On the gradience of island effects in Spanish relative clauses: experimental evidence

December 15, 2020

Abstract

Research on islands has been central to linguistic theory since Ross (1967)’s work. Its importance relies on the theoretical consequences islands posit for movement and long distance dependencies. In this paper we aim to explore the contrast between different islands in Spanish relative clauses to reveal whether there is any gradience in the strength of the island effects. In order to tease apart fine-grained contrasts we run an acceptability judgment study based on the factorial definition of island effects (Sprouse et al., 2012), an experimental paradigm that aims to isolate the different factors that can affect the acceptability of a sentence involving island violations.

Keywords: island effects, relative clauses, gradience, Spanish, experimental
1. **Introduction**

Research on islands has been central to linguistic theory since Ross (1967)’s work. Its importance relies on the theoretical consequences islands posit for movement and long distance dependencies. In this paper we report and discuss the findings from an acceptability judgment task that tested five different island structures in Spanish. We explore the contrast between different islands in the context of relative clauses, and the variation of island effects across these different constructions. Our main goal is to reveal which structures are indeed islands in this language and whether there is any evidence of gradience in the strength of island effects. In order to tease apart fine-grained contrasts we run an acceptability judgment study based on the factorial definition of island effects (Sprouse et al., 2012), an experimental paradigm that aims to isolate the different factors that can affect the acceptability of a sentence.

The structure of this paper is as follows. In the reminding of this Section we introduce the previous claims and proposals that motivate our study: in Section 1.1 we provide a brief background on island effects, in Section 1.2 we summarize the main features of the factorial definition of islands, and in Section 1.3 we outline previous findings on islands in Spanish. In Section 2 we describe our study on islands in Spanish and in Section 3 we summarize its results. In Section 4 we discuss our results and conclude.

1.1. **Brief background on island effects**

The processing of long-distance dependencies requires relating two syntactic elements that are not in a local relation with each other. This demands different resources from the parser. For instance, the parser needs to maintain the uninterpreted element in the working memory and it also needs to determine the specific point in the sentence at which that element is interpreted. The relations between the moved phrases and their associated verbs are often referred to as ‘unbounded dependencies’. In English, for instance,
a wh-phrase may be separated from the verb by any number of elements, as shown in (1). In the theoretical literature, these dependencies are called A’-dependencies, and the psycholinguistic literature refers to them as filler-gap dependencies. For illustrative purposes, in all the examples in this paper, the filler will be **bolded**, and the gap will be represented by the underscore ‘__’.

(1)  
   a. **What** did Sonia buy __?  
   b. **What** does Bruno think that Sonia bought __?  
   c. **What** does Bruno think that Ana said that Sonia bought __?

Although wh-phrases can be extracted across multiple embedded clause boundaries, there are a number of syntactic environments where extraction is judged as unacceptable. These syntactic environments are know as islands, a term coined by Ross (1967). In this respect, island sensitivity is often considered a diagnostic of movement. The literature on this topic recognizes different islands, which can be referred to based on the structure that creates them, as in (2a)-(2g), or based on the constraint they violate, as in (2h)-(2i):

(2)  
   a. **Whether**-islands  
      *What** did Sonia wonder [CP whether Bruno bought __]?  
   b. Wh-islands  
      *What** does Sonia ask [CP when Bruno bought __]?  
   c. Complex Noun Phrase (NP) islands  
      *What** did you hear [NP the rumor that Sonia bought __]?  
   d. Subject islands  
      *Who** did [NP pictures of __] fell on Sonia’s head?  
   e. Adjunct islands  
      *What** does Sonia get happy [CP if Bruno buys __]?
f. Relative Clause (RC) islands

*What did Sonia meet \([\text{RC} \text{ the author who wrote } \_ \_]?)?

g. Sentential Subject islands

*Who did \([\text{CP} \text{ that Sonia help } \_ \_] \text{ surprised Bruno?} )\?

h. Coordinate Structure Constraint violations

*What did Sonia buy \([\text{ConjP} \text{ a book and } \_ \_]?)?

i. Left Branch Extraction violations

*How tall did Sonia meet \([\text{NP} \_ \_ \text{ a woman}]?)?

Importantly, in addition to wh-dependencies, island effects are observed in other constructions that involve movement such as relative clauses (3a), topicalizations (3b), and scrambling (in languages that allow it, such as Japanese). For this reason, any successful account of the grammar, processing, or learning of island effects must be extended to account for all of them.

(3)  

a. *Sonia likes the car that Bruno wonders \([\text{CP} \text{ whether Ana bought } \_ \_] \).  

b. *Sonia doesn’t know who bought most of these cars, but that car, she wonders \([\text{CP} \text{ whether Ana bought } \_ \_] \).  

