Isogloss
Open Journal of Romance Linguistics

Received: 30-09-2024
Accepted: 05-02-2025
Published: 14-07-2025

Comparing two experimental designs for the study of subject islands in Spanish

Laura Stigliano
The Ohio State University
stigliano.4@osu.edu

Matías Verdecchia
IIF-SADAF-CONICET / Universidad de Buenos Aires
mnverdecchia@filo.uba.ar

Marisol Murujosa
Instituto de Lingüística, Facultad de Filosofía y Letras (UBA) / CONICET
marisolmurujosa@conicet.gov.ar

How to cite: Stigliano, Laura, Verdecchia, Matías, & Marisol Murujosa. 2025. Comparing two experimental designs for the study of subject islands in Spanish. RLLT 25, eds. Adam McBride, James Law & Willis Fails. Special Issue of *Isogloss. Open Journal of Romance Linguistics* 11(5)/7, 1–28.
DOI: https://doi.org/10.5565/rev/isogloss.510

Abstract

This paper investigates the phenomenon of subject islands in Spanish through the lens of two experimental designs, which we refer to as the *subject/object design* and the *simple/complex subject design*. Our study is the first one to directly compare these proposed designs in a single, controlled experiment. By comparing these methodologies, we aim to evaluate their effectiveness in capturing the nuances of subject island effects to establish the basis for future investigations in this area. Our findings reveal that both designs yield a significant degradation in acceptability for extractions from subjects, which is consistent with theoretical proposals that aimed to account for subject islands. We question, however, the assumption that a significant interaction, and by extension super-additivity, is the sole criterion for identifying island structures, given that the different factors tested (i.e., the position of the gap, the object/subject distinction, and the extraction/sub-extraction distinction) did not contribute to the degradation observed in subject islands.

Keywords: island, subject island, subject, experimental syntax, Spanish.

1. Introduction

Since Sprouse's (2007) seminal work, factorial designs for measuring island effects have gained widespread acceptance in experimental research (see Sprouse 2023 for an updated overview). While this methodology has been consistently applied to examine various island domains (e.g., relative clause islands, adjunct islands, or wh-islands), the specific design for investigating subject islands remains a topic of ongoing debate. Recently, two different designs have been proposed: on the one hand, what we will refer to as the *subject/object design* (SOD; Sprouse et al. 2012: Experiment 1, Sprouse et al. 2016)—which compares the effect of (sub)extractions in subjects vs. objects; on the other hand, what we will refer to as the *simple/complex subject design* (SCSD; Sprouse et al. 2012: Experiment 2, Kush et al. 2018)—which compares simple and complex subjects from embedded vs. matrix positions. Although both designs aim to isolate a super-additive effect, they differ in the factors they test and the predictions they make.

To evaluate the scope and adequacy of the two designs, we conducted a single, controlled acceptability judgment experiment that examines subject islands in Spanish employing both methodologies. In this paper, we report and discuss the key findings of the study. In short, the results indicate that both designs show a significant decrease in acceptability for sentences involving extractions from subjects (i.e., subject islands). However, contrary to expectations, none of the factors tested plays a significant role in explaining the degradation observed in subject islands.

The structure of the paper is as follows. In Section 2, we provide a brief overview of syntactic approaches to subject islands. In Section 3, we introduce the factorial design for the study of islands, with a special focus on the specific design developed for subject islands and its subsequent modifications (Section 3.1). Additionally, we review the results and key outcomes from two experimental studies on islands in Spanish (Section 3.2). In Section 4, we describe our study, and in Section 5, we present the results. In Section 6, we discuss the main findings of the study, and their relevance to the discussion of subject islands. Finally, in Section 7, we provide some concluding remarks.

2. Subject Islands

The so-called *long-distance dependencies* require linking two syntactic elements—a *filler*, represented in italics, and a *gap*, represented by the underscore—that are not in a local relation with each other:[1]

(1) *What* did Sonia buy ___?

[1] This type of dependencies is also referred to in the literature as *filler-gap* dependencies.

\The relation between the moved wh-phrase (*what*, in (1)) and its associated verb (*buy*, in (1)) is often referred to as an *unbounded dependency*, so called because the wh-phrase can be separated from the verb by any number of intervening elements or clauses, as seen in (2):

(2) a. *What* does Bruno think [that Sonia bought ___]?
b. *What* does Bruno think [that Ana said [that Sonia bought ___]]?
c. *What* does Bruno think [that Ana said [that Pete believes [that Sonia bought ___]]]?

Although wh-phrases can be moved across multiple embedded clause boundaries, there are several syntactic environments where such extraction is judged as strongly unacceptable/ungrammatical. These syntactic environments are known as *islands*, a term coined by Ross (1967). The different types of islands are named after the structure that gives rise to them. For instance, examples (3) and (4) illustrate *interrogative islands* and *complex noun phrase islands*, where extraction crosses the boundary of an interrogative clause (Int-Cl) and a complex noun phrase (CNP), respectively:

(3) *_What_ does Bruno ask [$_{\text{Int-Cl}}$ when Sonia bought ___]?

(4) *_What_ did you hear [$_{\text{CNP}}$ the rumor that Sonia bought ___]?

Research on islands has been central to linguistic theory since Ross's work. The importance of islands lies in their theoretical implications for syntactic movement and long-distance dependencies, as well as for the theories that study the processing of these structures. The constraints islands impose on certain syntactic operations provide valuable insights into the underlying mechanisms that govern human language. In this paper, we study a particular type of syntactic island, the so-called *subject island*, which typically refers to the impossibility of extracting elements from subjects, as illustrated in (5):

(5) a. English, Chomsky (1977:106)
*_Who_ did [$_{\text{Subject}}$ stories about ___] terrify John?
b. English, Postal (1974:189)
*_Who_ did [$_{\text{Subject}}$ pictures of ___] lay on the table?
c. English, Kayne (1981:114)
*_Who_ was [$_{\text{Subject}}$ a picture of ___] lying there?

One of the earliest attempts to account for subject islands was Chomsky's (1973) *Subject Condition*, a syntactic constraint that categorically bans the extraction of elements from subjects:[2]

[2] It is important to note that syntactic accounts are not the only approaches to explaining island effects. The so-called *grammatical* explanations posit the existence of grammatical constraints. In contrast, the so-called *processing* accounts attribute island effects to constraints of the parser (see Hofmeister et al. 2013, Kluender & Gieselman 2013, among others). Additionally, the so-called *discourse-based* explanations (see Abeillé et al. 2020, Winckel &

(6) Subject condition
No rule can involve X, Y in the structure
. . . X . . . [$_{\alpha}$ . . . Y . . .] . . .
where α is a subject phrase properly containing the minimal major category containing Y, and Y is subjacent to X.

This condition was later redefined as a restriction on extractions from elements not in a complement position, which were considered "not selected" (in terms of Cattell 1976 and Cinque 1977, among others), or "not properly governed" (in terms of Huang 1982). In more modern terms, it has been claimed that extractions from elements sitting in the Spec-TP position (such as subjects) were banned for linearization reasons (Uriagereka 1999), or because Spec-TP is a derived position (assuming that subjects' original position is VP-internal, or in modern terms, that subjects externally merge in Spec-*v*P, and subsequently move to Spec-TP), as proposed by the *Freezing Principle* (see, e.g., Wexler & Culicover 1980). In summary, these approaches not only posit a categoric ban on extractions from subjects, but they argue that the main reason for this illicit extraction is the position subjects occupy in the syntactic structure.

However, subsequent research has challenged the notion that subjects are categorically islands for extraction, suggesting a more nuanced view. For instance, Chomsky (2008) noted that wh-extractions seem to be permitted out of passive/unaccusative subjects (7a), but not out of transitive subjects (7b). Crucially in both cases, the subject is in a moved position, specifically Spec-TP, rather than a complement position:

(7) Engish, Chomsky (2008:147)
a. *Of which car* was [$_{\text{Subject}}$ the driver ___] awarded a prize?
b. **Of which car* did [$_{\text{Subject}}$ the driver ___] cause a scandal?

To address the limitations of approaches that relied on categorical bans on extractions from subjects, Haegeman et al. (2014) propose to "deconstruct" Chomsky's Subject Condition. They question the notion of subject islands as a categorical constraint and suggest instead that the (im)possibility of extractions from subjects depends on a range of interacting factors. Their approach builds on an analysis of various factors and draws on introspective judgments and judgments reported in the (theoretical) literature.

To briefly illustrate some of the factors discussed by Haegeman et al., they discuss how thematically prominent arguments, such as agents (8a), tend to block extraction more than less prominent arguments, such as goals (8a):

Abeillé 2020, among others) suggest that island effects arise from the information structure of the sentence, such as focus-background conflicts. Due to space limitations and our focus being on testing experimental designs that target grammatical explanations, we will not delve into the processing and discourse approaches here. For a more comprehensive discussion, we refer readers to the sources cited above, and references therein.

