

For which speakers do you wonder whether *si* is a syntactic island?



Photo by Maysam Yabandeh

Bradley Hoot
DePaul University

Shane Ebert
University of Illinois Chicago



[tinyurl.com/
LSRL2023](https://tinyurl.com/LSRL2023)

53rd Linguistic Symposium
on Romance Languages
June 27, 2023

Our experiment finds island effects for Spanish *si* just like English *whether* but with substantial individual variation, which is poorly understood.



2

Our presentation reports the results of an experiment investigating syntactic island effects with the complementizer *si* in Spanish, and what we will show you today is that with these constructions in Spanish we observe island effects of similar magnitude to those of English, alongside substantial variation both within and across participants.

To contextualize our project, we'll first talk about the impetus for studying islands, then what we did and our general findings, and then investigate the individual variation.

Long-distance dependencies are generally unbounded but are restricted in some contexts, known as *islands*.

Non-island (unbounded)

- (1) ¿Qué tarea escuchaste que Mateo copió ___?
'Which homework did you hear that Mateo copied ___?'



Complex NP Island

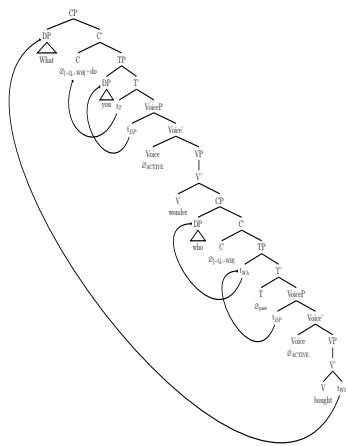
- (2) * ¿Qué tarea escuchaste [el rumor de que Mateo copió ___]?
'Which homework did you hear the rumor that Mateo copied ___?'

3

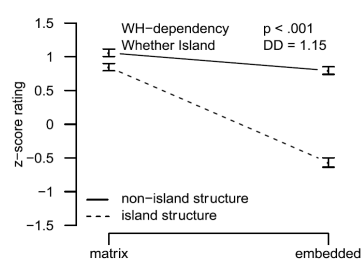
Wh-phrases can be extracted out of their clause quite productively, such as in (1).

But in some contexts, there are constraints on how local that movement or filler-gap dependency needs to be. For example, you cannot move it out of a clause that's part of a complex NP, as in (2).

Islands raise questions core to linguistic theory and have been intensely studied, with a recent focus on experimental methods.



Ross (1967)
Chomsky (1973)
Bresnan (1976)
Huang (1982)
Kayne (1984)
Chomsky (1986)
Rizzi (1990)
etc.

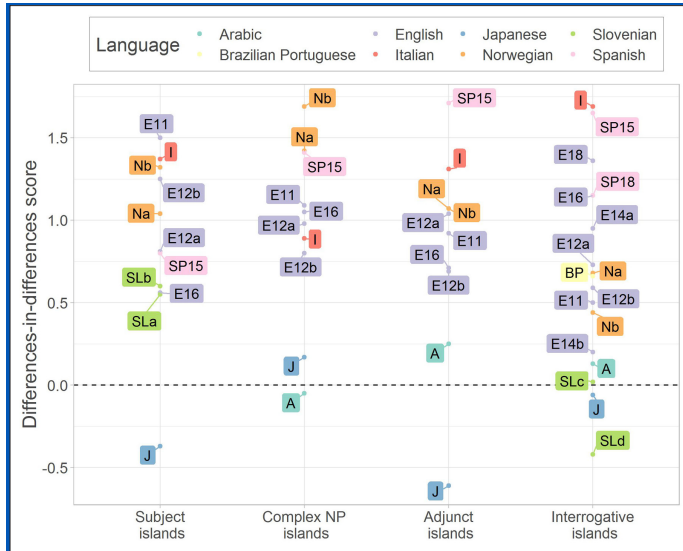


4

Because understanding filler-gap dependencies is central to any linguistic theory (Chomsky et al. 2019), restrictions on displacement have been of great interest to linguists, resulting in intense research on the topic over the past 50 years (see Boeckx 2012; Citko 2016; Szabolcsi & Lohndal 2017 for an overview).