(Examples adapted from Sprouse et al., 2016)

Different approaches have been proposed since Ross (1967)’s seminal work. For instance, Chomsky (1973)’s first account of island effects was based on the so-called Subjacency Condition, a constraint on wh-movement operations that required that A’-movement do not cross more than one bounding node, which were defined in terms of a list subject to crosslinguistic variation. Then, Chomsky (1986) redefined bounding nodes as barriers, which were considered ‘blocking categories’. This account proposed that the filler and its gap shouldn’t be separated by more than one barrier. Later, Chomsky (1999) recasted his theory in terms of phases, and proposed the Phase Impenetrability Condition which claims that only the head and the edge of a phrase are accessible
to syntactic operations (phases include vP, CP, and DP). Many other authors, within different theoretical backgrounds, have attempted to explain islandhood and have proposed different theories to account for island effects crosslinguistically; for reasons of space, we will not summarize all this literature here.\footnote{For an overview on this topic, the reader can refer to Boeckx (2012) and references therein.}

What is of interest for this paper are the distinctions that have been proposed to classify different islands. For instance, it has been claimed that islands come in two different ‘varieties’ (see e.g. Cinque, 1990; Postal, 1998; Szabolcsi, 2006). Descriptively, \textit{strong islands} are domains in which no extraction can take place, and \textit{weak islands} are domains from which some phrases, but not others, can be extracted. In this respect, the distinction between strong and weak islands can be understood as the distinction between \textit{absolute} and \textit{selective} islands. This is illustrated in (4). As these examples show, arguments can be extracted from a \textit{whether} clause (4a), but adjuncts cannot (4b):

\begin{enumerate}[\setlength\itemsep{0pt}]  
  \item Which topic did John ask [whether to talk about __]? \label{ex:4a}
  \item *How did John ask [whether to behave __]? \label{ex:4b}
\end{enumerate}

(Examples adapted from Szabolcsi, 2006)

Related to the strong/weak distinction, some authors (i.a. Cinque, 1990; Boeckx, 2012) have drawn yet another distinction regarding how strong the island effect is ‘felt’. This is shown in the following examples of \textit{whether} islands; according to this author, extraction of an argument gives rise to a mild degradation, represented by ‘??’; but extraction of an adjunct gives rise to ungrammaticality:

\begin{enumerate}[\setlength\itemsep{0pt}]  
  \item ??Which student do you wonder whether Bruno praised __? \label{ex:5a}
  \item *How do you wonder whether Bruno praised Sue __? \label{ex:5b}
\end{enumerate}

(Examples adapted from Boeckx, 2012)

This variation in the strength of the island effect is what we focus on in this paper. It’s worth noting that this contrast is not completely straightforward, depending on a
number of factors. Some of these factors are: the distinction between D-linked and non-D-linked wh-phrases (see, e.g. Pesetsky (1987), and subsequent work), the argument/adjunct distinction mentioned above, the type of island involved, and the language under study, among others. What is more, this distinction between different degrees of deviance has important consequences for a theory of islands (and for the overall architecture of grammar in general), and in particular, it has led some authors to propose different explanations for islandhood in each case. For instance, according to Rizzi (1990)’s Relativized Minimality analysis, a milder deviance is analyzed as an intervention effect, that is, the relation between a filler and its gap is interrupted by the intervention of a constituent of the same/similar type. In addition, to explain why some configurations give rise to a strong deviance in one language but only to a mild deviance in another language Rizzi claimed that there were parametric differences between languages, and that the relevant syntactic constraints were ranked differently across languages.

Besides grammatical explanations, like the ones summarized above, which assume that islands arise as a consequence of grammatical constraints, processing explanations have also been put forth. These proposals claim that islands arise as a consequence of certain constraints of the parser (see, e.g. Kluender and Kutas, 1993; Kluender and Gieselman, 2013; Hofmeister and Sag, 2010; Hofmeister et al., 2013). Sprouse et al. (2012) refer to purely processing approaches as reductionist explanations. Crucially, to be able to tease apart these two approaches, rigorous testing needs to be done. This implies collecting data in a systematic and rigorous way, using a methodology that makes it possible to isolate the different factors that can affect the acceptability of a sentence.

