. These seemingly different effects of the presuppositionality of the extraction domain were attributed to the fact that the presuppositional relative clause of my stimuli was always the

direct object of the verb *miru* ‘to see’, contrary to the study by Vincent et al. (2022) that utilized a wider range of less common transitive verbs.

Altogether, what appears to be a robust island effect exhibited by scrambling out of a relative clause in Japanese has turned out not to contradict the proposal that the penalty of violating a relative clause island is mild; even if a sentence satisfies processing biases and preferences such as ObS and ODT, it could still display a clear drop in acceptability due to other factors, such as extraction out of a presuppositional domain.

Chapter 6

General discussion

6.1 Summary

This dissertation focused on various instances of extraction out of a relative clause, traditionally thought to be a syntactic island, and the effect such extraction has on sentence acceptability. Considering the cross-linguistic variation in how sentences with a relative clause island violation have been judged, previous literature has proposed that languages differ in either the status of relative clauses as islands (i.e., some languages simply do not consider relative clauses as an island), or the possibility of deriving a structure that looks like it is violating a relative clause island but is not. The current dissertation has explored this proposal in depth, and by means of formal acceptability experiments and simulations with large language models, it has demonstrated that it may not be the correct strategy to account for the cross-linguistic variation. On the one hand, the findings from the previous chapters have confirmed what syntacticians have long thought: relative clauses are islands. On the other hand, however, the results suggest that the way that syntacticians have dealt with cross-linguistic variation in this domain, either by saying that languages vary as to the island status of relative clauses or that some languages offer ways to circumvent the potential island violation, is insufficient in the Japanese case. Instead, this dissertation has shown that relative clauses on their own yield only mild island effects,

contrary to what has traditionally been thought. In cases where the island effects appear to be more severe, as they often do, I proposed that this is because of independent factors that combine with the island violation to produce a much greater degradation in acceptability, and it is the presence or absence of these independent factors that results in the cross-linguistic variation observed.

Let us now summarize the findings from each of the chapter in further detail. Chapter 2 presented a series of acceptability experiments featuring Japanese double relatives, first with only the double relatives that have been judged to be well-formed (subject-extraction double relatives), and then with both subject-extraction and object-extraction double relatives, the latter of which have been judged ill-formed. The experiments revealed that even subject-extraction double relatives exhibit a significant relative clause island effect, and that the significant island effect was observed only when the dependency involves a gap (i.e., an A'-movement dependency), but not when it involves an anaphor *jibun* in place of the gap. Additionally, the size of the island effect turned out to be similar between subject-extraction and object-extraction double relatives. The lower acceptability of object-extraction double relatives was attributed to the fact that they involve resolving the subject dependency before the object dependency, which was formulated as the *Object before Subject Bias* (ObS), an idea originally proposed by Nakamura and Miyamoto (2013).

Chapter 3 took a computational approach to better understand the nature of the relative clause island effect, which was observed to be relatively mild. Leveraging the characteristic that language models do not undergo adaptation to sentences that are initially judged unacceptable, this chapter revealed a pattern of "acceptability judgments" among Japanese language models that closely resembles those among human participants in Chapter 2. In particular, most of the

language models tested in the chapter reproduced the small size of island effect among sentences with subject-extraction double relatives, which challenges the idea that the small island effect observed among humans is due to adaptation or aggregating over varying judgments.

Chapter 4 investigated the case of double relatives in English, which have mostly been judged to be clearly ill-formed. The chapter demonstrated that some additional factors lower the acceptability of English double relatives, including the subject island effect and the COMP-trace effect. In addition to these factors, the chapter proposed the *One Dependency at a Time Bias* (ODT), which predicts an additional processing cost when multiple long-distance dependencies are resolved simultaneously.

Chapter 5 turned to another type of dependency across a relative clause: scrambling, which has been shown to display a clear island effect when it takes place out of a relative clause. The chapter proposed, however, that scrambling out of a relative clause is another case in which additional factors (in this case, the presuppositionality of extraction domain) gives rise to what looks like a strong relative clause island effect. Similar to the relatively well-formed instances of relative clause island violation in languages such as English, scrambling out of a relative clause led to only a small superadditive drop in acceptability when the relative clause was in a non-presuppositional environment (e.g., pivot of an existential construction).

Altogether, the findings from the prior chapters suggest that relative clauses themselves are comparatively mild islands, but they can result in more severe effects when the structure violates additional syntactic constraints or causes additional processing difficulties. The overall picture is that while extraction out of a relative clause results in significant degradation in acceptability, producing a structure that goes against other constrains and biases can degrade the sentence even further. In the following sections, I discuss the implications of the current

dissertation on the status of relative clause island across different languages as well as across different types of dependency within a language. I then conclude with the future directions.

6.2 Implications across languages and dependency types

This dissertation has focused on a variety of factors that have the potential of making extraction out of a relative clause more difficult for one language than another, or that cause variability in acceptability within a language. Those factors are grouped into categories and summarized with examples below.

A. Processing-oriented factors

(1) *Object before Subject Bias (ObS)*

Evidence (from Experiment 3b, Chapter 2 Section 2.5.2): In Japanese, object-extraction double relatives (=b) are less acceptable than subject-extraction double relatives (=a) as the latter involves a delay in assigning the semantic role to the subject.

a.	[$_{RC2}$ [$_{RC1}$ __$_i$ __$_j$	kai-ta]	SF-shoosetsu$_j$-ga	saikin
		write-PST	Sci-Fi novel-NOM	recently
	shoten-de	tokushuu-sa-re-ta	gakusha$_i$]-wa	hokorashige-da.
	bookstore-at	feature-do-PASS-PST	professor-TOP	looks.proud-COP

'The professor$_i$ [$_{RC2}$ who the sci-fi novel$_j$ [$_{RC1}$ that __$_i$ wrote __$_j$] was recently featured in a bookstore] looks proud.'

b.	[$_{RC2}$ [$_{RC1}$ __$_i$ __$_j$	kai-ta]	gakusha$_i$-ga	saikin
		write-PST	professor-NOM	recently
	shoten-de	tokushuu-sa-re-ta	SF-shoosetsu$_j$]-wa	
	bookstore-at	feature-do-PASS-PST	Sci-Fi novel-TOP	
	daigakusee-ni		ninki-da.	
	college.students-DAT		popular-COP	

'The sci-fi novel$_j$ [$_{RC2}$ that the professor$_i$[$_{RC1}$ who __$_i$ wrote __$_j$] was recently featured in a bookstore] is popular among college students.'