(8) English, Chomsky (2008; as cited in Haegeman et al. 2014:83)
a. *Of which books* did [$_{\text{Subject}}$ the authors ___] receive a prize?
b. **Of which* car did [$_{\text{Subject}}$ the driver ___] cause a scandal? (=7b)

This has been attributed to the "structural integrity" of the extracted constituent, though this explanation may inaccurately predict that extractions from agentive objects in passive constructions should be restricted, which is not supported by the data. Additionally, they discuss the *Specificity Condition* (Fiengo & Higginbotham 1981), which in turn posits that specific DPs are more resistant to extraction, suggesting that all specific DPs, regardless of their syntactic role or position, should be uniformly opaque.

In summary, Haegeman et al. (2014) argue that the extent to which extractions from subjects are degraded depends on both the internal and external properties of subjects, as well as some characteristics of the extracted element. This already complex empirical picture is further complicated when we look at cross-linguistic variation. For example, Starke (2001) observes that languages like French and Italian permit extractions from both pre- and post-verbal subjects, whereas languages like Czech and Slovak restrict extractions to post-verbal subjects only. Similarly, Ordóñez and Treviño (1999) identify the same asymmetry between pre- and post-verbal subjects in Spanish.[3] Haegeman et al.'s proposal, then, shifts the focus from a rigid, universal constraint to a more nuanced understanding of multiple interacting factors, offering new insights into the variability of extractions from subjects. Additionally, they suggest that the relative weight of each factor may influence the acceptability of extractions and that this weighting likely differs across languages. However, this remains an open question, since determining the precise contribution of each factor requires experimental investigations that allow us to uncover fine-grained distinctions and capture cross-linguistic differences. With the present study, we take a first step toward addressing this empirical gap.

It is important to note that all theoretical accounts of island effects discussed so far in this section have relied on introspective judgments, often lacking systematic comparisons across different structures. In contrast, controlled experimental studies are shown to be able to reveal subtle distinctions between various sentence types, offering a more precise understanding of these effects and the phenomena that underlie them (see, e.g., Lewis & Phillips 2015). Furthermore, research on subject islands has uncovered significant inter- and intra-speaker variation (see, e.g., Lu et al. 2024). Experimental studies are essential for capturing this variability, providing objective data collection, and enabling controlled comparisons across structures. They detect fine-grained distinctions that informal judgments might overlook and allow for statistical analysis to validate observed patterns. These studies also offer a robust means of testing theoretical predictions.

In the following sections, we summarize the methods and results of experimental studies on islands, with a focus on subject islands (Section 3.1) and subject islands in Spanish (Section 3.2).

[3] As one anonymous reviewer points out, it is also worth noting that there are scope asymmetries associated with different subject positions in Spanish (see Uribe-Etxebarria (1995) for discussion).

3. Experimental studies on islands

To adjudicate between different accounts of island effects, and, in particular, to subject islands, it is important to define these effects in a way that allows for theory comparisons. A useful definition of island effects is the *relative* definition, which compares the island-violating sentence with a minimally different grammatical sentence (i.e., a minimal pair), as shown below:

(9) a. *What* does Bruno think [$_{\text{Decl-Cl}}$ that Sonia bought ___]?
b. **What* does Bruno ask [$_{\text{Int-Cl}}$ when Sonia bought ___]?

However, just by looking at these two sentences, it is not possible to determine what is the source of unacceptability for the second one. For instance, it could be the case that the unacceptability of (9b) is due to the mere presence of *when* rather than by the location of the gap inside the island structure. To control for this possibility, Sprouse (2007) proposed a specific design to experimentally test islands. In Section 3.1, we describe Sprouse's design for the study of islands, focusing particularly on subject islands, as well as the modifications that have been proposed, and the rationale behind them. In Section 3.2, we briefly summarize the results and key findings obtained in the two experimental studies of islands in Spanish.

3.1. Two experimental designs for subject islands

Sprouse (2007 and subsequent works; see Sprouse et al. 2012, Sprouse et al. 2016, among others) proposes a 2×2 factorial design to measure and quantify island effects.[4] The primary goal of this type of factorial design in the study of islands is to isolate the contribution of the island constraint from two other factors that are present in extractions from island domains and that may independently decrease their acceptability. These factors are, on the one hand, the long-distance dependency between the filler and the gap, and, on the other hand, the island configuration, which arguably involves a complex syntactic structure. To do so, Sprouse's 2×2 design crosses two conditions: STRUCTURE and GAP-POSITION, each with two levels (*non-island/island* and *matrix/embedded*, respectively). This is illustrated in (10) for *whether*-islands:

(10) English Sprouse et al., (2012:86)

a. *Who* __ thinks [that John bought a car]?	*non-island* | *matrix*
b. *What* do you think [that John bought __]?	*non-island* | *embedded*
c. *Who* __ wonders [whether John bought a car]?	*island* | *matrix*
d. *What* do you think [whether John bought __]?	*island* | *embedded*

The main advantage of this design is that it can properly set apart the three effects mentioned above, that is, the GAP-POSITION effect, the STRUCTURE effect, and, most importantly, the island effect. First, the influence of the gap position is quantified

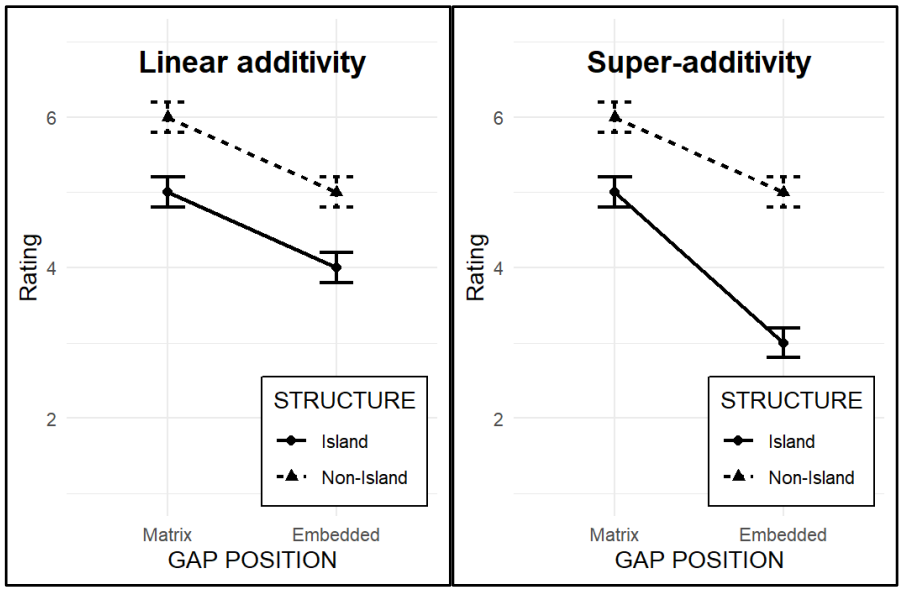
[4] It is worth noting, however, that the use of 2×2 designs to isolate the contribution of specific constraints was already well established and widely utilized (see Cowart 1997).

by comparing (10a) with (10b), i.e., extraction from a matrix clause versus extraction from an embedded clause. Second, the influence of the structural complexity is captured by contrasting (10a) with (10c), i.e., a sentence that contains an island structure with one that doesn't contain such structure. Finally, the island effect is obtained by calculating the total effect, i.e., (10a) – (10d), and subtracting the GAP-POSITION and the STRUCTURE effects. The quantification of the three effects can be summarized in (11), as follows:

(11) a. GAP-POSITION effect = (10a) – (10b)
b. STRUCTURE effect = (10a) – (10c)
c. Island effect = ((10a) – (10d)) – ((10a) – (10b)) – ((10a) – (10c))

According to Sprouse et al. (2012, 2016), an island effect is defined as a significant (negative) interaction, given that (11c) is equivalent to calculating the difference between the effect of gap position in islands versus non-islands (i.e., (10a – 10b) – (10c – 10d)), or, alternatively, the difference between the effect of structure in matrix versus embedded extractions (i.e., (10a – 10c) – (10b – 10d)).[5] In other words, an island effect can be defined as a *super-additive* interaction between STRUCTURE and GAP-POSITION, where the combined effect of these two factors is larger than their linear sum.[6] This can be visually identified in an interaction plot (Figure 1), where parallel lines signal the absence of an island effect (i.e., linear additivity), while nonparallel lines indicate the presence of an island effect (i.e., super-additivity):[7]

Figure 1. Super-additive versus linear additive effects.



[5] See also the so-called *differences-in-differences* (DD) scores (Maxwell & Delaney 2003).

[6] Sprouse et al. (2012) attribute this super-additivity to the presence of an independent grammatical constraint (i.e., the island constraint). However, see Gieselman et al. (2013), Hofmeister et al. (2014), Keshev & Meltzer-Asscher (2019), among others, for an explanation in terms of processing difficulties.

[7] It is important to mention that, although not explicitly stated, this definition of super-additivity as a significant interaction presupposes significant effects of the manipulated factors.