1.2. Testing islandhood with a factorial design

Many theoretical proposals regarding the origin of islands were based on informal judgments, often times lacking systematic comparisons between different structures. Addi-
tionally, some previous studies only consider the deviance of the island itself, informally comparing different islands without taking into consideration the base-line, non-island condition, which might show a difference to begin with. However, it has been shown that controlled experimental studies have the advantage of capturing fine-grained distinctions between different types of sentences. For example, Sprouse et al. (2012) have created an experimental paradigm that aims to isolate the different factors that can affect the acceptability of a sentence—i.e., the distance between the antecedent and the gap (‘length’) and the structural complexity of the sentence containing the dependency (‘structure’), controlling for possible confounding factors that might obscure the judgments. This experimental paradigm has been applied to different studies in different languages such as Brazilian Portuguese (Almeida, 2014), English (Sprouse et al., 2012, 2016; Michel, 2014), Hebrew (Keshev and Meltzer-Asscher, 2019), English and Italian (Sprouse et al., 2016), Japanese (Fukuda and Sprouse, 2019), L1 and L2 English and Spanish (Ortega-Santos et al., 2018), Norwegian (Kush et al., 2018, 2019), Spanish (López Sancio, 2015; Pañeda et al., 2020), Slovenian (Stepanov et al., 2018), among others.

In particular, this paradigm consists on manipulating different factors that can affect the acceptability of a sentences: the distance between the filler and the gap, and the structural complexity of the sentence containing the dependency. Each factor has two levels: SHORT and LONG for the former factor, and ISLAND and NON-ISLAND for the latter factor. This gives rise to a 2×2 experimental design. An example is given in (6) for a whether-island, sentences are presented without grammatical judgments:

(6)  
  a. **Who __** thinks [that Sonia bought a car]?  
      NON-ISLAND, SHORT 
  
  b. **What** does Bruno think [that Sonia bought __]?  
      NON-ISLAND, LONG 
  
  c. **Who __** wonders [whether Sonia bought a car]?  
      ISLAND, SHORT 
  
  d. **What** does Bruno wonder [whether Sonia bought __]?  
      ISLAND, LONG 

(Examples adapted from Sprouse et al., 2012)
In this design, the first condition (6a) is considered the baseline, as its the combination of the two ‘unmarked’ conditions. Conditions (6b)-(6c) manipulate only one factor each. The fourth condition (6d) is the combination of the marked conditions in both factors. Using this type of design, two effects can be obtained. The first one is simply linear additivity, which is the linear sum of individual costs. The second one is supper-additivity, which is a combined effect that is greater than the linear sum of individual costs, in other words: \((6b - 6a) + (6c - 6a) > (6d - 6a)\). A supperadditivy effect is evidenced by an interaction between distance and structural complexity. Additionally, a Differences-In-Differences (DD) score is calculated to assess the strength of the supperadditive interaction between those two factors (see Section 3 for details on how to calculate the DD scores). That is, a positive DD score reflects a superadditive interaction, and the larger the DD score, the larger the interaction, which is interpreted as a stronger island effect. On the contrary, a DD score of zero represents no interaction at all, which is interpreted as no island effect.

The current study applies this paradigm to Spanish relative clauses, a language in which not many experimental studies have been conducted, specially in the context of relative clauses. We turn in the next section to briefly summarize the previous theoretical claims and experimental findings on islands in Spanish.

### 1.3. Islands in Spanish

Research on islands in Spanish (see, e.g. Torrego, 1984; Suñer, 1991; Gallego et al., 2007; Jiménez Fernández, 2009; Gallego, 2011; Haegeman, 2014) was mostly based on informal judgments, which gave rise to a lot of variation with regard to the examples used and the judgments reported. Most of the literature on this topic in this language analyze Subject islands, discussing Torrego (1984)’s original observation that wh-movement out of subject in Spanish is possible and interacts with the subject-inversion rule in this language. As for the discussion of other islands, Torrego (1984) and Suñer (1991) claim that extractions out of indirect questions are possible in this
On the gradience of island effects in Spanish relative clauses

language (contrasting with languages like English), as the following examples show
(judgments are the author’s):

(7) a. Quién no sabes cuánto pesa __?
    who   not you.know how.much weights
    ‘Who don’t you know how much weights?’

    b. Quién no sabes qué es __ en esta empresa?
    who   not you.know what is   in this firm
    ‘Who don’t you know what is in this firm?’

    (Examples adapted from Torrego, 1984)

What is more, this author claims that extraction out of non-selected question or a
whether-clause, are also possible:

(8) a. Qué no te explicas por qué Juan habrá comprado __?
    what not you understand why   Juan will.have bought
    ‘What don’t you understand why Juan will have bought __?’

    b. Qué diccionario no sabías si Celia había devuelto __?
    what dictionary   not you.know if Celia had returned
    ‘What dictionary didn’t you know whether Celia had returned __?’