(2) *One Dependency at a Time Bias (ODT)*

Evidence (from Experiment 2, Chapter 4 Section 4.3): English double relatives tend to have low acceptability (=a,b) when they involve simultaneous or nearly simultaneous resolution of multiple long-distance dependencies, regardless of which argument undergoes extraction of a longer distance.

a. I praised the teacher$_j$ [$_{RC}$ that I remember the book$_i$ [$_{RC}$ that I gave __$_i$ to __$_j$ last year]].

b. I praised the book$_i$ [$_{RC}$ that I remember the teacher$_j$ [$_{RC}$ that I gave __$_i$ to __$_j$ last year]].

(3) *Crossing versus nested dependency*

Evidence (from Chapter 4, Section 4.4): English double relatives are less acceptable when they form crossed dependencies as a result of relativization (=b) compared with when they form nested dependencies, but the contrast in acceptability is perceptible only when double relatives involve relativization out of an infinitival relative clause.

a. I just bought a violin$_j$ [that I composed a sonata$_i$ [to play __$_i$ on __$_j$]].

b. I just composed a sonata$_i$ [that I bought a violin$_j$ [to play __$_i$ on __$_j$]].

B. Syntax-oriented factors

(4) *Subject island effect*

Evidence (from Chapter 4, Section 4.1): English double relatives where the inner relative clause is in the subject position (=b) are less acceptable than the ones where the inner relative clause is in the object position (=a).

a. The professor$_i$ [$_{RC2}$ who I liked the sci-fi novel$_j$ [$_{RC1}$ that __$_i$ wrote __$_j$]] looks proud.

b. The professor$_i$ [$_{RC2}$ who the sci-fi novel$_j$ [$_{RC1}$ that __$_i$ wrote __$_j$] sold well] looks proud.

(5) *COMP-trace effect*

Evidence (from Experiment 1, Chapter 4 Section 4.2): English double relatives where the subject gets relativized out of another relative clause (=b) are less acceptable than ones where the object gets relativized (=a).

a. I waved at the children$_j$ [$_{RC}$ who I know the teacher$_i$ [$_{RC}$ who __$_i$ is running with __$_j$]].

b. I waved at the children$_i$ [$_{RC}$ who I know the teacher$_j$ [$_{RC}$ that __$_i$ are running with __$_j$]].

C. Semantics-oriented factor

(6) *Presuppositionality of extraction domain*

Evidence (from the experiment in Chapter 5, Section 5.3): Japanese sentences with scrambling out of a relative clause (as well as English double relatives) are less acceptable when extraction takes place out of a presuppositional environment (=b) compared with a non-presuppositional environment (=a).

a. Amari shira-re-tei-nai SF-shoosetsu-o$_j$ zemi-no seito-wa
not.well know-PASS-PRS-NEG Sci-Fi novel-ACC seminar-GEN student-TOP
[*to* [$_{RC}$ __$_i$ __$_j$ asa-no jyouhoubangumi-de zessan-shi-ta]
morning-GEN talk.show-at praise-do-PST
wakate hyouronka$_i$-ga iru-to] it-ta.
junior critic-NOM there.is-that say-PST
'A student from the seminar said [*to* that there is a junior critic [$_{RC}$ who praised the Sci-Fi novel that was not well-known at a morning talk show]].'

b. Amari shira-re-tei-nai SF-shoosetsu-o$_j$ zemi-no seito-wa
not.well know-PASS-PRS-NEG Sci-Fi novel-ACC seminar-GEN student-TOP
[*to* [$_{RC}$ __$_i$ __$_j$ asa-no jyouhoubangumi-de zessan-shi-ta]
morning-GEN talk.show-at praise-do-PST
wakate hyouronka$_i$-o mi-ta-to] it-ta.
junior critic-ACC see-PST-that say-PST
'A student from the seminar said [*to* that they saw a junior critic [$_{RC}$ who praised the Sci-Fi novel that was not well-known at a morning talk show]].'

Recall that, while some of the factors above such as ODT (=(2)) seem to be at play only in the condition with the extraction out of a relative clause (i.e., cases involving multiple long-distance dependencies), other factors such as ObS (=(1)), COMP-trace effect (=(5)), and the presuppositionality of extraction domain (=(6)) seem to have a broader impact; for instance, ObS is active for long-distance relativization out of not only a relative clause island, but also a *koto*-clause (a non-island). Likewise, in both Vincent et al.'s (2022) English experiment and my

Japanese experiment from Chapter 5, having a presuppositional environment itself appears to lower acceptability, with or without an extraction out of it.

The following is another summary of how each of the factors listed above has played out (excluding (3), for which we did not find experimental evidence):

(7) *Japanese double relatives* (Chapter 2)

ObS (=(1)): This is possible to satisfy, as exemplified in (1a).
ODT (=(2)): This is always satisfied, as exemplified in (1a,b).
Subject island (=(4)): This is either non-existent or extremely small.
COMP-trace (=(5)): This is not relevant.
Presuppositionality of extraction domain (=(6)): At least in the stimuli presented in Experiment 2, it was always the *ga*-marked passive subject that was relativized, which appears to count as a non-presuppositional environment.

Outcome: So long as Japanese double relatives satisfy (1) and (6), their acceptability should be relatively high despite the relative clause island violation.

(8) *English double relatives* (Chapter 4)

ObS (=(1)): This is impossible to satisfy, because of its word order property.
ODT (=(2)): This is possible to satisfy, as exemplified below (from Kuno, 1976).

This is the child who there is nobody who is willing to accept.

Subject island (=(4)): This is possible to avoid, as exemplified in (4a).
COMP-trace (=(5)): This is possible to avoid, as exemplified in (5a).
Presuppositionality of extraction domain (=(6)): This is possible to control, as demonstrated in Vincent et al. (2022), while most of the English sentences with double relatives that have been featured in this dissertation (e.g., (2), (4), (5)) violate this constraint.

Outcome: Some of the factors such as (1) are unavoidable, and though English double relatives of relatively high acceptability do exist, there are a number of factors that need to be taken into account in order to obtain the acceptability.

(9) *Japanese scrambling out of a relative clause* (Chapter 5)

ObS (=(1)): This is always satisfied, as exemplified in (6a,b), and scrambling of the subject (which would violate the ObS) is disallowed on independent grounds.
ODT (=(2)): This is always satisfied, as exemplified in (6a,b),
Subject island (=(4)): This is either non-existent or extremely small.
COMP-trace (=(5)): This is not relevant.
Presuppositionality of extraction domain (=(6)): This is possible to control, as exemplified in (6a). Previous observations regarding scrambling out of a relative clause involved a presuppositional extraction environment, in violation of this constraint.