While Sprouse and colleagues adopt this design to test different types of islands (e.g., *whether*-islands, complex NP islands, adjunct islands), they propose a slightly modified version for subject islands (Sprouse et al. 2012: 94, Sprouse et al. 2016: 318). To begin with, the STRUCTURE factor is evaluated by comparing *subjects* and *objects*, which correspond to the *island* and *non-island* conditions, respectively. Regarding the GAP-POSITION factor, rather than contrasting the extraction of matrix and embedded elements, as in the canonical design in (10), Sprouse et al. propose contrasting extraction *of* embedded subjects/objects with extraction *from* embedded subjects/ objects, that is, the TYPE OF EXTRACTION (i.e., *extraction* vs. *subextraction*). Overall, crossing these conditions results in four sentences as in (12) (note that (12d) contains the putative island violating structure). For expository reasons, we refer to this experimental design as the *subject/object design*:

(12) Subject/object design, Sprouse et al. (2012:94)

a. *What* do you think the speech interrupted ___?	*of* | *object*
b. *What* do you think ___ interrupted the TV show?	*of* | *subject*
c. *What* do you think the speech about global warming interrupted [the TV show about ___]?	*from* | *object*
d. *What* do you think [the speech about ___] interrupted the TV show about global warming?	*from* | *subject*

Sprouse and colleagues employ this design in acceptability rating tasks and indeed find the expected super-additive effect (see Sprouse et al. 2012: Experiment 1 and Sprouse et al. 2016). However, the authors themselves raise concerns regarding a potential underestimation of the island effect in this design. In subsequent work, Kush et al. (2018) point out that extractions from objects could give rise to a filled-gap effect after the embedded verb (e.g., in (12c), given that *what* can be posited as a filler after *interrupted*).[8] They argue that since this effect is absent in the other conditions, it may lower the putative subject island penalty. Kush et al. further observe that the fact that both subjects and objects involve complex DPs in (12c) and (12d) decreases the acceptability of these conditions. In consequence, this could produce a floor effect that, once again, reduces the impact of the subject island degradation.

To avoid these confounding factors, an alternative experimental design for subject islands was developed (Sprouse et al 2012: Experiment 2, Kush et al. 2018). Like the original design in (10), this one also isolates the factor GAP-POSITION by contrasting extractions from *matrix* clauses (e.g., (13a) and (13c)) with extractions from *embedded* clauses (e.g., (13b) and (13d)). However, here the factor STRUCTURE varies between *simple-subject* (i.e., embedded subjects without a modifier, as in (13a) and (13b)), and *complex-subject* (i.e., embedded subjects that contain a modifier, as in (13c) and (13d)), taking the latter as the island configuration. For expository reasons, we refer to this experimental design as the *simple/complex subject design*:

[8] Note that this observation only holds for languages that allow P-stranding, like English. In languages that do not permit leaving the preposition stranded after movement, like Spanish, the fronted wh-phrase pied-pipes the preposition. In consequence, the extracted element cannot be interpreted as occupying a potential gap position after the verb (i.e., the presence of the preposition prevents the wh-phrase from functioning as the direct object of the verb). See Section 6 below.

(13) Simple/complex subject design, Sprouse et al. (2012:102)

a. *Who* ___ thinks the speech interrupted the primetime TV show?	*matrix* | *simple*
b. *What* do you think ___ interrupted the primetime TV show?	*embedded* | *simple*
c. *Who* ___ thinks the speech about global warming interrupted the primetime TV show?	*matrix* | *complex*
d. *What* do you think [the speech about ___] interrupted the primetime TV show?	*embedded* | *complex*

Kush et al. (2018) acknowledge that this design is not without problems. One disadvantage that the authors point out is that the sentence that contains the alleged island violation (i.e., (13d)) could also exhibit an independent effect of sub-extraction. Given that this is absent in the other conditions, the design would overestimate the size of the subject island effect. Nevertheless, even if this may obscure the results, they do not consider it a significant shortcoming to rule out the design. In fact, they argue that there is no independent evidence of the existence of a sub-extraction effect, and if it exists, it should be significantly smaller than an island effect. For these reasons, they conclude that it is still better to run the risk of overestimating the effect, rather than underestimating it, as the subject/object design arguably does. The results of their experiments, then, seem to support their proposal: applying this alternative design, both Sprouse et al. (2016: Experiment 2) and Kush et al. (2018) find a super-additive effect in English and Norwegian subject islands, respectively.

3.2. Subject islands in Spanish

Research on islands in Spanish has predominantly relied on introspective judgments, resulting in considerable variation in both the examples provided and the judgments reported (see, e.g., Torrego 1984, Suñer 1991, Gallego & Uriagereka 2007, Jiménez Fernández 2009, Gallego 2011, Haegeman et al. 2014). Only recently have experimental studies begun to examine islands in a more systematic way, using controlled experimental methods that reveal more nuanced differences between constructions. Regarding subject islands, only two studies have specifically investigated these structures: López Sancio (2015) and Pañeda et al. (2020), and each employed one of the two proposed experimental designs described in the previous section: the subject/object design and the simple/complex subject design, respectively. As we will summarize below, there are important differences between the two studies, making direct comparisons between the results obtained difficult. Moreover, these studies examined subject islands in relation to other island types, which further complicates interpreting the results for subject islands alone, since participants might be inadvertently comparing the acceptability of one island against the others. Finally, upon closer examination of the items tested, some confounding factors arise. In the remainder of this section, we outline each study, the types of stimuli used, the issues they raised, and the results obtained.

López Sancio (2015) employed a factorial design to test four types of islands: wh-islands with *por qué* (‘why’), *cuándo* (‘when’), and *cómo* (‘how’), complex NP islands, subject islands, and adjunct islands, across two conditions: wh-extractions (i.e., wh-questions), and extractions out of relative clauses. The study was divided into

four smaller studies, with participants exposed to only two island types: one for the wh-extraction condition and one for the relative clause condition. In the case relevant here, subject islands in wh-dependencies (i.e., wh-questions) were tested using the subject/object design, alongside wh-islands in relative clauses (López Sancio's Study 2). The experiment employed an acceptability judgment task on a 1-to-7 Likert scale. Study 2 included 51 participants, who were monolingual speakers of Peninsular Spanish.

López Sancio's experimental items were modeled after the Italian materials used by Sprouse et al. (2012) (i.e., the subject/object design, as described above), though modifications were made specifically for subject islands. In Sprouse et al. (2012), both subjects and objects were specific. However, as the author notes, and as already pointed out in Section 2 above, specificity has been reported in the literature as a factor that makes DPs more resistant/opaque to extractions (see, e.g., Fiengo & Higginbotham 1981). To address this, López Sancio opted to use non-specific subject DPs. Below, in (14), is an example of the stimuli used in López Sancio's Study 2 for subject islands:

(14) Spanish, López Sancio (2015: ex. (42))

a. *¿De qué revista* crees que [varios redactores ___] han escrito artículos de Lady Gaga?
'*Of which magazine* do you think that [several editors ___] have written articles about Lady Gaga?' — *from | subject*

b. *¿Quiénes* dices que ___ han escrito artículos sobre Lady Gaga?
'*Who* do you say that have ___ written articles about Lady Gaga?' — *of | subject*

c. *¿Sobre quién* dices que ha escrito [artículos __] varios redactores de Vogue?
'*About whom* do you say that some Vogue editors have written [articles about ___]?' — *from | object*

d. *¿Sobre quién* dices que han escrito ___ varios redactores?
'*About whom* do you say that some editors have written ___?' — *of | subject*

Each study tested eight experimental items in a 2×2 design, meaning each participant saw two items per condition. Additionally, each study included eight experimental items from another island type (in this case wh-islands in relative clause extractions) and 32 fillers, resulting in a total of 54 sentences rated. Although there was a 2:1 ratio of fillers to experimental items, there was a 1:1 ratio of acceptable to unacceptable items, assuming three of the four experimental conditions were considered acceptable (conditions (14b), (14c), and (14d)) and one unacceptable (condition (14a)).

The results for the island condition, with a mean rating of 3.39 (SD = 1.69), revealed significant main effects of TYPE OF EXTRACTION (i.e., *of* vs. *from*; $p < 0.001$) and STRUCTURE (i.e., *subject* vs. *object*; $p < 0.001$), as well as a significant interaction

($p < 0.001$).[9] The author interprets these findings as evidence of island effects for subject islands in wh-extractions (it is worth mentioning that such effects were not observed in relative clause extractions).