    (Examples adapted from Torrego, 1984)

Additionally, Fábregas (2013) claims that relativization from a whether-clause is
possible, and relativization of certain adjuncts, like temporal adjunct is also possible,
but relativization of cause are not (although this author claim that these gaps are in fact
pro):

(9) a. el juguete que Ana se pondría contenta si su padre le compra __
    the toy   that Ana SE would.get happy if her father to.her buys
    ‘the toy that Ana would get happy if her father buys for her’

    b. el tipo que los vecinos se enfadaron cuando __ tocó la guitarra
    the guy   that the neighbours got.mad when   played the guitar
    ‘the guy that the neighbours got mad when he played the guitar’

    c. *el tipo que María se puso contenta porque __ vino
    the guy   that María SE got happy because   came
    ‘the guy that María got happy because he came’

    (Examples adapted from Fábregas, 2013)
With respect to experimental studies in this language, we summarize here three previous studies. In the first place, López Sancio (2015) followed a factorial design which tested four islands with both wh-extractions and extractions out of relative clauses with extracted PPs. An example is provided in (10) (judgment based on the study’s result). The islands tested in this study were wh-islands, complex NP islands, subject islands and adjunct islands. It’s worth noting that this classification correspond to different islands to the labels used in the current study (see below). López Sancio found island effect for all the wh-dependencies tested, but for the relative-clause dependencies, island effects were only found for wh-, complex NP and adjunct islands, and not for subject islands.

(10) *Aurora adora a la cantante con la que Patri se pregunta cuándo se casará el rapero __.

Int: ‘Aurora loves the singer that Patri wonders when the rapper will marry __.’

(Examples adapted from López Sancio, 2015)

Another study, by Ortega-Santos et al. (2018), also applied the factorial definition of islands and tested extraction of wh-phrases out of embedded interrogative clauses (i.e. wh-islands), as shown in in (11) (judgment based on the study’s result). These authors tested L1 Spanish and English speakers, and L2 Spanish and English speakers as well. With regard to L1 Spanish speakers, they found that these constructions show indeed island effects, contra previous claims in the literature, such as the ones mentioned above.

(11) *Quién no sabes por qué __ escribió el informe?

Int: ‘Who don’t you know why __ wrote the report?’

Finally, Pañeda et al. (2020) examined islands in Spanish using a speeded acceptability task. They tested wh-extractions out of four islands (subject islands, complex
On the gradience of island effects in Spanish relative clauses

NP islands, adjunct islands and *wh*-islands). The authors tested these structures using a word-by-word presentation procedure followed by with a forced-choice task (i.e. participants had to choose between *acceptable* and *unacceptable*). Using a Bayesian framework, their study found that all four constructions tested display a superaditive effect. They further analyzed the strength of the island effect for each construction, and found that it varied in a way that complex NP islands showed the weakest effect, subject islands showed the strongest effect, and *wh*-islands and adjunct islands showed an intermediate effect.

2. The current study

The current study tested six different constructions in Spanish, which included a non-island structure and five different islands. We tested extractions out of relative clauses, which differs from Ortega-Santos et al. (2018) and Pañeda et al. (2020); although López Sancio (2015) tested relative clause dependencies, our study differs from the constructions López Sancio employed. Our main goal was to compare five different islands, some of them have not been previously tested in the same experiment and on the same participant population before, such as *wh-* and *whether*-islands. We aimed to reveal which structures give rise to island effects in Spanish by carrying out a controlled study. In addition, we aimed to tease apart fine-grained distinctions to reveal any gradability in the judgments obtained for different islands. An acceptability judgment experiment was designed following the factorial definition of island effects (Sprouse et al., 2012). The factorial design incorporated two factors that are known to affect the acceptability of a sentence independently of grammatical constraints: the distance between the antecedent and the gap (‘length’) and the structural complexity of the sentence containing the dependency (‘structure’). We tested five construction types and two extraction types for relative clauses (see experimental design and sample stimuli below).
2.1. Participants

80 participants were recruited on social media. 3 participants were removed since they reported to be bilingual speakers (2 participants reported being Spanish-English bilinguals and 1 participant reported being a Spanish-French bilingual). The remaining 77 participants were native speakers of Spanish: 58 participants from Buenos Aires, Argentina; 11 participants from other Argentinian provinces, 4 participants from Spain, 6 participants from other Latin American countries (2 participants from Mexico, 2 participants from Ecuador, 1 participant from Chile, and 1 participant from Venezuela). Furthermore, 2 participants were excluded due to their failure in following the instructions. The analysis was carried over the data from the remaining 75 participants, which had a mean age of 30.85 years old (range: 18-57 years). 57 participants were self-identified as female, and 24 were self-identified as male. All participants provided their consent to participate in the study. As compensation, each participant received a $5 gift-card.