Recently, this research has undergone an experimental boom, following work especially by Sprouse and colleagues (e.g., Sprouse et al. 2012, 2016). Yet many questions remain.

There is debate about the empirical record and how to interpret it.



Pañeda et al. (2020)
Fig. 1

5

Despite the scrutiny by experimentalists, questions remain about the evidence. You can observe the wide range of effects observed for different islands in different languages in this figure created by Pañeda et al. (2020), which plots effect size measured by DD score on the y-axis and four types of islands on the x-axis. Each point is an experiment, marked with the language examined.

We can see a substantial range of effects for all island types. Furthermore, those effects are interpreted differently by different authors.

For example, among the interrogative islands at the right, of which *whether* islands are a subset, some authors interpret the effects they find to support analyses of these structures as islands, some argue they show these are not islands at all, and some have analyses in between.

In this context, data from Spanish *whether*-islands is relevant because Spanish has been claimed not to instantiate this restriction.

English → Weak islands = Extraction selectively possible (Szabolcsi & Lohndal 2017)

- (3) Which man are you wondering <whether to invite ____>?
- (4) * What did he wonder <whether the pearls cost ____>?
- (5) * How many books are you wondering <whether to write ____ next year>?

Spanish → Not islands = Extraction always possible (Torrego 1984)

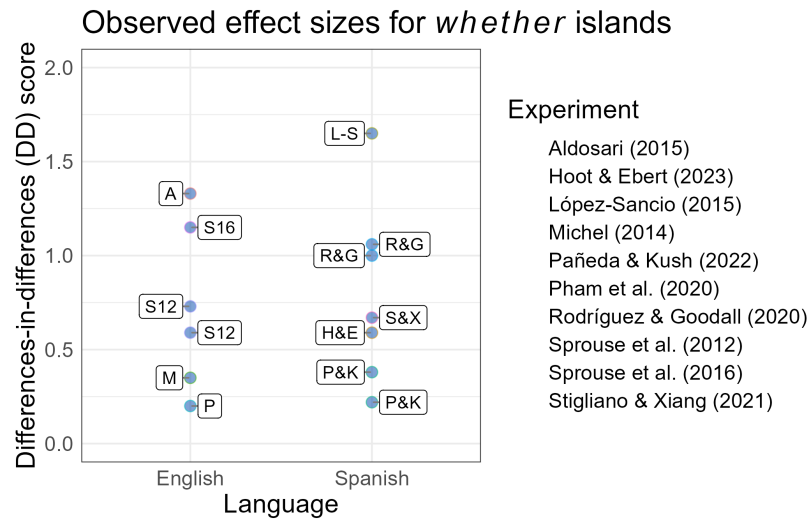
- (6) ¿A quién se pregunta Juan <si María quiere ____>?
Who does Juan wonder whether María loves?
- (7) ¿Qué diccionario no sabías <si Celia había devuelto ya ____>?
Which dictionary didn't you know whether Celia had returned already?

6

In this landscape, Spanish is of special interest. Unlike English, which bars extraction of *whether* in at least some cases, as you can see in (3-5), Spanish has been claimed to uniformly allow extraction over *si* 'whether.'

This contrast has been of special interest to linguists working on Spanish, and *si* islands have been the most intensely studied of all islands in Spanish.