Upon closer examination of the items, we found that they were not uniformly constructed, which could introduce confounding factors. For instance, three out of the eight items used D-linked wh-phrases, while the remaining five used non-D-linked ones, which is relevant because D-linked wh-phrases are easier to extract from (see, e.g., Pesetsky 1987, Cinque 1990, among many others). More critically, in the island condition (condition (14a)), extractions were made from pre-verbal subjects, which are dispreferred in Spanish, whereas the object conditions (conditions (14c) and (14d)) involved post-verbal subjects, which are preferred in wh-questions in Spanish (Torrego 1984).[10] Moreover, not all item sets were constructed as minimal pairs. For instance, a given set could include some items with the embedding verb *creer* ('to believe'), and some other items with the embedding verb *decir* ('to say'), as the example in (14) showed. Finally, the condition that was supposed to test extractions of objects (i.e., condition (14d) above) included half of the items with wh-DP extractions (i.e., *qué* 'what'), and the other half with wh-PP extractions (e.g., *en quién* 'in who' or *contra quién* 'against who'). Most of these PPs were not selected by the verb, casting doubt on their status as true objects (as opposed to adjuncts) and creating an imbalance, as the subject extractions were always wh-DPs (*quién* 'who.SG' or *quiénes* 'who.PL'). In our design, we aimed to control for these factors.

The other experimental study that tested subject islands in Spanish is Pañeda et al. (2020), which used a speeded acceptability judgment task with a binary response (that is, acceptable/not acceptable). Data from 80 participants was included in the analysis. Each participant rated all island types, with two items per island type (32 items in total, as four island types were tested). Subject islands were based on the design outlined by Sprouse et al. (2012) in their Experiment 2 and Kush et al. (2018), that is, the simple/complex subject design. The embedding verb was *creer* ('to believe'). Embedded subjects were all definite subjects, which is problematic, for the reasons listed above. An example item set for subject islands is provided in (15):

(15) Spanish, (adapted from) Pañeda et al. (2020:ex.(10))[11]

a. ¿*Quién* ___ cree que el discurso ofendió tanto a Julia ayer?
'*Who* ___ believes that the discourse offended Julia so much yesterday?' — *simple* | *matrix*

b. ¿*Quién* ___ cree que el discurso del director ofendió a Julia?
'*Who* ___ believes that the director's discourse offended Julia?' — *complex* | *matrix*

[9] In López Sancio (2015), TYPE OF EXTRACTION is called GAP and STRUCTURE is called TYPE.

[10] In condition (b), the entire subject is extracted. Since both pre- and post-verbal subjects are possible in Spanish, it is unclear whether the extraction originates from a pre-verbal or post-verbal position.

[11] Note that Pañeda et al. (2020) use a different terminology: condition (a) is called *non-island/short*; condition (b) is called *island/short*; condition (c) is called *non-island/long*; and condition (d) *island/long*.

c. ¿*Quién* crees que ___ ofendió tanto a Julia con el discurso?
'*Who* do you believe that offended Julia so much with the discourse?' *simple* | *embedded*

d. ¿*De quién* crees que [el discurso ___] ofendió tanto a Julia]?
'*Of who* do you believe that [the discourse ___] offended Julia so much?' *complex* | *embedded*

Pañeda et al.'s (2020) experimental items were mixed with 48 fillers and 24 items from an unrelated experiment. The fillers maintained a balanced ratio of acceptable to unacceptable sentences, as well as an equal number of questions and declarative sentences. After completing the acceptability judgments, participants also performed an operation span task to assess their working memory capacity.

With regards to their proposed analysis, what is of particular interest here is that the fixed effects for this design were STRUCTURE (*simple* vs. *complex*) and GAP-POSITION (*matrix* vs. *embedded*),[12] and they also analyzed their interaction. These effects were coded with treatment contrasts. In the case of STRUCTURE, the level *simple* was treated as the reference level. In the case of GAP-POSITION, the *matrix* level was treated as the reference level. Therefore, in this design, the effect of GAP-POSITION explored the role of both linear and structural distance by comparing the *simple/embedded* and *simple/matrix* conditions. According to the authors, the effect of STRUCTURE addressed the cost associated with an island configuration in the absence of increased filler-gap distance by comparing the *complex/matrix* and *simple/matrix* conditions. However, it is important to note that this is not an island *per se*, but rather a more complex structure (i.e., a 'complex' DP; we will come back to this point in our discussion in Section 6). The interaction between STRUCTURE and GAP-POSITION assessed whether these two factors combine in an interactive way: the presence of a negative interaction shows that the acceptability of the *complex/embedded* condition (i.e., the putative island structure) was lower than expected by the mere addition of the two factors, which they equate to a super-additive effect.

Pañeda et al. (2020) found that subject islands exhibited the strongest super-additive effects compared to the other island types tested. As the authors note, this result is somewhat unexpected, given that, as discussed earlier, subject islands are typically reported to show weaker effects (similar to what López Sancio found in his experiments).

However, it is important to highlight that the subjects in this study were pre-verbal and referential/specific, introduced by a definite article. In this respect, Pañeda et al. observe that definiteness could have contributed to the degradation of this condition. However, since their specific design did not investigate the impact of definiteness on subextraction (i.e., whether definiteness affects subextraction from objects and subjects equally), it is difficult to fully endorse their conclusion that the results "support the prediction that subject island effects are strong when the island sentences bear these characteristics" (Pañeda et al. 2020: 20). Furthermore, as noted by the authors, the prominence of subject islands showing the largest super-additive

[12] In Pañeda et al. (2020), STRUCTURE is called STRUCTURE (i.e., *island/non-island*) and GAP-POSITION is called DISTANCE (i.e., *short/long*).

effect might be influenced by the fact that the *simple*/*matrix* condition (15a) was accepted less frequently than the *simple*/*embedded* condition (15c), which was unexpected.

Finally, it is important to note that the authors acknowledge the impact of using two different designs on evaluating the islandhood of subjects and the strength of the observed effects when comparing their results with those of López Sancio (2015), especially since the simple/complex subject design has the potential to overestimate the effect of the island, as discussed by Kush et al. (2018) and mentioned above. This consideration is one of the reasons our study aimed to compare these two designs in a single experiment.

3.3 Interim summary

Two different designs have been proposed for studying subject islands, each with distinct hypotheses, assumptions, and predictions, which we now summarize for clarity. Although the hypotheses and predictions are not explicitly stated by their proponents, they can be inferred from the designs themselves.

First, the subject/object design operates under two main hypotheses: (1) subjects are more costly to extract than objects, possibly due to the specifier versus complement distinction, and (2) subextraction (extraction *from* a subject or object) is more costly than extraction (extraction *of* a subject or object). From these hypotheses, the subject-object design predicts that subextractions from subjects will lead to lower ratings. This effect could either be linearly additive (suggesting no true island effect) or super-additive (indicating the presence of an island effect).

In contrast, the simple/complex subject design posits that: (1) a complex DP (e.g., a DP with a PP modifier) is more costly than a simple DP (e.g., a DP without a PP modifier), and (2) extractions from embedded clauses are more costly than extractions from matrix clauses. Accordingly, this design predicts that extractions from embedded complex structures will result in lower ratings.

In the present study, we aim to test these hypotheses and their predictions, while addressing several concerns with existing experiments in order to compare the two designs for investigating subject islands. This serves as a first step towards a broader goal: assessing the contribution of various factors, as discussed by Haegeman et al. (2014), in making extractions from subjects deviant, which we do in independent work. To achieve this, we first need to determine the most appropriate methodology for studying subject islands.

4. Our Study

Previous experimental research on subject islands in Spanish is limited, with only two studies specifically investigating these structures, as outlined in the previous section. These studies, however, are based on different designs that may present methodological limitations. Our study aims to begin addressing this gap by being the first to directly compare the two experimental designs for subject islands within a single, controlled experiment. In doing so, we seek to answer the open question of how

best to isolate the effect of subject islands, providing a foundation for future research in this area.

4.1. Participants

Participants were recruited on social media. First, they completed a brief background questionnaire covering age, place of birth, and native language to ensure eligibility. Inclusion in the final analysis was determined post-test, based on these responses. While all participants were allowed to finish the survey, only those meeting the specified criteria were included in the analysis.

A total of 129 participants responded to the survey. From this pool, we excluded individuals who fell outside the 18 to 65 age range, self-reported as bilingual, or indicated that more than one language was spoken in their home during childhood. Additionally, we retained only participants born and raised in regions of Argentina where Rioplatense Spanish is spoken, to control for dialectal variation. Following these criteria, 99 participants remained for analysis.

4.2 Procedure

The acceptability judgment task was administered using PCIbex (Zehr & Schwarz 2018). Participants used their computers or phones. After providing informed consent and completing a brief demographic questionnaire, they rated the acceptability of sentences on a 1-to-7 scale, where 1 meant "completely unacceptable" and 7 meant "completely acceptable". Participants were instructed to rely on their intuitions as native speakers of Spanish, without considering prescriptive grammar rules or the plausibility of the described scenarios. After reviewing the instructions, they completed three practice trials. Each experimental item (context+question) appeared individually on the screen, and participants rated it by either typing a number on the keyboard or selecting it on the screen. Once a rating was submitted, the sentence disappeared and the next one was displayed.