Outcome: So long as Japanese scrambling out of a relative clause satisfies (6), its acceptability should be relatively high despite the relative clause island violation.

Given the list of factors that need to be satisfied (and potentially other factors that we have not explored here), the cross-linguistic variability in the status of relative clause may boil down to how feasible it is to satisfy as many of those factors as possible. In languages such as Japanese, their word order properties (namely having pre-nominal relative clauses) make it possible to create double relatives that satisfy both ObS and ODT. In addition, both subject island effects and COMP-trace effects seem to be absent and thus do not need to be controlled for in Japanese. It is perhaps for these reasons that double relatives seem to be particularly well known in Japanese, Korean and Chinese, all of which have pre-nominal relative clauses, as exemplified in (10) below.

(10) a. [$_{RC2}$ [$_{RC1}$ __$_i$ __$_j$ ki-tei-ru] fuku$_j$-ga yogoretei-ru] shinshi$_i$
wear-PROG-PRS clothes-NOM dirty-PRS gentleman
'the gentleman$_i$ [$_{RC2}$ who the clothes$_j$ [$_{RC1}$ that __$_i$ is wearing __$_j$] are dirty]'
(Japanese; Kuno, 1973)

b. [$_{RC2}$ [$_{RC1}$ __$_i$ __$_j$ cohaha-nun] kangaci$_j$-ka cwuk-un] ai$_i$
like-ADN dog-NOM die-ADN kid
'The kid$_i$ [$_{RC2}$ who the dog$_j$ [$_{RC1}$ that __$_i$ liked __$_j$] die]'
(Korean; Han & Kim, 2004)

c. [$_{RC2}$ [$_{RC1}$ __$_i$ chuan __$_j$ de yifu$_j$ hen piaoliang] de na-ge ren$_i$
wear REL dress very pretty REL that-CL person
'The person$_i$ [$_{RC2}$ who the dress$_j$ [$_{RC1}$ that __$_i$ is wearing __$_j$] is very pretty]'
(Chinese; Aoun & Li, 2003)

All of the examples resolve the object dependency and then the subject dependency, in accord with both ObS and ODT. Furthermore, all of the extraction environments involve the so-called *be*-type predicate, which may correspond to the predicates that accommodate the existential reading of their argument (see Section 5.4.3 for details).

Conversely, while we have also observed relatively well-formed instances of relative clause island violations among languages such as English (with post-nominal relative clauses), as exemplified in (11) below.

(11) a. This is the child who there is nobody$_j$ [$_{RC}$ who __$_i$ is willing to accept _$_j$].
(English; Kuno, 1976)

b. Suppe$_j$ kender jeg mange$_i$ [$_{RC}$ der __$_i$ kan lide __$_j$].
soup know I many who can like
'Soup, I know many people who like.'
(Danish; Erteschick-Shir, 1973)

c. Al lexem Saxor$_j$, yeS rak gvina axat$_i$
on bread black be only cheese one
[$_{RC}$ Se-keday limraox __$_i$ __$_j$].
that-worthy to.spread
'On black bread, there is only one cheese that's worth spreading.'
(Hebrew; Sichel, 2018)

d. Gianni$_j$, al quale non c'è nessuno [$_{RC}$ che
Gianni for whom NEG there.is nobody who
__$_i$ sia in grado di resistere__$_j$],
is able to resist
'Gianni, whom there is nobody that is able to resist, …'
(Italian; Cinque, 2010)

Languages exemplified appear to have no way of controlling ObS, and need to take additional factors such as the subject island and the COMP-trace effect into account, and thus such instances are less frequent than in languages like Japanese.

Based on the cross-linguistic pattern observed above, as well as in the previous chapters, I propose that the factors listed in (1) through (6) have a cumulative effect; the more factors a sentence with a relative clause island violation further violates/fails to control for, the lower their acceptability is. Consequently, the more factors a language needs to take into account in order to derive a relatively well-formed instance of relative clause island violation, the less likely it may be for the speakers of the language to encounter and produce those instances. At the same time, the different factors may not have equal weights; for instance, even though English double relatives can never satisfy ObS (=(1)), it is still possible to derive relatively well-formed instances (as exemplified in (11)) so long as the presuppositionality of extraction domain (=(6)) is controlled for. I speculate, however, that the acceptability of those instances would resemble that of Japanese object-extraction double relatives (=(1b); not satisfying ObS) rather than that of subject-extraction double relatives (=(1a); satisfying ObS) if it were possible to set up an experiment that warrants a direct comparison of the results of the two languages.

Note that we are able to account for the cross-linguistic variation in the status of relative clause island effects without resorting to the special mechanisms that have been invoked in the past to deal with the variation. That is, we have not needed to say that languages like Japanese versus English differ in the island status of relative clauses, or that the former have a special structure that allows it to give the appearance of extracting from a relative clause without actually doing so. Instead, my analysis claims that languages are identical with regard to the island status of relative clauses; the superficial differences that arise are due to the ease of

satisfying all the additional relevant factors that amplify the penalty of a relative clause island violation.

Similar to the uniformity of relative clause island effects across languages, the analysis put forth in this dissertation suggests that the same size of relative clause island effect is invoked regardless of the type of extraction. For Japanese, three types of dependency across a relative clause have been discussed; *wh*-dependency, relativization, and scrambling, as illustrated in the following:

(12) a. *wh-dependency across a relative clause*
Momoko-wa [RC__$_i$ nani-o kat-ta] otokonohito$_i$-o mimashi-ta ka?
Momoko-TOP what-ACC buy-PST man-ACC see-PST Q
'What did Momoko see the man$_i$ [RC that __$_i$ bought <what>]]?'
(Tanaka & Schwartz, 2018)

b. *relativization across a relative clause* (=(10a))
[RC2 [RC1 __$_i$ __$_j$ ki-tei-ru] fuku$_j$-ga yogoretei-ru] shinshi$_i$
wear-PROG-PRS clothes-NOM dirty-PRS gentleman
'the gentleman$_i$ [RC2 who the clothes$_j$ [RC1 that __$_i$ is wearing __$_j$] are dirty]'
(Kuno, 1973)

c. *scrambling across a relative clause*
Ano hon$_j$-o John-ga [RC __$_i$ __$_j$ kai-ta] hito$_i$-o
that book-ACC John-NOM write-PST person-ACC
sagashi-tei-ru rasii.
look.for-PROG-PRES seem
'That book, John seems to be looking for the person who wrote (it).'
(Saito, 1985)

For a *wh*-dependency across a relative clause (=(12a)), Tanaka and Schwartz (2018) revealed that it exhibits a small yet significant superadditive drop in acceptability in a factorial-design acceptability experiment, despite its having been considered well-formed previously. Likewise, the experiments from Chapter 2 showed that relativization of the subject (=(12b))

across another relative clause leads to the same size of superadditive effect as the relativization of the object (=(12b)) across another relative clause, the latter of which has been judged less acceptable than the former. For scrambling out of a relative clause (=(12c)), though the previous studies have judged it to be ill-formed without exception, taking presuppositionality into account has enabled us to observe more acceptable cases, as in (6a). Again, my analysis makes it unnecessary to claim that the size of a relative clause island effect varies across dependency types, similar to the way the analysis treats languages equally when it comes to the existence and size of the penalty of violating a relative clause island.