2.2. Materials

Experimental items consisted of 36 sentence sets. There were twelve conditions per item, and each participant saw three items of each condition. We tested six different construction types (non-islands, *wh*-islands, *whether* islands, complex NP islands, adjunct islands and relative clause islands) and two lengths (short and long), which resulted in a $6 \times 2$ design (twelve conditions). Our experimental design differed from previous studies in that we compared all island structures to the same baseline non-island condition. This decision was based on the preference to have the same baseline for comparison. As we will show below, each condition minimally differ from the baseline and from other conditions, which allowed us to carry the desired comparisons between these structures. Experimental items were intermixed with a set of 36 fillers that contained an equal number of acceptable and unacceptable sentences. The order of
presentation of fillers and experimental trials was randomized on a by-participant basis and experimental items were distributed across Latin square lists, such that each participant only saw one condition of each experimental item. Each participant rated a total of 72 items (36 experimental and 36 fillers), and the task took around 20 minutes. Sample stimuli are shown below for each construction type (examples are provided without a grammatical judgment). All the sentences were relative clauses, where the direct object of the main verb was the (extracted) relativized argument. All the extracted DPs were human. In the examples below, the examples in (a) are the short conditions for each structure, and the examples in (b) are the long condition for each structure. Following previous studies, the short condition was generated by extracting the subject of the first embedded verb, and the long condition was generated by extracting the object of the second embedded verb. The non-island condition (12) was the baseline of comparison for all the other island structures and involved a bridge verb (i.e. escuchar ‘to hear’) as first embedded verb:

(12) Non-island

a. Sonia vio al **profesor** que __ escuchó que Bruno contrató a
   Sonia saw DOM.the professor that heard that Bruno hired DOM
   Ana.
   Ana
   ‘Sonia saw the professor that heard that Bruno hired Ana.’

b. Sonia vio al **profesor** que Ana escuchó que Bruno contrató __.
   Sonia saw DOM.the professor that Ana heard that Bruno hired
   ‘Sonia saw the professor that Ana heard that Bruno hired.’

In order to create a whether-island (13), the first embedded verb was changed to preguntar ‘to ask’, followed by the complementizer si ‘if/whether’:

(13) **Whether** island

a. Sonia vio al **profesor** que __ preguntó si Bruno contrató a
   Sonia saw DOM.the professor that asked if Bruno hired DOM
   Ana.
   Ana
   ‘Sonia saw the professor that asked whether Bruno hired Ana.’
b. Sonia vio **al profesor** que Ana preguntó si Bruno contrató __.
   ‘Sonia saw DOM.the professor that Ana asked if Bruno hired
   ‘Sonia saw the professor that Ana asked whether Bruno hired.’

As for **wh**-islands, the first embedded verb was also **preguntar** ‘to ask’, but in these cases it was followed by a **wh**-word **quién** ‘who’, which was the subject of the second embedded verb:

(14) **Wh**-island

   a. Sonia vio **al profesor** que __ preguntó quién contrató a Ana.
   ‘Sonia saw DOM.the professor that asked who hired DOM Ana
   ‘Sonia saw the professor that asked who hired Ana.’

   b. Sonia vio **al profesor** que Ana preguntó quién contrató __.
   ‘Sonia saw DOM.the professor that Ana asked who hired __.
   ‘Sonia saw the professor that Ana asked who hired.’

With respect to complex NP islands (15), the verb **escuchar** ‘to hear’ was used, as in the non-island condition, but it was followed by **el rumor de que** ‘the rumor that’, which was the NP from which the argument was extracted:

(15) **Complex NP island**

   a. Sonia vio **al profesor** que __ escuchó el rumor de que Bruno contrató a Ana.
   ‘Sonia saw DOM.the professor that heard the rumor of that Bruno hired DOM Ana
   ‘Sonia saw the professor that heard the rumor that Sonia hired Ana.’

   b. Sonia vio **al profesor** que Ana escuchó el rumor de que Bruno contrató __.
   ‘Sonia saw DOM.the professor that Ana heard the rumor of that Bruno hired __.
   ‘Sonia saw the professor that Ana heard the rumor that Bruno hired.’

To create adjunct islands we used an intransitive verb as the first embedded (**alegrarse** ‘get.happy’, in this case), followed by a temporal adjunct headed by **cuando** ‘when’.