4.3 Materials

Our experiment aimed to test both the subject/object and the simple/complex subject designs (as described in Section 3.1). Thus, the experimental design manipulated the following factors. On the one hand, following the subject/object design, the two factors were STRUCTURE (*subject* vs. *object*) and TYPE OF EXTRACTION (*of*, that is, extraction, vs. *from*, that is, subextraction). On the other hand, the *simple/complex subject* design had the following two factors: STRUCTURE (*simple* embedded subject vs. *complex* embedded subject) and GAP POSITION (*matrix* vs. *embedded*). This yields two 2×2 designs, resulting in eight conditions. However, two of these conditions were identical across designs, reducing the total number to six (unique) conditions. For ease of exposition, since two different factors are called the same (i.e., STRUCTURE), we refer to the STRUCTURE factor in the simple/complex subject design as COMPLEXITY. For clarity, Table 1 provides a summary of the condition labels as proposed in the subject/object and the simple/complex subject designs.

Table 1. Summary of experimental conditions of both designs.

Example	**Subject/object design** TYPE OF EXTRACTION/STRUCTURE	**Simple/complex subject design** COMPLEXITY/GAP POSITION
A Ex. (16a)	*of / object*	—
B Ex. (16b)	*from / object*	—
C Ex. (16c)	*of / subject*	*simple / embedded*
D Ex. (16d)	*from / subject*	*complex / embedded*
E Ex. (16e)	—	*simple / matrix*
F Ex. (16f)	—	*complex / matrix*

We tested 18 sets of items, each containing six context+question pairs, for a total of 108 target items. Each participant was thus exposed to three items per condition. Experimental items were mixed with 18 filler items in a 1:1 ratio, with fillers equally distributed between acceptable and unacceptable items.[13] As a result, each participant rated a total of 36 sentences, plus three additional sentences that served as practice items. Four versions of the survey were created using a Latin Square design to ensure that each participant saw only one condition from each set. The order of fillers and experimental trials was randomized individually for each participant. An example of an experimental item set is provided in (16), where only the question is shown for reasons of space; however, all questions were preceded by a supportive context:[14]

(16) Spanish
a. SOD: *of / object* SCSD: –
¿*Qué nota* dijo Sonia que un periodista del diario La Nación escribió ___ el domingo?
'*What article* did Sonia say that a journalist from La Nación wrote on Sunday?'
b. SOD: *from / object* SCSD: –
¿*Sobre qué actor* dijo Sonia que un periodista del diario La Nación escribió [una nota ___] el domingo?
'*About which actor* did Sonia say that a journalist from La Nación wrote an article on Sunday?'

[13] The unacceptable fillers included sentences with wh-movement out of strong islands (i) and sentences with filled gaps (ii). Importantly, in no case they were 'word salad' sentences:
(i) ¿*Qué libro* dijo Federico que se enojó con su primo porque todavía no le devolvió ___?
'*What book* did Federico say he got angry with his cousin because he still hadn't returned?'
(ii) ¿*Qué película* dijo Agustín que volvió a ver esa película en la tele?
'*What movie* did Agustín say he watched again that movie on TV?'

[14] The rest of the experimental materials, as well as the raw data and the R code, can be found at https://osf.io/s5cxf/?view_only=6a66a2e4ece7414e8647974390a34530.

c. SOD: *of / subject* SCSD: *simple / embedded*
¿Qué periodista dijo Sonia que ___ escribió una nota sobre el actor Ryan Gosling?
'*Which journalist* did Sonia say wrote an article about the actor Ryan Gosling?'
d. SOD: *from / subject* SCSD: *complex / embedded*
¿De qué diario dijo Sonia que [un periodista ___] escribió una nota sobre el actor Ryan Gosling?
'*About which newspaper* did Sonia say that a journalist wrote an article about the actor Ryan Gosling?'
e. SOD: – SCSD: *simple / matrix*
¿Qué estudiante ___ dijo que un periodista escribió una nota sobre el actor Ryan Gosling?
'*Which student* said that a journalist wrote an article about the actor Ryan Gosling?'
f. SOD: – SCSD: *complex / matrix*
¿Qué estudiante ___ dijo que un periodista del diario La Nación escribió una nota sobre el actor Ryan Gosling?
'*Which student* said that a journalist from La Nación wrote an article about the actor Ryan Gosling?'

As we just mentioned, all the questions (both experimental items and fillers) were preceded by a supportive context to make the questions sound more natural. We aimed to create the 'best-case scenario' in terms of considering what factors could negatively influence the judgments, in order to isolate, as much as possible, the island effect. To this end, in addition to providing a supportive context, all the extracted elements were D-linked wh-phrases, which are generally reported to yield higher acceptability rates (see, e.g., Frazier & Clifton 2002, Goodall 2015, among others). To prevent specificity from also negatively affecting the ratings, all relevant DPs (i.e., embedded objects and subjects) were indefinite, headed by the indefinite determiner *un/una* ('a.MASC/a.FEM). Furthermore, all embedded subjects were pre-verbal; while this might not be the preferred position for embedded subjects in wh-questions in Spanish (see, e.g., Torrego 1984, among others), since this was the case for all conditions, any potential negative effects would apply equally across conditions. Finally, conditions (a) and (b) included an adjunct on the right edge of the sentence to avoid having a gap in final position.[15]

4.4 Analysis

To further refine the dataset, participants whose mean ratings for grammatical and ungrammatical fillers deviated by more than 2.5 standard deviations from the overall mean for these conditions were excluded from the analysis (N = 5 excluded; N = 94 remained).

Mixed-effects ordinal logistic regressions, from the 'ordinal' R package (Christensen, 2022), were used as they are recommended for discrete ordinal response

[15] We used different prepositions for extractions *from* objects (i.e, condition (b)) such as *de* ('of'), *sobre* ('about'), and *para* ('for') to avoid participants getting used to seeing always the same preposition, and to off-set any effects that a particular preposition might have. All participants saw the same number of items for each preposition (that is, one per preposition).

variables with non-normal distributions (Bürkner & Vuorre 2019, Veríssimo 2021). This type of analysis was preferred over mixed-effects linear regressions with *z*-scores as treating Likert-scale data as continuous and standardizing it (i.e., using a *z*-score transformation) would assume interval-scale properties, which may not be appropriate and could lead to misleading inferences (Agresti 2010, Liddell & Kruschke 2018). For the simple/complex subject design, the fixed effects in the model included COMPLEXITY (*simple* vs. *complex*) and GAP POSITION (*matrix* vs. *embedded*), as well as their interaction. These effects were coded with treatment contrasts. In the case of COMPLEXITY, the level *simple* was treated as the reference level; and in the case of GAP POSITION, the *matrix* level was treated as the reference level. Therefore, we replicated the simple effects and the interaction configuration proposed by Pañeda et al. (2020) (see Section 3.2).

For the subject/object design, the fixed effects in the model included STRUCTURE (*subject* vs. *object*) and TYPE OF EXTRACTION (extractions *of* vs. extractions *from*); these effects were also coded with treatment contrasts. For the TYPE OF EXTRACTION factor, the *of* level was treated as the reference level; and for STRUCTURE, the *object* level was treated as the reference level. In this configuration, the effect of TYPE OF EXTRACTION isolated the cost associated with sub-extraction (i.e., extractions *from*), by comparing the *of/object* and *from/object* conditions. The effect of STRUCTURE assessed the differential cost of extracting a subject versus an object, by comparing the *of/object* and *of/subject* conditions. The model also included the interaction between both factors.

All models had maximal random effects structures, including intercepts and slopes by participants and items for all fixed effects and their interactions (Barr et al. 2013). The inclusions of these random effect structures were permitted by the designs. We report effect sizes with model coefficients in log-odds (β), standard errors (SE), the *z*-statistic, and the *p*-value.

5. Results

Participants overall responded to filler items as expected: grammatical sentences had a mean rating of 6.66 (SE = 0.04), whereas ungrammatical sentences were rated, on average, 2.32 (SE = 0.11). Mean results per condition (±SE) for each design can be observed in Figures 2 and 3. Figure 2 summarizes data obtained from the subject/object design; while Figure 3, from the simple/complex subject design:

Figure 2. Mean ratings per condition in the SOD. Error bars indicate standard errors (SE).