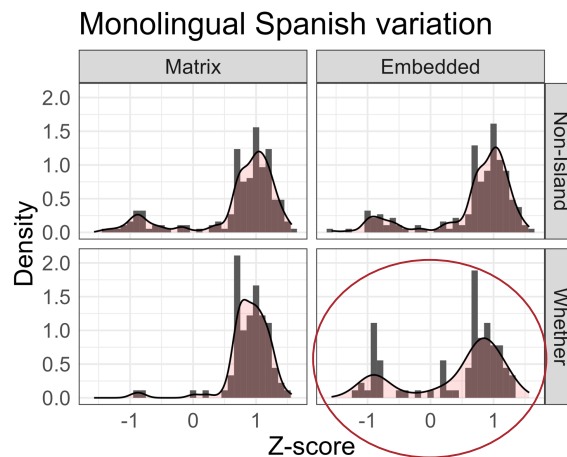
Experimental results for *whether* islands in Spanish range widely and mirror those from English; both languages have mixed and disputed results.



But if we plot the existing studies of Spanish *whether* islands using the same format Pañeda used, and if we compare them to studies of *whether* in English, we observe the same situation with Spanish as we observe cross-linguistically: a large range of effect sizes, from very small to very large, with correspondingly different interpretations.

What's more, one of these previous studies is our own, and when we looked a little closer, we observed curious individual variation.

We previously observed variation in judgments of *whether* islands in Spanish, with many participants rating these island violations high.



8

Our previous study found a statistically significant island effect and a medium effect size, yet when we examined the distribution of scores, we found a bimodal distribution, with some low ratings and rather more high ratings, casting doubt on the meaning of the island effect.

Although there was undoubtedly an effect under the common factorial definition of islands, can we really say these sentences are ruled out when the majority of speakers rate them high?

We set out to systematically test *whether* islands in Spanish to contribute to clarifying the empirical picture.

“...there is real empirical value in systematically re-testing languages for island effects ... to establish the range of variation across languages and ... dependency types.”

(Sprouse & Villata, 2021)

9

In this landscape, the purpose of the study we are reporting today is to add clarity to the empirical picture by examining Spanish *whether* islands again and comparing them directly to English, in the spirit of this quote.

Our experiment was designed to mitigate some sources of variation that can limit comparability across studies. Different experiments generally test different materials and populations, but we tested the same participants across languages by recruiting Spanish/English bilinguals, and we used uniform materials that were translation equivalents to reduce possible confounds from task effects.

Additionally, we examine individual variation for a more complete picture.

We used a 3x2 factorial design to isolate island effects.

3 Island Conditions

2 Gap Positions

Non-Island

Complex NP Islands

Whether Islands

Matrix Clause

Embedded Clause

10

This design adapts the factorial design created by Sprouse and colleagues (Sprouse et al. 2012, 2016). We follow Stigliano & Xiang (2021) in comparing multiple islands against a single non-island condition in a larger design.

We investigated two islands, compared to a non-island condition, with materials designed to favor extraction.

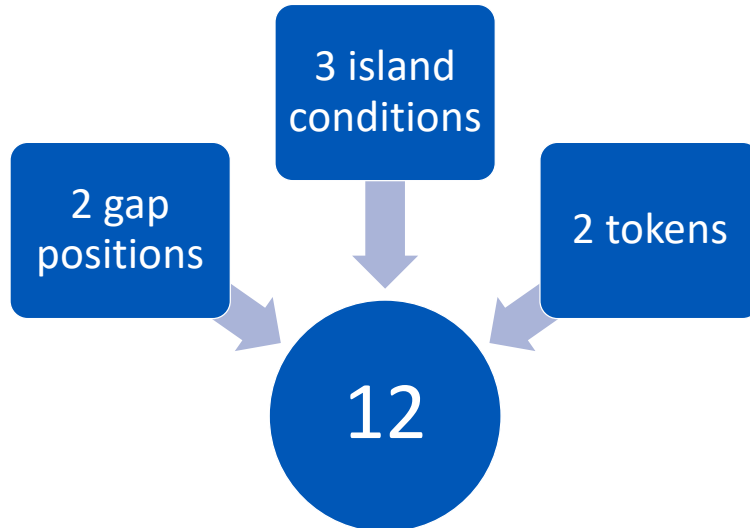
- | | |
|--|--------------------------|
| (8) ¿Qué tarea escuchaste que Mateo copió ____?
'Which homework did you hear that Mateo copied ____?' | Non-island |
| (9) ¿Qué tarea escuchaste el rumor de que Mateo copió ____?
'Which homework did you hear the rumor that Mateo copied ____?' | Complex NP Island |
| (10) ¿Qué tarea quieres saber si Mateo copió ____?
'Which homework do you want to know whether Mateo copied ____?' | Whether Island |