The extracted argument was extracted from that adjunct:

(16) **Adjunct island**
a. Sonia vio al profesor que ___ se alegró cuando Bruno contrató a Ana.
   ‘Sonia saw the professor that got happy when Bruno hired Ana.’

b. Sonia vio al profesor que Ana se alegró cuando Bruno contrató a D. renunciaron ayer.
   ‘Sonia saw the professor that Ana got happy when Bruno hired D. resigned yesterday.’

To create relative clause islands we relativized the subject of the second embedded verb; this relative clause was the one from which the filler was extracted:

(17) Relative Clause island

a. Sonia vio al profesor que ___ escuchó que los empleados que contrataron a D. renunciaron ayer.
   ‘Sonia saw the professor that heard that the employees that hired D. resigned yesterday.’

b. Sonia vio al profesor que D. escuchó que los empleados que contrataron renunciaron ayer ___.
   ‘Sonia saw the professor that Ana heard that the employees that hired resigned yesterday.’

2.3. Procedure

The acceptability judgment task was run on IbexFarm (Drummond, 2013), a web-based, online tool designed for linguistics studies. Each participant used their own computer. Participants first signed a consent formed and completed a short demographic questionnaire. Then, they rated the acceptability of a set of sentences on a 1-7 scale, with 1 indicating that the sentence was completely unacceptable, and 7 indicating that it was completely acceptable. Participants were instructed to base their judgments on their intuitions as native speakers of Spanish, and not to base their judgments on any prescriptive rules or the plausibility of the situations described. After reading the instructions,
participants were given three practice trials. Participants were allowed to take one or two breaks during the task. Participants saw one sentence at a time, providing their rating either by entering the number with the keyboard or by clicking on the number on the screen. Once they chose a number, the sentence disappeared and the next sentence was displayed.

2.4. Analysis

The analysis was carried using the R software (R Core Team, 2013). First, we fitted mixed-effects regression models on the acceptability ratings using the lme4 package (Baayen et al., 2008). The critical dependent measure was the acceptability response given for each sentence, which range from 1 (completely unacceptable) to 7 (completely acceptable). Before data analysis, we transformed each individual’s raw acceptability ratings into z-scores. Each island type was analyzed separately against the non-island condition; this resulted in five 2×2 models (one for each island). For each model, fixed effects were ‘length’ (short vs. long), ‘structure’ (non-island vs. island), and their interaction. We also included by-participant and by-item random intercepts. Random slopes were added if the resulting models successfully converged. According to the current design, island effects should emerge as a significant interaction between ‘length’ and ‘structure’ (visually, a significant interaction can be identified by non-parallel lines).

3. Results

Mean acceptability ratings by condition (including grammatical and ungrammatical fillers) are reported in Table 1. Figure 1 show plots by island type, displaying mean normalized ratings (z-scores). Each island is compared to the baseline, non-island conditions. Visually, a linear additive effect (i.e. absence of an island effect) is represented by two parallel lines; on the contrary, a superadditive effect is represented by two non-parallel lines. As can be observed in the five plots in Figure 1, all islands show suppre-
On the gradience of island effects in Spanish relative clauses

<table>
<thead>
<tr>
<th></th>
<th>SHORT</th>
<th>LONG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-islands</td>
<td>5.41 (0.138)</td>
<td>4.63 (0.142)</td>
</tr>
<tr>
<td>Adjunct islands</td>
<td>6.13 (0.106)</td>
<td>2.41 (0.118)</td>
</tr>
<tr>
<td>Wh-islands</td>
<td>6.15 (0.107)</td>
<td>2.91 (0.132)</td>
</tr>
<tr>
<td>Complex NP islands</td>
<td>5.80 (0.123)</td>
<td>2.12 (0.099)</td>
</tr>
<tr>
<td>Rel Clause islands</td>
<td>4.74 (0.158)</td>
<td>1.68 (0.087)</td>
</tr>
<tr>
<td>Whether islands</td>
<td>6 (0.112)</td>
<td>3.48 (0.141)</td>
</tr>
<tr>
<td>Grammatical fillers</td>
<td>6.78 (0.022)</td>
<td></td>
</tr>
<tr>
<td>Ungrammatical fillers</td>
<td>1.74 (0.038)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Mean acceptability rating by condition. SE shown in parenthesis.

 additive effects. Furthermore, we ran linear-mixed effects models on z-scores, including ‘Length’ and ‘Structure’ as fixed factors, and by-participant and by-item as random intercepts.\(^2\) Overall, we found a significant effect of Structure (p<0.0001), a significant effect of Length (p<0.0001), and a significant interaction (p<0.0001). We also ran a separate model\(^3\) for each island type, comparing it with the baseline, non-island conditions. To do this, we subset the data from each island type from the overall data set to create data sets that only contained one island type and the baseline, non-island conditions (i.e. four conditions for each). For all island types we found a significant effect of Structure (p<0.0001 for all islands, except for whether islands, with p<0.05), a significant effect of Length (p<0.0001), and a significant interaction (p<0.0001).