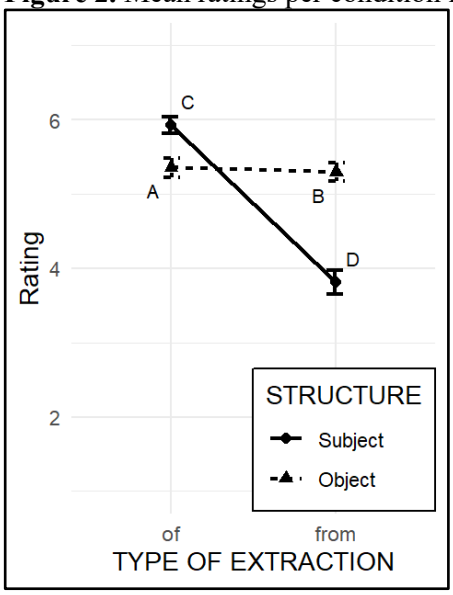
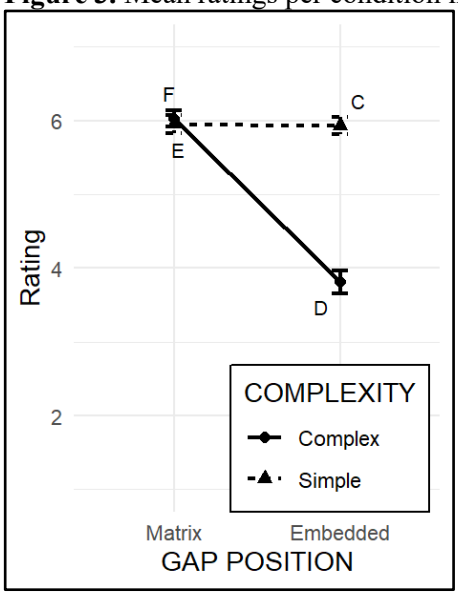


Figure 3. Mean ratings per condition in the SCSD. Error bars indicate standard errors (SE).



Tables 2 and 3 summarize the main findings from the models fitted to the data obtained. In the case of the simple/complex subject design, both fixed effects were found to be not significant, while a significant negative interaction was indeed found. In the case of the subject/object design, a significant effect of the factor STRUCTURE was found but not a significant effect of TYPE OF EXTRACTION. Also, a significant negative interaction between both factors was observed.

Table 2. Results of the model fitted to data obtained from the simple/complex subject design. Model estimates are expressed in log-odds and significant effects at the α = .05 level are bolded.

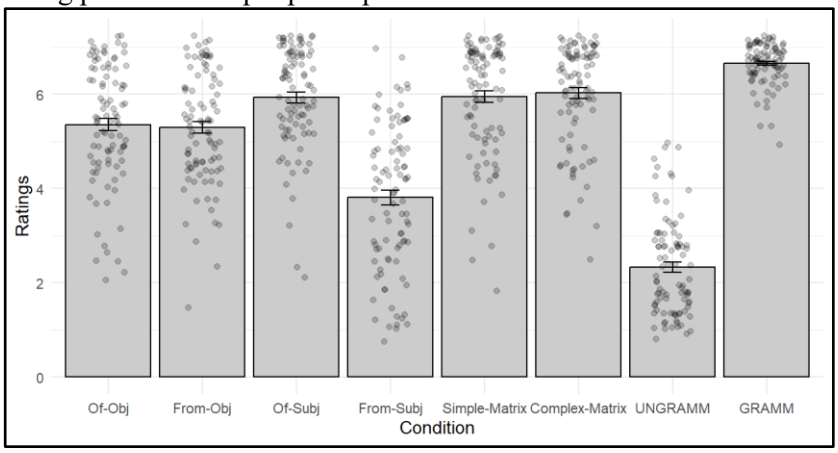
Effects and interactions	$\hat{B}$	SE	*z*	*p*
Complexity	−0.06	0.25	−0.22	0.823
Gap position	−0.31	0.36	−0.84	0.40
Gap position × *Complexity*	**−3.18**	**0.41**	**−7.71**	**1.22e−14**

Table 3. Results of the model fitted to data obtained from the subject/object design. Model estimates are expressed in log-odds and significant effects at the α = .05 level are bolded.

Effects and interactions	$\hat{B}$	SE	*z*	*p*
Structure	0.96	0.31	3.09	0.02
Type of extraction	−0.08	0.34	−0.24	0.81
Structure ×*Type of extraction*	**−3.04**	**0.47**	**−6.53**	**6.62e−11**

Upon closer examination of the experimental data, a more complex picture emerges. First, the raw average rating for the putative island structure (i.e., extractions from subjects—condition (16d)) is 3.81 (SE = 0.16), while the raw average rating for ungrammatical fillers is 2.32 (SE = 0.11). This suggests that while the purported island-violating structure is indeed degraded compared to its baselines, it is still relatively acceptable compared to ungrammatical sentences. This middle-range result for the island-violating structure prompted us to investigate whether participants were uniformly rating these structures near the middle of the scale or whether there might be a bimodal distribution, with some participants judging these structures as highly degraded while others found them relatively acceptable. As shown in Figure 4, there is considerable variation in the ratings of this structure, indicating substantial divergence in participants' judgments compared to the other structures tested.

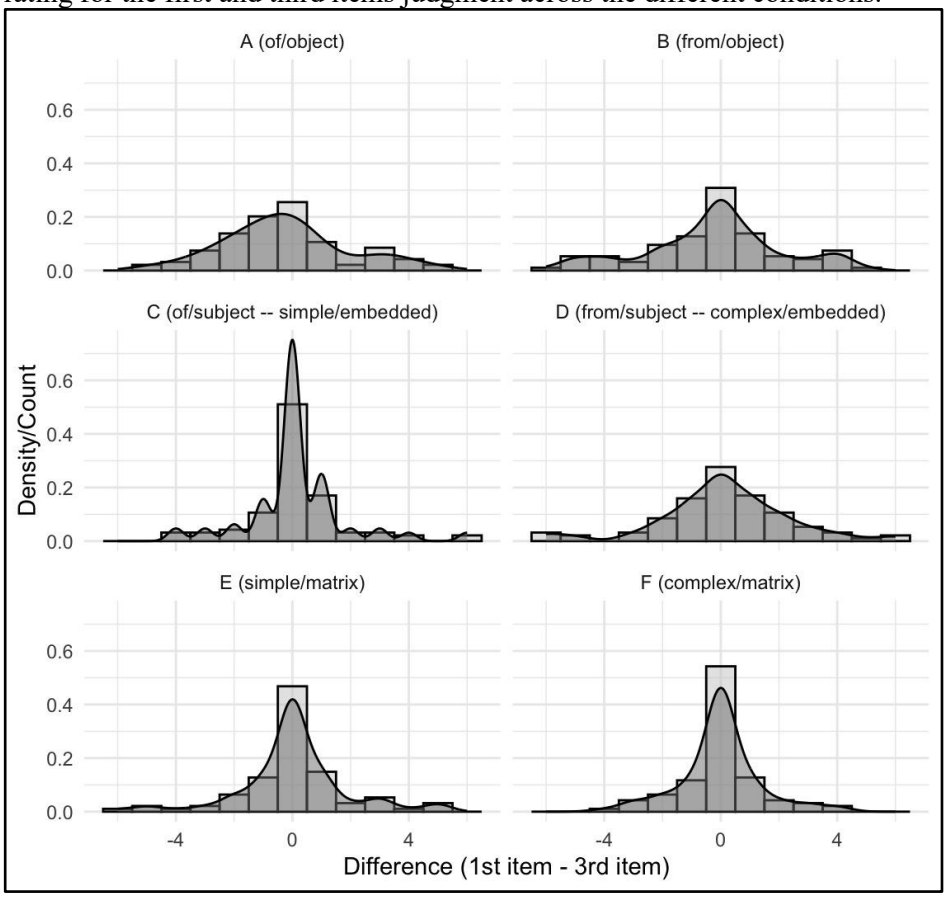
Figure 4. Bars show mean rating per condition (±SE). Scattered data points show mean rating per condition per participant.



One might wonder whether, in addition to inter-speaker variation, we also observe intra-speaker variation, and, if so, whether there are indications of the so-called *satiation effects* (see Lu et al. 2024 for an overview; see especially Goodall 2011 for a study on satiation in Spanish and English, which examines the interaction between wh-inversion and island effects), even though this experiment was not specifically designed to test the presence of those effects. To explore this possibility, we analyzed the ratings for the first and third judgments for each condition (remember that each participant saw three items per condition).

Figure 5 displays a combined histogram and density plot illustrating the distribution of differences between the rating for the first and third items judgment across the different conditions. A difference closer to 0 suggests consistency in the participants' ratings, while a difference closer to -6 suggests potential satiation. In other words, if satiation were to occur, we would expect the histogram and density plot to be skewed to the left. However, a visual inspection of the plots reveals no such skewness. Instead, all conditions appear to exhibit a normal distribution centered around 0, leading us to conclude that there is no evidence of intra-speaker variation.

Figure 5. Histogram and density plot illustrate the distribution of differences between the rating for the first and third items judgment across the different conditions.



We calculated the skewness and kurtosis for the difference values across the six different conditions. The skewness values ranged from −0.239 to 0.613, indicating that the distributions of difference are generally close to symmetrical, with some mild left-skew (negative values) or right-skew (positive values). The kurtosis values ranged

from −0.0428 to 2.97, suggesting that the distributions are mostly mesokurtic, with only a few distributions showing heavier tails (i.e., in particular, extractions *of subjects*). These results are summarized in Table 4. A Kruskal-Wallis rank sum test was conducted to determine if there were significant differences in the difference values across the six conditions. The results indicated no statistically significant difference between the groups ($\chi^2(5) = 6.15$, $p = 0.292$).

Table 4. Skewness and kurtosis test results for difference values across conditions.