6.3 The nature of the "residual" effect

In addition to the factors influencing the acceptability of sentences with a relative clause island violation (including (1) through (6)), the previous chapters also revealed that a small yet significant "residual" effect of violating the island on acceptability is detected even when a sentence successfully satisfies all of those factors. One such example was seen in the subject-extraction double relatives in Japanese (e.g., (1a)), where such a residual effect was observed among human participants in Experiment 1 of Chapter 2 (Section 2.3), as well as large language models in Chapter 3. Another example was from the case of scrambling out of a relative clause in Japanese, when the extraction domain was non-presuppositional. One possible explanation for such a residual effect is that there is a factor that is currently overlooked, and the residual effect would be eliminated upon controlling for it. Another possibility, which this dissertation has assumed, goes as follows: Although one manages to follow various constraints as laid out in (1) through (6) that come into play in extracting out of a relative clause, the simple act of extracting out of the relative clause triggers a small penalty, and this is what accounts for the "residual"

effect. Critically, the latter possibility challenges the long-standing observation that a relative clause exhibits a very large island effect when violated. The traditional analysis of the relative clause island contrasts with the other types of islands, most famously the *wh*-island, which has been known to be "selective" (or "subliminal"; Almeida, 2014); factors such as argument (as opposed to adjunct) extraction as well as discourse-linking are known to facilitate extraction, as the following exemplifies:

(13) a. How$_j$ do you wonder [who solved __$_j$]? [adjunct extraction; non-D-linked]
b. What$_j$ do you wonder [who solved __$_j$]? [argument extraction; non-D-linked]
c. Which problem$_j$ do you wonder [who solved __$_j$]? [argument extraction; D-linked]

Even in cases like (13c), however, studies have shown that a small *wh*-island effect persists in acceptability experiments (Goodall, 2015; Sprouse et al., 2016) completely.

Findings from this dissertation about the relative clause island thus paint a potentially more coherent picture than the traditional analysis, with regard to the other types of island; on one hand, there are cases (subject-extraction double relatives, *wh*-island violation as in (13c)) that have been thought to allow extraction out of an island but still exhibit a small residual effect; on the other hand, what has been known as a strong island effect (object-extraction double relatives, *wh*-island violation as in (13a)) can be seen as a combination of the penalty for the island violation per se and any additional factors.

The analysis that the penalty of violating an island itself is small, even for islands such as relative clauses, opens up a host of questions, including the source of island effects and the mechanism of A'-movement dependency at large. While the exact nature of the penalty of relative clause island violation revealed in this dissertation needs to be left for future research

(see Sprouse & Villata, 2021 for a summary on the source of island effects), the fact that the penalty was observed in a novel structure (Japanese double relatives) that was traditionally considered well-formed, that the penalty persisted even after removing a number of factors, and that it was only detected in an A'-movement dependency (it was not detected once the dependency involved the anaphor *jibun*) hopefully contribute to the ongoing debate about what it means to be an island effect.

6.4 Future directions

The present dissertation has focused primarily on relative clause islands, proposing that what has been referred to as a relative clause island can in fact be split apart into a number of factors, only one of which is the island constraint per se. It is possible that a similar scenario holds with other types of islands, and there have been some proposals to this effect. For example, Haegeman et al. (2014) have argued that subject islands can also be "deconstructed" into several independent constraints on movement, including, but not limited to, the following:

(14) *Freezing Principle*: A moved constituent is frozen for extraction.

a. Which candidate were there [posters of __] all over the town?
b. *Which candidate were [posters of __] all over the town?

Edge Condition: The edge of a phase is opaque for extraction.

a. Which candidate were there [posters of __] all over the town?
b. *Who do you wonder [which posters of __] are all over the town?

In each of the examples above, the (b) sentences are judged to be less acceptable than the (a) sentences due to the violation, under their analysis, of the constraint mentioned. Haegeman et

al. (2014) propose that the constraints exemplified in (14) have a cumulative effect, where an extraction out of the subject is judged the less acceptable the more relevant constraints it violates.

The proposal that the subject island effect can be "deconstructed" thus has some similarity with this dissertation's proposal regarding the relative clause island effect. Although a study by Greco et al. (2017) attempted to put Haegeman et al.'s proposal to the test through acceptability experiments and failed to find evidence for a cumulative effect, a follow-up study with an updated experimental design would be necessary.

Moreover, with regard to the study with large language models presented in Chapter 3, there are multiple opportunities for a follow-up study given that it was one of the first studies to probe the models' knowledge of long-distance dependencies and their relevant constraints in a language that is not English. In particular, featuring other types of extraction (e.g., scrambling) as well as other types of islands (e.g., a complex noun phrase that is not a relative clause) would be important in order to confirm not only the conclusion from Chapter 3 that language models trained with Japanese text can learn long-distance dependencies, but also the experimental design and metrics (e.g., licensing interaction) employed in the chapter to reach the conclusion. In addition, the experiment in Chapter 3 used four versions of a pre-trained Japanese language model, which varies in the number of parameters, while keeping the amount of training data consistent (approximately 553 million sentences). It is of interest to vary the size of training data to see whether it affects the learning outcome; Wilcox et al. (2023) probed and demonstrated the island knowledge of several English language models, including one (pre-trained by Gulordava et al., 2018) trained with the amount of language data that they estimated to be equivalent to a quantity of linguistic experience of an 8-year-old (approximately 90 million tokens). Learnability is one of the major topics of debate among researchers when it comes to islands (Chomsky,

1975; Otsu, 1981; Goodluck et al., 1992; Richards, 2001; Pullum & Scholz, 2002; Pearl & Sprouse, 2013), and if the findings of Wilcox et al. (2023) could be replicated with a Japanese language model that was trained with the amount of data that a Japanese-speaking child could reasonably be exposed to, that could lend support to the argument that grammatical generalizations are learnable from input. At the same time, we could explore how to improve the architecture of language models so that it would learn the grammatical constraints with much less amount of data than what the state-of-the-art language models currently require.[1]

All in all, it remains to be seen to what extent the analysis and findings (both behavioral and computational) presented in this dissertation about relative clause island effects can be extended to other languages, other island types, and other forms of A'-movement. I believe, however, that they provide an attractive and promising approach to cross-linguistic variability in island phenomena.