11

(8) is uncontroversially grammatical, and (9) is uncontroversially ungrammatical, so they serve as baselines of comparison against which we can compare sentences like (10).

We took several measures to favor extraction in order to give the fairest test of the proposition that (10) is grammatical in Spanish. We used D-linked wh-phrases (i.e., *which book* rather than *what*), which facilitates extraction. We also avoided rogative verbs like *preguntar* 'ask' and *preguntarse* 'wonder,' which are used in most previous studies of wh-islands in Spanish, but which have been argued to independently prohibit extraction (Suñer 1991; Torrego 1984; see also Pañeda & Kush 2022 for some experimental support for this view). We also controlled lexical items for frequency and semantic properties.

Participants judged 12 target sentences in each language.

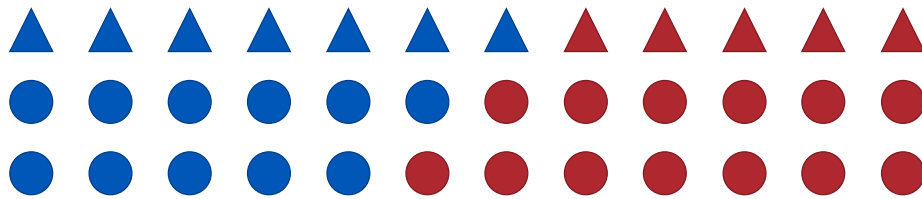


12

Each person judged 12 sentences per language, distributed by Latin square.

Language blocks were rotated by participant, so some judged English before Spanish and some vice versa. This was part of a larger study of Spanish-English code-switching, so all participants judged mixed sentences with code-switching first. In Grosjean's (1998) terms, all participants were certainly in a bilingual mode, with both languages activated.

They also judged 24 fillers in each language.



13

Additionally, they judged 24 fillers per language for a 2:1 ratio of fillers to target, a 1:1 overall ratio of grammatical to ungrammatical, and items with the full range of acceptability, along with anchor items demonstrating the full range of the scale and practice items.

US-raised Spanish-English bilinguals (n = 60) completed a written acceptability judgment task (AJT) via Prolific.



¿Qué edificio escuchaste la noticia de que Víctor diseñó?

(mal) 1 2 3 4 5 6 7 (bien)

Haz clic en los cuadritos para contestar.

14

Participants judged the sentences on a seven-point scale without accompanying contexts, as shown here.

We assessed proficiency with a lexical decision task, working memory with a backward digit span task, and language history with a background questionnaire.

Estimulo	Palabra?	Estimulo	Palabra?	Estimulo	Palabra?
terzo		pandilla	✓	seda	✓
pellizcar	✓	fatacidad		flaco	✓
pulmones	✓	pauca		esposante	
batillón		aviso	✓	orguloso	✓
zapato	✓	rompido		bizcocho	✓
tergiversar	✓	loro	✓	hacido	
pésimo	✓	granuja	✓	cabello	✓
cadena		estornudar	✓	alegre	✓
hacha	✓	torpe	✓	engatusar	✓
antar		alfombra	✓	temblo	
cenefa	✓	rebuscar	✓	polvoriento	✓
asesinato	✓	cadallo		pemición	
hellar	✓	canela	✓	hervidor	✓
yunque	✓	cuchara	✓	cintra	
regar	✓	jilguero	✓	yacer	✓
abracar		martillo	✓	átar	✓
floroso		cartinar		tiburón	✓
arsa		ladrón	✓	frondoso	✓
brecedad		ganar	✓	tropaje	
ávido	✓	flamida		hormiga	✓
capillo		candado	✓	pozo	✓
lacayo	✓	camisa	✓	empirador	
lampera		vegada		guante	✓
látigo	✓	fomentar	✓	escuto	
bisagra	✓	nevar	✓	laúd	✓
secuestro	✓	musgo	✓	barato	✓
acutación		tacaño	✓	grodó	
merodear	✓	plaudir		acantilado	✓
decar		besar	✓	prisa	✓
alardio		matar	✓	clavel	✓