Condition	Skewness	Kurtosis
A. of / object	0.348	−0.0428
B. from / object	−0.185	−0.0549
C. of /subject – simple / embedded	0.613	2.97
D. from / subject – complex / embedded	−0.239	1.12
E. simple / matrix	−0.122	2.17
F. complex / matrix	0.138	1.62

6. Discussion

The results obtained show a negative interaction between the fixed effects in both designs, hence we found super-additivity, as defined in Sprouse et al. (2012, 2016) and explained in Section 3.1. However, we challenge the exclusive reliance on a significant interaction—and, by extension, on super-additivity—as the sole criterion for determining which factors are relevant to our understanding of island structures. To further elaborate, Sprouse's original design aims to isolate the factors that might independently contribute to the degradation of a sentence, in order to assess whether there is a super-additive effect (as opposed to just a linear additive effect) in the condition that combines the two degrading factors, that is, in the purported island-violating condition (see footnote 7). However, we find that in both designs, the factors that were expected to produce a decrease in acceptability either had no effect or even the opposite effect (i.e., that factor yielded significantly higher ratings). More specifically, in the simple/complex subject design, the complexity of the embedded subject did not show a significant effect on the extraction of matrix subjects, that is, the *complex/matrix* and *simple/matrix* conditions were rated similarly. Moreover, the GAP POSITION factor didn't trigger any degradation either, that is, the s*imple/embedded* and *simple/matrix* were rated similarly, and no significant effect was found. In the subject/object design, the situation was even more striking: extractions of subjects were rated significantly higher than extractions of objects. Additionally, no significant effect was found for TYPE OF EXTRACTION, that is, extraction *of* object versus extraction *from* objects wasn't significantly different. This means that we didn't find this factor to be a degrading one. It is worth mentioning that we are analyzing simple effects and not main effects but, following the analysis in Pañeda et al. (2020), we believe this makes more sense given the hypotheses that we are trying to test, as explained in Section 3.3. Given that all these factors do not appear to produce an additional

degradation, we cannot say that the island-violating condition exhibits the same super-additive effect as in the original proposal by Sprouse et al. (2012, 2016). Nevertheless, it is noteworthy that both designs reveal an important drop in the acceptability of extractions from subjects compared to all the other conditions. These results fit the traditional definition of subject islands adopted in the theoretical syntax literature, that is, a constraint on movement from subjects.

Regarding the subject/object design in particular, our experimental design does not exhibit the potential shortcomings observed by Kush et al. (2018). As pointed out in Section 3.1, the authors claim that extractions from objects could trigger a filled-gap effect after the embedded verb, which would produce an underestimation of the putative subject island violation. Now, given that Spanish does not allow P-stranding, movement from objects always involves extracted PPs. Hence, there is no risk of a filled-gap effect in our stimuli: pied-piping precludes the interpretation of the extracted wh-phrase as the direct object of the embedded predicate. Moreover, Kush et al. further point out that the presence of complex DPs in both subject and object positions in extractions from these domains decreases the acceptability of these conditions, which may lead to a floor effect that reduces the extent of the subject island effect. However, contrary to this prediction, our results do not show a floor effect. Thus, according to our findings, none of the concerns raised by Kush et al. seem to apply to the subject/object design, that is, it does not seem to underestimate the size of the subject island effect in Spanish.

Furthermore, the results obtained from the subject/object design do not show a general effect of sub-extraction, as noted before. This conclusion follows from comparing extractions *of* objects with extractions *from* objects, which were not significantly different. By contrast, extractions *of* subjects were clearly more acceptable than extractions *from* subjects. This suggests that the substantial decrease observed in this condition might be due to a specific grammatical constraint (i.e., the subject island itself), and not to the interplay between the extraction domain and the sub-extraction operation. In other words, sub-extraction does not appear to independently lower acceptability.

As for the simple/complex subject design, as mentioned above, our experiment reveals that the complexity of the embedded subject is not significant when extracting matrix subjects. This means that adding a PP modifier within an embedded subject is not equivalent to having a true island configuration, as in the canonical factorial design proposed by Sprouse (2007, and subsequent work). Additionally, our experiment also reveals that the position of the extracted subject is also not significant. This means that extracting matrix subjects is not significantly different from extracting embedded subjects. Upon closer examination of the stimuli, the four sentences resulting from crossing the two relevant factors in this design do not constitute strict minimal pairs, which might explain the results obtained. Although this design doesn't check for an effect of sub-extraction, the results from the subject/object design reveal that this is not an overall relevant effect, but it only affects extractions from subjects, which is in line with its assumptions.

In sum, neither of the two designs seem to be adequate in explaining which factors contribute to the degradation of extractions from subjects (i.e., subject islands). However, both seem equally adequate in showing that subjects are indeed islands. This leads us to conclude that, unlike other islands, subject islands might arise from the

interplay and cumulative effects of other factors, as noted by Haegeman et al. (2014), but these factors might be related to the properties of the subject itself (e.g., definiteness/specificity, thematic role, among others) and the properties of the extractee (e.g., D-linking, thematic role, among others), rather than related to the overall structure of the sentence.

Finally, when we look at variability, we found substantial inter-speaker variation, as shown by a visual inspection of the data. In recent research in this area, the question of inter-speaker variation and speaker consistency has been brought up as a new area of exploration (Hoot & Ebert, 2024). We found that, in our sample, none of the conditions seem to show a not normal distribution, and the distributions of the different conditions are not significantly different. By comparing the first item that participants rated with the third item that participants rated for a given condition, we aimed to answer the question of whether there is evidence of a satiation effect, especially in the purported island violation, which has been reported to show such effects (see, e.g., Lu et al. 2024 for an overview). We did not find any indication of a satiation effect in our data. However, given that our experiment wasn't specifically designed to test these effects, these results should be taken as preliminary.

7. Conclusions

This study aimed to compare two experimental designs for investigating subject islands: the subject/object design and the simple/complex subject design. Our goal was to assess how these designs capture island effects, particularly focusing on the differences in their predictions and assumptions to establish a baseline for future investigations in this area.

Our findings reveal that both designs show a statistically significant interaction for the island-violating structure, supporting the definition of an island in a factorial design. However, we question the assumption that a significant interaction, and by extension super-additivity, should be the sole criterion for identifying island structures experimentally, given that we found an interaction between the manipulated factors even when they did not show the predicted degrading effects on their own. In other words, we challenge the idea that islands necessarily should be defined as super-additive interactions between the factors tested, as proposed in previous studies, as we can find super-additivity without significant simple effects.

The results obtained in our experiment set the basis for the study of subject islands and discuss important methodological considerations with respect to what are the most appropriate comparisons that should be tested to understand these islands, which seem to be qualitatively different from other islands. This emphasizes the role of experimental methods in helping shape linguistic theory (see Lewis & Phillips 2015, and references therein), and, in particular, it opens new avenues for research by identifying which factors are relevant in the study of subject islands. In on-going research, we explore some of the factors proposed by Haegeman et al. (2014) such as the thematic role of the subject and the pre- and post-verbal position.

Acknowledgments

We are grateful to the audiences at the *54th Linguistic Symposium on Romance Languages* (LSRL54) for their valuable feedback and discussion. We also thank Eszter Ronai for engaging with us on the initial ideas. Finally, we extend our appreciation to the editor, Adam McBride, and the two anonymous reviewers for their insightful comments and suggestions. All remaining errors and shortcomings are our own.

References

Abeillé, Anne, Hemforth, Barbara, Winckel, Eloide, & Edward Gibson. 2020. Extraction from subjects: differences in acceptability depend on the discourse function of the construction. *Cognition* 204: 1-23. https://doi.org/10.1016/j.cognition.2020.104293

Agresti, Alan. 2010. *Analysis of ordinal categorical data* (2nd ed.). Chichester, England: John Wiley & Sons. https://doi.org/:10.1002/9780470594001

Barr, Dale J., Levy, Roger, Scheepers, Christoph, & Harry J. Tily. 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

Bürkner, Paul-Christian, & Vuorre, Matti. 2019. Ordinal Regression Models in Psychology: A Tutorial. *Advances in Methods and Practices in Psychological Science* 2(1): 77-101. https://doi.org/10.1177/2515245918823199

Cattell, Ray. 1976. Constraints on Movement Rules. *Language* 52: 18-50. https://doi.org/10.2307/413206

Chomsky, Noam. 1973. Conditions on transformations. In S. Anderson, & P. Kiparsky (eds), *A festschrift for Morris Halle*, 232-286. NY: Academic Press.

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

Chomsky, Noam. 2001. Derivation by Phase. In M. Kenstowicz (ed.), *Ken Hale: A Life in Language*, 1-52. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/4056.003.0004

Chomsky, Noam. 2008. On phases. In R. Freidin, C. P. Otero, & M. L. Zubizarreta (eds), *Foundational Issues in Linguistic Theory*, 133-166. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/7713.003.0009

Christensen, Rune Haubo Bojesen. 2022. *Ordinal-regression models for ordinal data. R package version 2022,* 11-16. https://cran.r-project.org/web/packages/ordinal

Cinque, Guglielmo. 1977. Towards a unified treatment of island constraints. In W. Dressler, & W. Meid (eds), *Proceedings of the Twelfth International Congress of Linguists*, 344-348. Innsbruck: Innbrucker Beiträge zur Sprachwissenschaft.