Following Sprouse et al. (2012), we measured the strength of islands effects with Differences-in-Differences (DD) scores (Maxwell et al., 2017). The DD scores is a measure that captures the strength of the superadditive interaction. To calculate it, first, we calculated the difference (D1) between the NON-ISLAND/LONG condition and the ISLAND/LONG condition. Second, we calculated the difference (D2) between the scores in the other two conditions (i.e. NON-ISLAND/SHORT and ISLAND/SUBJECT). Finally, we calculated the difference between these two difference scores. According to Sprouse et al., a positive DD score reflects a super additive interaction, and the larger the DD, the

\(^2\)lmer(zscore ∼ Structure*Length + (1|Participant) + (1|Item), data)
\(^3\)lmer(zscore ∼ Structure*Length + (1|Participant), data). More complex models (e.g. including item as random intercept) did not converge successfully.
Figure 1: Plots by Island Type. Error bars correspond to ± 1 SE.
On the gradience of island effects in Spanish relative clauses

<table>
<thead>
<tr>
<th>DD score</th>
</tr>
</thead>
<tbody>
<tr>
<td>DD Adjunct islands</td>
</tr>
<tr>
<td>DD Complex NP islands</td>
</tr>
<tr>
<td>DD Wh-islands</td>
</tr>
<tr>
<td>DD Relative Clause islands</td>
</tr>
<tr>
<td>DD Whether islands</td>
</tr>
</tbody>
</table>

Table 2: DD scores

larger the interactions, which is interpreted as a stronger island effect; on the contrary, a DD score of zero represents no interaction at all (i.e. no island effect). Our DD scores are summarized in Table 2, and show that there is a superadditive effect in all constructions, henceforth, we can conclude that all constructions are indeed islands.

In addition, we were interested in testing whether there was a difference in the strength of the island effect across different constructions, i.e. whether some islands would give rise to a stronger effect. That is, we were interested in testing if some kind of gradability in the island effects can be supported statistically. We first ran a model only on the five island constructions (excluding the baseline, non-island condition).4 We found a significant effect of Structure (p<0.0001), a significant effect of Length (p<0.0001), and a significant interaction (p<0.0001). These results indicate that there is indeed a significant difference in the strength of the island effect for the different island types tested here. Furthermore, we ran a statistical analysis on DD scores,5 which showed that there is a main effect on DD type (p<0.0001), namely, there seems to be a difference regarding the strength of the island effect across the different constructions tested. Finally, we ran a pair comparison on DD scores and found that there was a significant difference between whether islands and three other islands: adjunct islands (p<0.001), complex NP islands (p<0.001), and wh-islands (p<0.05), but not relative clause islands (p>0.1). All the other islands didn’t show a significant difference compared to each other. From this data, we conclude that there is a difference regarding the

4lmer(zscore ∼ Structure*Length + (1|Participant) + (1|Item), data)
5lmer(meanzscore ∼ DDType + (1|Participant), data) More complex models (e.g. including Item as random intercept) did not converge successfully.
strength of island types overall, but this is a difference triggered by only one type of islands, that is *whether*-islands.

4. Discussion & Conclusions

Our study tested six constructions in Spanish. They all involved extractions out of a relative clause, and one of them was considered the baseline, non-island condition. We aimed to show whether the other five constructions show island effects and whether there was a difference in the strength of this effect. Our results show that the five constructions tested exhibit superadditive effects, which means that the five constructions should be considered islands in Spanish, under a factorial definition of islands. With regard to *wh*- and *whether* islands, our results challenge previous claims in the theoretical literature (see e.g. Torrego, 1984; Suñer, 1998; Fábregas, 2013) who argue that those constructions do not trigger island effects.

The results obtained in our study are only partially in line with previous experimental studies in this language (in particular, López Sancio 2015 and Pañeda et al. 2020). First, in line with López Sancio (2015), we also found an interaction between ‘length’ and ‘structure’ for extractions out of *wh*-clauses, complex NPs, and adjuncts show, which indicates that those structures should be considered islands in this language. However, we differ from López Sancio (2015) in the size of the effect, as will be discussed below. In addition, Pañeda et al. (2020) found an interaction for extractions out of adjuncts and *whether* clauses, but not for complex NPs; however, it’s worth noticing that they tested extraction out of *wh*-questions, not relative clauses, as we did.