[1] GPT-3 (Brown et al., 2020) was trained on approximately 114 billion words, which means that if a person is consistently exposed to ≈30,000 words per day (≈11 million words per year) throughout their lifetime, GPT-3's training data is equivalent of ≈100 human lifetimes (assuming that 1 lifetime = 100 years).

Bibliography

Alexopoulou, T., & Keller, F. (2007). Locality, cyclicity, and resumption: At the interface between the grammar and the human sentence processor. *Language, 83*(1), 110-160.

Allwood, J. (1982). The Complex NP Constraint in Swedish. In Engdahl, E., & Ejerhed, E. (eds.), *Readings on Unbounded Dependencies in Scandinavian Languages*. Almqvist & Wiksell International, 15-32.

Allwood, J. S. (1976). The complex NP constraint as a non-universal rule and some semantic factors influencing the acceptability of Swedish sentences which violate the CNPC. *University of Massachusetts occasional papers in linguistics*, *2*(1), 2.

Almeida, D. (2014). Subliminal wh-islands in brazilian portuguese and the consequences for syntactic theory. *Revista da ABRALIN*, *13*(2), 55-93.

Andersson, L. G. (1982). What is Swedish an exception to? Extractions and island constraints. *Readings on unbounded dependencies in Scandinavian languages*, 33-45.

Aoshima, S., Yoshida, M., & Phillips, C. (2009). Incremental processing of coreference and binding in Japanese. *Syntax*, *12*(2), 93-134. https://doi.org/10.1111/j.1467-9612.2009.00123.x

Aoun, J., & Li, Y. H. A. (2003). *Essays on the representational and derivational nature of grammar: The diversity of wh-constructions*. MIT press.

Bach, E., Brown, C., & Marslen-Wilson, W. (1986). Crossed and nested dependencies in German and Dutch: A psycholinguistic study. *Language and Cognitive Processes*, *1*(4), 249-262.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bernardy, J. P., & Lappin, S. (2017). Using deep neural networks to learn syntactic agreement. *Linguistic Issues in Language Technology, 15*, 1-15.

Bianchi, V., & Chesi, C. (2014). Subject islands, reconstruction, and the flow of the computation. *Linguistic Inquiry*, *45*(4), 525-569.

Borer, H. (1981). The definition of variables. *Journal of Linguistic Research*, *1*, 16-40.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877-1901.

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software*, *80*, 1-28.

Chomsky, N. (1964). [the development of grammar in child language]: discussion. *Monographs of the Society for Research in Child development*, 35-42.

Chomsky, N. (1973). Conditions on transformations. In Anderson, S. R. & Kiparsky, P. (eds.), *A Festschrift for Morris Halle*, 232-286. Holt, Rinehalt, and Winston.

Chomsky, N. (1975). Questions of form and interpretation. *Scope of American linguistics*, 159-196.

Chomsky, N. (1977). On Wh-movement. In Culicover, P., Akmajian, A., & Wasow, T. (eds.), *Formal Syntax*, Academic Press, 71-132.

Chomsky, N. (1981). *Lectures on government and binding*. Foris.

Chomsky, N. (1986). *Barriers*. MIT Press.

Chung, S., & McCloskey, J. (1983). On the Interpretation of Certain Island Facts in GPSG. *Linguistic Inquiry*, *14*, 704-13.

Cinque, G. (1990). *Types of A'-dependencies*. MIT Press.

Cinque, G. (2010). On a Selective 'Violation' of the Complex NP Constraint. In Zwart, J. W., & de Vries, M. (eds.), *Structure Preserved: Studies in Syntax for Jan Koster*. John Benjamins Publishing Company, 81-89.

Cowart, W. (1997). *Experimental syntax: Applying objective methods to sentence judgments*. SAGE Publications.

Cowart, W., & McDaniel, D. (2021) The *that*-trace effect. In: Goodall, G. (ed.), *The Cambridge handbook of experimental syntax*. Cambridge University Press 258-77.

Culicover, P. W., & Rochemont, M. S. (1990). Extraposition and the complement principle. *Linguistic Inquiry*, *21*(1), 23-47.

Diesing, M. (1992). Bare plural subjects and the derivation of logical representations. *Linguistic Inquiry*, *23*(3), 353-380.

Drummond, A. (2013). Ibex farm. Online server: http://spellout.net/ibexfarm

Dryer, M. S. (1980). The positional tendencies of sentential noun phrases in universal grammar. *The Canadian Journal of Linguistics, 25*(2), 123-195. https://doi.org/10.1017/S0008413100009373.

Engdahl, E. (1982). Restrictions on unbounded dependencies in Swedish. *Readings on unbounded dependencies in Scandinavian languages*, 151-174.

Engdahl, E. (1985). Parasitic gaps, resumptive pronouns, and subject extractions. *Linguistics*, *23*, 3-44.

Engdahl, E. (1997). Relative Clause Extractions in Context. *Working Papers in Scandinavian Syntax, 60*, 51-79.

Erteschik-Shir, N. (1973). *On the nature of island constraints*. Doctoral dissertation, MIT, Cambridge, MA.

Erteschik-Shir, N., & Lappin, S. (1979). Dominance and the functional explanation of island phenomena. *Theoretical linguistics, 6*, 41-86.

Fanselow, G. (2021). Acceptability, grammar and processing. In Goodall, G. (ed.), *The Cambridge Handbook of Experimental Syntax*. Cambridge University Press.

Fiengo, R., & Higginbotham, J. (1981). Opacity in NP. *Linguistic Analysis*, *7*(4), 395-421.

Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry*, *9*, 427-473.

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and language*, *140*, 1-11.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6*(4), 291-325. https://doi.org/10.1016/0010-0277(78)90002-1

Frazier, L., & Clifton Jr, C. (1989). Successive cyclicity in the grammar and the parser. *Language and cognitive processes*, *4*(2), 93-126.

Fukuda, S., & Sprouse, J. (2017). *Are there islands in Japanese?* Ms., University of Hawaii at Manoa and University of Connecticut.