Cinque, Guglielmo. 1990. *Types of A'-dependencies*. Cambridge, MA: MIT Press.

Cowart, Wayne. 1997. *Experimental syntax*. CA: SAGE Publications.

Fiengo, Robert, & James Higginbotham. 1981. Opacity in NP. *Linguistic Analysis* 7: 347-373.

Frazier, Lyn, & Charles Clifton. 2002. Processing "d-Linked" Phrases. *Journal of Psycholinguist Research* 31: 633-659. https://doi.org/10.1023/A:1021269122049

Gallego, Ángel J. 2011. Successive cyclicity, phases, and CED effects. *Studia Linguistica* 65(1): 32-69. https://doi.org/10.1111/j.1467-9582.2010.01175.x

Gallego, Ángel J., & Juan Uriagereka. 2007. Sub-extraction from subjects: A phase theory account. In J. Camacho, N. Flores-Ferrán, L. Sánchez, V. Deprez, & M. J. Cabrera (eds), *Romance linguistics 2006: Selected papers from the 36th linguistic symposium on Romance languages (LSRL)*, 155-168. Philadelphia, PA: John Benjamins. https://doi.org/10.1075/cilt.287.12gal

Gieselman, Simone, Kluender, Robert, & Ivano Caponigro. 2013. Isolating processing factors in negative island contexts. In Y. Fainleib, N. LaCara, & Y. Park (eds), *Proceedings of NELS 41*, 233-246. Amherst, MA: Graduate Linguistic Student Association.

Goodall, Grant. 2011. Syntactic Satiation and the Inversion Effect in English and Spanish *wh*-Questions. *Syntax* 14: 29-47.

Goodall, Grant. 2015. The D-linking effect on extraction from islands and non-islands. *Frontiers in Psychology* 5: 1493. https://doi.org/10.3389/fpsyg.2014.01493

Haegeman, Liliane, Jiménez-Fernández, Ángel L., & Andew Radford. 2014. Deconstructing the Subject Condition in terms of cumulative constraint violation. *The Linguistic Review* 31(1): 73-150. https://doi.org/10.1515/tlr-2013-0022

Hofmeister, Philip, Casasanto, Laura Staum, & Ivan A. Sag. 2013. Islands in the grammar? Standards of evidence. In J. Sprouse, & N. Hornstein (eds), *Experimental Syntax and Island Effects*, 42-63. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139035309.004

Hofmeister, Philip, Casasanto, Laura Staum, & Ivan A. Sag. 2014. Processing effects in linguistic judgment data: (Super-)additivity and reading span scores. *Language and Cognition* 6: 111-145. https://doi.org/10.1017/langcog.2013.7

Hoot, Bradley, & Shane Ebert. 2025. Syntactic island effects in Spanish: Experimental evidence. *Natural Language and Linguistic Theory* 43: 629-667. https://doi.org/10.1007/s11049-024-09633-5

Huang, Cheng-Teh James. 1982. *Logical relations in Chinese and the theory of grammar*. Ph.D. thesis, Massachusetts Institute of Technology.

Jiménez Fernández, Ángel L. 2009. On the composite nature of subject islands: A phase-based approach. *SKY Journal of Linguistics* 22: 91-138.

Kayne, Richard S. 1981. ECP extensions. *Linguistic Inquiry* 12: 93-133.

Keshev, Maayan, & Aya Meltzer-Asscher. 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

Kluender, Robert. 2004. Are subject islands subject to a processing account? In V. Chand, A. Kelleher, A. J. Rodríguez, & B. Schmeiser (eds), *West Coast Conference on Formal Linguistics (WCCFL)* 23: 101-125. Somerville, MA: Cascadilla Press.

Kluender, Robert, & Simone Gieselman. 2013. What's negative about negative islands? A re-evaluation of extraction from weak island contexts. In J. Sprouse, & N. Hornstein (eds), *Experimental syntax and island effects*, 186-207. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139035309.010

Kush, Dave, Lohndal, Terje, & Jon Sprouse. 2018. Investigating variation in island effects: A case study of Norwegian wh-extraction. *Natural Language & Linguistic Theory* 36: 743-779. https://doi.org/10.1007/s11049-017-9390-z

Lewis, Shevaun, & Colin Phillips. 2015. Aligning grammatical theories and language processing models. *Journal of Psycholinguistic Research* 44(1), 27-46. https://doi.org/10.1007/s10936-014-9329-z

Liddell, Torrin M., & John K Kruschke. 2018. Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology* 79: 328-348. https://doi.org/10.1016/j.jesp.2018.08.009

López Sancio, Sergio. 2015. *Testing syntactic islands in Spanish*. M.A. thesis, Euskal Herriko Unibertsitatea.

Lu, Jiayi, Frank, Michael C., & Judith Degen. 2024. A meta-analysis of syntactic satiation in extraction from Islands. *Glossa Psycholinguistics* 3(1): 1-33. http://doi.org/10.5070/G60111425

Maxwell, Scott E., & Harold D. Delaney. 2003. *Designing experiments and analyzing data: A model comparison perspective*. Mahwah: Lawrence Erlbaum Associates.

Ordóñez, Francisco, & Esthela Treviño. 1999. Left Dislocated Subjects and the Pro-Drop Parameter: A Case Study of Spanish. *Lingua* 107: 39-68. https://doi.org/10.1016/S0024-3841(98)00020-5

Pañeda, Claudia, Lago, Sol, Vares, Elena, Veríssimo, João, & Claudia Felser. 2020. Island effects in Spanish comprehension. *Glossa: a journal of general linguistics*, 5(1): 1-30. https://doi.org/10.5334/gjgl.1058

Pesetsky, David. 1987. *Wh*-in-situ: Movement and unselective binding. In E. J. Reuland, & A. G. B. ter Meulen (eds), *The Representation of (in)definiteness*, 98-129. Cambridge, MA: MIT Press.

Postal, Paul M. 1974. *On Raising*. Cambridge, MA: MIT Press.

Ross, John R. 1967. *Constraints on variables in syntax*. Ph.D. thesis, Massachusetts Institute of Technology.

Sprouse, Jon. 2007. *A program for experimental syntax: Finding the relationship between acceptability and grammatical knowledge*. Ph.D. thesis, University of Maryland.

Sprouse, Jon. 2023. Acceptability judgments. In J. Sprouse (ed.), *The Oxford Handbook of Experimental Syntax*, 3-28. Oxford: Oxford University Press.

Sprouse, Jon, Wagers, Matt, & Colin Phillips. 2012. A test of the relation between working-memory capacity and syntactic island effects. *Language* 88: 82-123. https://doi.org/10.1353/lan.2012.0004

Sprouse, Jon, Caponigro, Ivano, Greco, Ciro, & Carlo Cecchetto. 2016. Experimental syntax and the variation of island effects in English and Italian. *Natural Language and Linguistic Theory* 34: 307-344. https://doi.org/10.1007/s11049-015-9286-8

Starke, Michal. 2001. *Move dissolves into Merge: A theory of locality*. Ph.D. thesis, University of Geneva.

Suñer, Margarita. 1991. Indirect questions and the structure of CP: Some consequences. In H. Campos, & F. Martínez-Gil (eds), *Current studies in Spanish linguistics*, 283-312. Washington, DC: Georgetown University Press.

Torrego, Esther. 1984. On inversion in Spanish and some of its effects. *Linguistic Inquiry* 15(1): 103-129.

Uriagereka, Juan. 1999. Multiple spell-out. In S. Epstein, & N. Hornstein (eds), *Working Minimalism*, 251-282. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/7305.003.0012

Uribe-Etxebarria, Myriam. 1995. On the nature of Spec/IP and its relevance for scope asymmetries in Spanish and English. In J. Amastae, G. Goodall, M. Montalbetti, & M. Phinney (eds), *Contemporary Research in Romance Linguistics*, 355-366. Amsterdam: John Benjamins.

Veríssimo, João. 2021. Analysis of rating scales: A pervasive problem in bilingualism research and a solution with Bayesian ordinal models. *Bilingualism: Language and Cognition* 24(5): 842-848. https://doi.org/10.1017/S1366728921000316

Wexler, Kenneth, & Peter W. Culicover. 1980. *Formal principles of language acquisition*. Cambridge, MA: MIT Press.

Winckel, Elodie, & Anne Abeillé. 2020. French subject island constraint? A discourse-based HPSG account. *Lingvisticæ Investigationes* 43(1): 95-128. https://doi.org/10.1075/li.00041.win

Zehr, Jérémy, & Florian Schwarz. (2018). PennController for Internet Based Experiments (IBEX). https://doi.org/10.17605/OSF.IO/MD832