The discrepancies between different studies call for more rigorous and systematic experimental investigations and emphasize the prominent place that formal experimental work deserves in linguistic theory. An important difference between the current study and previous studies is that we used the same baseline condition for all islands, and that we minimally modified this condition for each island, which allowed us, on the
On the gradience of island effects in Spanish relative clauses

one hand, to have a more accurate standard of comparison, and on the other hand, to
carry out a comparison between the different islands tested.

With regard to extractions out of relative clauses, which were only tested in López Sancio (2015), who found a DD score of 2.11 for complex NP islands, whereas our DD score was 1.13. This led to a different conclusion with regard to the strength of the island effects. Upon a closer look, however, there is the stimuli tested in López Sancio (2015) and the stimuli tested in the current study. For example, the island and non-island conditions from López Sancio (2015) are not minimal pairs, as shown in (18) below. Similar issues could be raised for the other island conditions tested in López Sancio (2015).

(18) a. Conozco al entrenador con el que el jugador oyó los rumores sobre que Ana está saliendo __. I.know the trainer with that the player heard the rumors about that Ana is going out __.
   ‘I know the trainer with whom the player heard the rumors that Ana was going out __.’

b. Conozco al entrenador con el que dice el jugador que sale Ana __. I.know the trainer with that says the player that goes out Ana __.
   ‘I know the trainer with whom the player says that Ana goes out __.’

(Examples adapted from López Sancio, 2015)

On the other hand, our stimuli compared minimal pairs, as the examples in (12b)-(15b), repeated below in (19):

(19) a. Sonia vio al profesor que Ana escuchó que Bruno contrató __. Sonia saw DOM the professor that Ana heard that Bruno hired
   ‘Sonia saw the professor that Ana heard that Bruno hired.’

b. Sonia vio al profesor que Ana escuchó el rumor de que Bruno hired
   Sonia saw DOM the professor that Ana heard the rumor of that Bruno
   hired
   ‘Sonia saw the professor that Ana heard the rumor that Bruno hired.’
In addition, based on the analysis of DD scores, we found some limited differences in the size of island effects. In particular, we only found that *whether*-islands are different from adjunct, complex NP and *wh*-islands, but not relative clause islands. No other significant differences were found with regard to the strength of the island effects. Since the majority of the constructions we tested show island effects with similar strength, our findings seem to present a more categorical picture of island effects than some previous studies have suggested. More future research is clearly needed. But it is worth pointing out that in the current study, the judgments on the island constructions are always compared with their matched baseline conditions. This is important because a direct comparison between different island constructions may lead to erroneous conclusions about island strength. For instance, while relative clause islands give rise to the lower acceptability judgments, the strength island effect (measured by the DD scores and further confirmed by a pair comparisons) is not greater than the other islands. We claim that the ‘illusion’ that these islands show a strong degradation is due to the fact that there is already bigger degradation of the non-island condition to begin with. However, when careful statistical comparisons with a baseline are carried out, it’s possible to observe that the islands that showed the mildest degradation (i.e. *whether* islands, with a rating of 3.48) and the islands that shows the strongest degradation (i.e. relative clause islands, with a rating of 1.68) don’t actually differ in the strength of the island effect.

In conclusion, we explored the contrast between different constructions in extractions out of relative clauses in Spanish. We showed that all the constructions considered show a significant interaction between ‘length’ and ‘structure’, hence, they can all be considered islands in this language. Crucially, we argue that the differences between our results and previous claims in the theoretical literature is due to a lack of a proper baseline condition. We furthermore analyzed whether there is any effect of gradience across constructions, but we failed to find an across-the-board gradient effect. We found, however, some limited differences in the size of the effects between one type of islands (i.e. *whether* islands) and three other islands (adjunct, complex NP and *wh*-islands).
On the gradience of island effects in Spanish relative clauses

Crucially, these limited differences in the size of the effect do not correlate with the difference in the strength of the degradation (i.e. while whether islands and relative clause islands show the smallest and the biggest degradation respectively, they do not differ statistically in the size of the effect). Given the overall similarity in the strength of the island effects, our findings seem to point to a more categorical picture of island effects.

References


Fábregas, Antonio. 2013. Nota sobre unas construcciones que temblaría la gramática si fueran extracciones de isla. Signo y seña 175–188.

Fukuda, Shin, and Jon Sprouse. 2019. Islandhood of Japanese complex NPs and the factorial definition of island effects.

Gallego, Ángel J, Juan Uriagereka, José Camacho, Nydia Flores-Ferrán, Liliana Sánchez, Viviane Déprez, and María José Cabrera. 2007. Sub-extraction from subjects: A phase theory account. *Amsterdam studies in the theory and history of linguistic science series 4* 287:155.