Fukuda, S., Tanaka, N., Ono, H., & Sprouse, J. (2022). An experimental reassessment of complex NP islands with NP-scrambling in Japanese. *Glossa: A Journal of General Linguistics*, *7*(1), 1-35. https://doi.org/10.16995/glossa.5737

Fukui, N., & Y. Takano. (2000). Nominal structure: An extension of the symmetry principle. In Svenonius, P. (ed.), *The derivation of VO and OV*, 219-254.

Futrell, R., Wilcox, E., Morita, T., & Levy, R. (2018). RNNs as psycholinguistic subjects: Syntactic state and grammatical dependency. *arXiv preprint arXiv:1809.01329*.

Futrell, R., Wilcox, E., Morita, T., Qian, P., Ballesteros, M., & Levy, R. (2019). Neural language models as psycholinguistic subjects: Representations of syntactic state. *arXiv preprint arXiv:1903.03260*.

Goodall, G. (2015). The D-linking effect on extraction from islands and non-islands. *Frontiers in Psychology*, *5*, 1493-1493.

Goodall, G. (2021). Sentence acceptability experiments: What, how, and why. In Goodall, G. (ed.). *The Cambridge Handbook of Experimental Syntax* (Cambridge Handbooks in Language and Linguistics). Cambridge University Press, 7-38.

Goodkind, A., & Bicknell, K. (2018). Predictive power of word surprisal for reading times is a linear function of language model quality. In *Proceedings of the 8th workshop on cognitive modeling and computational linguistics (CMCL 2018)* (pp. 10-18).

Goodluck, H., Foley, M., & Sedivy, J. (1992). Adjunct islands and acquisition. In *Island Constraints: Theory, acquisition and processing* (pp. 181-194). Springer Netherlands.

Greco, C., Marelli, M., & Haegeman, L. (2017). External syntax and the Cumulative Effect in subject sub-extraction: An experimental evaluation. *The Linguistic Review*, *34*(3), 479-531.

Gulordava, K., Bojanowski, P., Grave, E., Linzen, T., & Baroni, M. (2018). Colorless green recurrent networks dream hierarchically. *arXiv preprint arXiv:1803.11138*.

Haegeman, L., Jiménez-Fernández, Á. L., & Radford, A. (2014). Deconstructing the Subject Condition in terms of cumulative constraint violation. *The Linguistic Review*, *31*(1), 73-150.

Haig, J. (1976). Shadow pronoun deletion in Japanese. *Linguistic Inquiry*, *7*, 363-371.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Second meeting of the north american chapter of the association for computational linguistics*.

Han, C. H. (2013). On the syntax of relative clauses in Korean. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, *58*(2), 319-347. https://doi.org/10.1017/S0008413100003054

Han, C. H., Elouazizi, N., Galeano, C., Görgülü, E., Hedberg, N., et al. (2012). Processing strategies and resumptive pronouns in English. In *Proceedings of the 30th West Coast Conference on Formal Linguistics*, 153-161.

Han, C.-H., & Kim, J.-B. (2004). Are there “double relative clauses” in Korean? *Linguistic Inquiry*, *35*(2), 315-337.

Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge University Press. https://doi.org/10.1017/CBO9780511554285.

Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. OUP Oxford.

Heestand, D., Xiang, M., & Polinsky, M. (2011). Resumption still does not rescue islands. *Linguistic Inquiry*, *42*(1), 138-152.

Heilbron, M., Armeni, K., Schoffelen, J. M., Hagoort, P., & De Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. *Proceedings of the National Academy of Sciences*, *119*(32), e2201968119.

Herculano-Houzel, S. (2009). The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in human neuroscience*, *3*, 857.

Hewett, M. R. (2023). Types of Resumptive Ā-Dependencies. Doctoral dissertation, The University of Chicago, Chicago, IL.

Heycock, C., (1993). Syntactic predication in Japanese. *Journal of East Asian Linguistics*, *2*, 167-211.

Hsu, C,-C. N., Phillips, C., & Yoshida, M. (2005). *Cues for head-final relative clauses in Chinese*. Poster presented at the 18th Annual CUNY Conference on Human Sentence Processing. Tucson, AZ: University of Arizona.

Huang, C. T. J. (1982a). Move-wh in a language without wh-movement. *The Linguistic Review, 1*, 369-416.

Huang, C.-T. J. (1982b). *Logical relations in Chinese and the theory of grammar*. Doctoral dissertation, MIT, Cambridge, MA.

Ishizuka, T. (2009). CNPC Violations and Possessor Raising in Japanese. In Potter, D. & Storoshenko, D. R. (eds.), In *the Proceedings of the 2nd International Conference on East Asian Linguistics*.

Jeffreys, H. (1939/1961). *Theory of probability*. Oxford University Press.

Jurka, J. (2010). *The importance of being a complement: The CED effects revisited*. Doctoral Dissertation, University of Maryland.

Jurka, J., Nakao, C., and Omaki. A. (2011). It’s not the end of the CED as we know it: Revisiting German and Japanese subject islands. In Washburn, M. B., McKinney-Bock, K., Varis,

E., Sawyer, A., and Tomaszewicz, B. (eds.), *West Coast Conference on Formal Linguistics (WCCFL) 28*. 124-132. Cascadilla Press.

Kaplan, T. I., & Whitman, J. B. (1995). The category of relative clauses in Japanese, with reference to Korean. *Journal of East Asian Linguistics*, *4*(1), 29-58.

Keshev, M., & Meltzer-Asscher, A. (2019). A processing-based account of subliminal wh-island effects. *Natural Language and Linguistic Theory*, *37*(2), 621-657. https://doi.org/10.1007/s11049-018-9416-1

Kishida, M. (2011). *Reflexives in Japanese*. Doctoral dissertation, University of Maryland, College Park, MD.

Kluender, R., & Kutas, M. (1993). Subjacency as a processing phenomenon. *Language and cognitive processes*, *8*(4), 573-633. https://doi.org/10.1080/01690969308407588

Kluender, R. (1998). On the distinction between strong and weak islands: A processing perspective. In *The limits of syntax* (pp. 241-279). Brill.

Koopman, H. (1984). *The Syntax of Verbs: From Verb-Movement Rules in the Kru Languages to Universal Grammar*. Foris.

Kuroda, S. Y. (1965). *Generative grammatical studies in the Japanese language*. Doctoral dissertation, Massachusetts Institute of Technology.

Kuroda, S. Y. (1965). *Japanese syntax and semantics*. *Studies in natural language and linguistic theory*. Kluwer Academic.

Kuroda, S. Y. (2005). Focusing on the matter of topic: A study of wa and ga in Japanese. *Journal of east asian linguistics*, *14*(1), 1-58.

Kuncoro, A., Dyer, C., Hale, J., Yogatama, D., Clark, S., & Blunsom, P. (2018). LSTMs can learn syntax-sensitive dependencies well, but modeling structure makes them better. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 1*, 1426-1436.

Kuno, S. (1973). *The structure of the Japanese language*. MIT Press.

Kuno, S. (1976). Subject, Theme, and the Speaker's Empathy—A Reexamination of Relativization Phenomena. In Li, C. N. (ed.) *Subject and Topic*. Academic Press, Inc., 417-444.

Kuno, S., & Robinson, J. J. (1972). Multiple wh questions. *Linguistic inquiry*, *3*(4), 463-487.

Kuroda, S. Y., (1986). Movement of noun phrases in Japanese. In Imai, T., & Saito, M. (eds.), *Issues in Japanese Linguistics*, Foris, 229-271.

Kush, D. (2011). *On the (in-)escapability of islands in Scandinavian*. Ms., University of Maryland.

Kush, D., Lohndal, T., & Sprouse, J. (2018). Investigating variation in island effects. N*atural language & linguistic theory*, *36*(3), 743-779. https://doi.org/10.1007/s11049-017-9390-z

Kush, D., Lohndal, T., & Sprouse, J. (2019). On the island sensitivity of topicalization in Norwegian: An experimental investigation. *Language*, *95*(3), 393-420.

Kush, D., Omaki, A., & Hornstein, N. (2013). Microvariation in islands?. In Sprouse, J., & Hornstein, N. (eds.), *Experimental syntax and island effects*, 239-264.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software*, *82*(13), 1-26. https://doi.org/10.18637/jss.v082.i13

Kwon, N., Kluender, R., Kutas, M., & Polinsky, M. (2013). Subject/object processing asymmetries in Korean relative clauses: Evidence from ERP data. *Language*, *89*(3), 537.

Lasnik, H., & Saito, M. (1992). *Move alpha: Conditions on its application and output*. MIT press.

Lau, J. H., Clark, A., & Lappin, S. (2017). Grammaticality, acceptability, and probability: A probabilistic view of linguistic knowledge. *Cognitive Science, 41*(5), 1202-1241.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126-1177.

Lin, C. J. C., & Bever, T. G. (2011). Garden path and the comprehension of head-final relative clauses. *Processing and producing head-final structures*, 277-297. https://doi.org/10.1007/978-90-481-9213-7_13

Linzen, T., Dupoux, E., & Goldberg, Y. (2016). Assessing the ability of LSTMs to learn syntax-sensitive dependencies. *Transactions of the Association for Computational Linguistics*, *4*, 521-535.

Lu, J., Thompson, C. K., & Yoshida, M. (2020). Chinese wh-in-situ and islands: A formal judgment study. *Linguistic Inquiry*, *51*(3), 611-623. https://doi.org/10.1162/ling_a_00343

Marantz, A. (1984). *On the nature of grammatical relations*. MIT Press.

Marvin, R., & Linzen, T. (2018). Targeted syntactic evaluation of language models. *arXiv preprint arXiv:1808.09031*.

Maxwell, D. N. (1979). Strategies of Relativization and NP Accessibility. *Language*, *55*(2), 352-371. https://doi.org/10.2307/412589

McCawley, J. D. (1981). The syntax and semantics of English relative clauses. *Lingua*, *53*, 99-149.

McDaniel, D., Cowart, W., McKee, C., & Garrett, M. F. (2015). The role of the language production system in shaping grammars. *Language*, *91*(2), 415-441.

Mikami, A. (1963). *Nihongo-No-Ronri, Wa-to-Ga* [Logic of Japanese, Wa and Ga]. Kurosio.

Miyamoto, E. T., & Nakamura, M. (2013). Unmet expectations in the comprehension of relative clauses in Japanese. In *Proceedings of the Annual Meeting of the Cognitive Science Society* 35.

Murasugi, K. (2000). An antisymmetry analysis of Japanese relative clauses. In Alexiadou, A., Law, P., Meinunger, A., & Wilder, C. (eds.), *The syntax of relative clauses*, 231-263.

Nakamura, M., & Miyamoto, E. T. (2013). The object before subject bias and the processing of double-gap relative clauses in Japanese. *Language and Cognitive Processes*, *28*(3), 303-334. https://doi.org/10.1080/01690965.2011.634179

Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational Ideas—Part II. *Language and Linguistics Compass*, *10*(11), 591-613.

Oh, B. D., & Schuler, W. (2023). Why does surprisal from larger transformer-based language models provide a poorer fit to human reading times?. *Transactions of the Association for Computational Linguistics*, *11*, 336-350.

Omaki, A., Fukuda, S., Nakao, C., & Polinsky, M. (2020). Subextraction in Japanese and subject-object symmetry. *Natural Language and Linguistic Theory*, *38*(2), 627-669. https://doi.org/10.1007/s11049-019-09449-8

Oshima, D. (2009). A note on backward anaphora in japanese. *Ibaraki Daigaku Ryuugaku-sentaa Kiyou* [Ibaraki University international student center bulletin], *7*, 79-89.

Otsu, Y. (1981). *Universal grammar and syntactic development in children: Toward a theory of syntactic development*. Doctoral dissertation, Massachusetts Institute of Technology.

Pearl, L., & Sprouse, J. (2013). Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem. *Language Acquisition*, *20*(1), 23-68.

Perlmutter, D. (1972) Evidence for shadow pronouns in French relativization. In Peranteau, P. M. et al. (eds.), *The Chicago Which Hunt: Papers from the Relative Clause Festival, Chicago Linguistic Society*, 73-105.

Pesetsky, D. (1982). Complementizer-trace phenomena and the nominative island condition. *The linguistic review*, *1*(3), 297-344.

Pesetsky, D. (1987). Binding problems with experiencer verbs. *Linguistic Inquiry*, *18*(1), 126-140.

Pesetsky, D. (2017). Complementizer-trace effects. *Blackwell Companion to Syntax.* Wiley.

Phillips, C. (2006). The real-time status of island phenomena. *Language*, 795-823.

Pickering, M., & Barry, G. (1991). Sentence processing without empty categories. *Language and cognitive processes*, *6*(3), 229-259.

Polinsky, M., Clemens, L. E., Milton, A., Morgan, M. X., & Heestand, D. (2013). Resumption in English. In Sprouse, J., & Hornstein, N. (eds.), *Experimental syntax and island effects*, 341-359.

Pullum, G. K., & Scholz, B. C. (2002). Empirical assessment of stimulus poverty arguments. *The linguistic review*, *19*(1-2), 9-50.

R Core Team (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Richards, N. (2001). *Movement in language: Interactions and architectures*. Oxford University Press.

Rizzi, L. (1990). *Relativized Minimality*. The MIT Press.

Rizzi, L. (2018). Intervention effects in grammar and language acquisition. *Probus*, *30*(2), 339-367. https://doi.org/10.1515/probus-2018-0006

Ross, J. R. (1967). *Constraints on variables in syntax*. Doctoral dissertation, MIT, Cambridge, MA.

Saito, M., (1982). Case Marking in Japanese: A Preliminary Study. Ms., MIT.

Saito, M. (1985). *Some asymmetries in Japanese and their theoretical implications*. Doctoral dissertation, MIT, Cambridge, MA.

Saito, M. (1992). Long distance scrambling in Japanese. *Journal of East Asian Linguistics*, *1*, 69-118.

Sakai, H. (1994). Complex NP constraint and case-conversions in Japanese. In Nakamura, M. (ed.), *Current topics in English and Japanese*, Hitsuji Shobo, 179-200.

Sawada, K., Zhao, T., Shing, M., Mitsui, K., Kaga, A., Hono, Y., & Mitsuda, K. (2024). Release of pre-trained models for the Japanese language. *arXiv preprint arXiv:2404.01657*.

Shain, C., Meister, C., Pimentel, T., Cotterell, R., & Levy, R. (2024). Large-scale evidence for logarithmic effects of word predictability on reading time. *Proceedings of the National Academy of Sciences*, *121*(10), e2307876121.

Sichel, I. (2018). Anatomy of a Counterexample: Extraction from relative clauses. *Linguistic Inquiry*, *49*(2), 335-78.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302-319.

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., Fukuda, S., Ono, H., & Kluender, R. (2011). Reverse island effects and the backward search for a licensor in multiple wh-questions. *Syntax*, *14*(2), 179-203.

Sprouse, J., Wagers, M., & Phillips, C. (2012). A test of the relation between working-memory capacity and syntactic island effects. *Language*, *88*(1), 82-123.

Sprouse, J., Caponigro, I., Greco, C., & Cecchetto, C. (2016). Experimental syntax and the variation of island effects in English and Italian. *Natural Language & Linguistic Theory*, *34*, 307-344.

Sprouse, J., & Hornstein, N. (2013). *Experimental syntax and island effects*. Cambridge University Press.

Sprouse, J., & Villata, S. (2021). Island Effects. In Goodall, G. (ed.). *The Cambridge Handbook of Experimental Syntax* (Cambridge Handbooks in Language and Linguistics). Cambridge University Press, 227-257.

Steedman, M. J. (1985). Dependency and Coordination in the Grammar of Dutch and English, *Language*, *61*, 523-68.

Stigliano, L., & Xiang, M. (2021). Experimental Evidence on Island Effects in Spanish Relative Clauses. *Probus*, *33*(2), 271-296. https://doi.org/10.1515/prbs-2021-0008

Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and cognitive processes*, *1*(3), 227-245.

Strawson, Peter F., (1950). On referring, *Mind*, *59*, 320-44.

Sutton, R. (2019). The bitter lesson. *Incomplete Ideas (blog)*, *13*(1), 38.

Szabolcsi, A. (2006). Strong vs. weak Islands. In Everaert, M. & van Riemsdijk, H. (eds.), *The Blackwell companion to syntax. vol. 4*. Blackwell.

Szabolcsi, A., & den Dikken, M. (2003): Islands. In Cheng, L. & Sybesma, R. (eds.), *The Second GLOT International State-of-the-Article Book*. Mouton de Gruyter, 213-240.

Szabolcsi, A., & Lohndal, T. (2017). Strong vs. weak islands. In Everaert, M. & van Riemsdijk, H. (eds.), *The Blackwell companion to syntax. vol. 8*. John Wiley and Sons, Inc. http://doi.org/10.1002/9781118358733.wbsyncom008

Takahashi, D. (2002). Determiner raising and scope shift. *Linguistic Inquiry*, *33*, 575-615.

Tanaka, N., & Schwartz, B. D. (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*, 750-763.

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, *35*(3), 454-475.

Vasishth, S., & Nicenboim, B. (2016). Statistical methods for linguistic research: Foundational ideas-Part I. *Language and Linguistics Compass*, *10*(8), 349-369.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Vermeulen, R. (2005). Possessive and adjunct multiple nominative constructions in Japanese. *Lingua*, *115*(10), 1329-1363.

Villata, S., Rizzi, L., & Franck, J. (2016). Intervention effects and relativized minimality: New experimental evidence from graded judgments. *Lingua*, *179*, 76-96. https://doi.org/10.1016/j.lingua.2016.03.004

Vincent, J. W., Sichel, I., & Wagers, M. W. (2022). Extraction from English RCs and cross-linguistic similarities in the environments that facilitate extraction. *Languages*, *7*(2), 117. https://doi.org/10.3390/languages7020117

Warstadt, A., & Bowman, S. R. (2022). What artificial neural networks can tell us about human language acquisition. In *Algebraic structures in natural language* (pp. 17-60). CRC Press.

Whitman, J. (2013). The prehead relative clause problem. *MIT Working Papers in Linguistics*, *67*, 361-380.

Wilcox, E. G., Futrell, R., & Levy, R. (2023). Using computational models to test syntactic learnability. *Linguistic Inquiry*, 1-44.

Wilcox, E. G., Gauthier, J., Hu, J., Qian, P., & Levy, R. (2020). On the predictive power of neural language models for human real-time comprehension behavior. *arXiv preprint arXiv:2006.01912*.

Wilcox, E., Levy, R., Morita, T., & Futrell, R. (2018). What do RNN language models learn about filler-gap dependencies?. *arXiv preprint arXiv:1809.00042*.

Yamashita, H. & Chang, F. (2001). Long-before-short preference in the production of a head final language. *Cognition*, *81*(2). B45–B55. https://doi.org/10.1016/S00100277(01)00121-4.

Yano, M. (2019). On the nature of the discourse effect on extraction in Japanese. *Glossa: a journal of general linguistics, 4*(1). 90. https://doi.org/10.5334/gjgl.822

Yoon, H. J. (2007). Raising of major arguments in Korean, In Davis, W. D., & Dubinsky, S. (eds.), *New Horizon in the Analysis of Control and Raising*, Springer, 71-107.

Yoshida, M., Kazanina, N., Pablos, L., & Sturt, P. (2014). On the origin of islands. *Language, Cognition and Neuroscience*, *29*(7), 761-770. https://doi.org/10.1080/01690965.2013.788196

Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX). https://doi.org/10.17605/OSF.IO/MD832