Por favor, escriba cada dígito en el orden inverso al que apareció.

Tan solo teclee los números, uno por uno.

Hasta ahora ha escrito lo siguiente:

6, 7, 1

15

We used a lexical decision task called the Lextale-Esp (Izura et al. 2014) to estimate vocabulary size as a rough measure of proficiency, and we measured working memory with a backward digit span task (Wechsler 1997), in which participants see a series of numbers and then must recall the numbers in the opposite order, as well as collecting linguistic background information.

More info:

Because we were interested in examining individual variation, we also assessed their vocabulary size using a lexical decision task called the Lextale-Esp (Izura et al. 2014). This task presents 60 real Spanish words and 30 plausible non-words and asks participants whether they are familiar with the word, producing a score that includes an adjustment to control for guessing. Although language proficiency is a complex and multi-faceted phenomenon that cannot be reduced to a single measure, vocabulary size can be taken as one reasonable correlate, so we use it as a stand-in for a more complex measure.

We also measured working memory with a backward digit span task (Wechsler 1997).

In this task, participants see a series of numbers and then must recall the numbers in the opposite order. We chose this task because it is especially apt for testing memory as related to language processing in the context of islands. Some early attempts to correlate working memory with island effects were criticized for using “simple span tasks which do not include both storage and processing components” (Pham et al. 2020:4), but the backward digit span task includes both recall and a processing operation.

Finally, we collected demographic and language history information with a questionnaire.

We excluded participants who demonstrated low Spanish proficiency and those deemed 'non-cooperative.'



16

Participants were all simultaneous or early sequential bilinguals who reported exposure to both English and Spanish before age 8, Spanish-speaking caregivers in childhood, spending the majority of their time before age 18 in the US, and living in the US now.

We excluded anyone with Spanish proficiency below what the Lextale's creators identified for beginning L2 learners, as well as participants who Tom Juzek calls 'non-cooperative' for various reasons. I'm happy to go over the specifics later, but the key fact is that we excluded participants who did not or could not complete the task appropriately, and we had 60 participants after all exclusions.

More info:

Beyond not meeting the inclusion criteria, we excluded participants for three reasons:

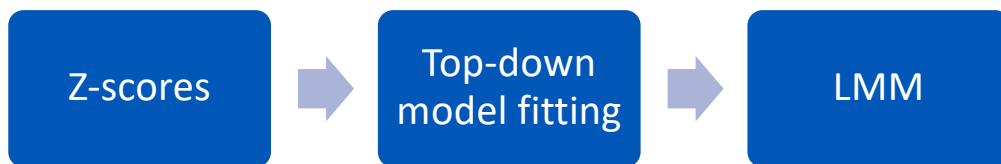
- Those whose Lextale-Esp score was below what Izura et al. (2014) report for beginning L2 learners, suggesting Spanish proficiency too low to complete the task successfully = 43
- 'Non-cooperative' participants who responded to more than 20% of sentences

below a reasonable threshold for the time required to read the sentence and make a judgment (calculated as 1200 ms, following Häussler & Juzek 2021, Juzek 2016) = 1

- Participants who did not complete the task correctly, as indicated by:
 - mean ratings for the ungrammatical filler sentences at the midpoint of the scale (4) or higher (following Pañeda & Kush 2022) = 17
 - ratings on the opposite side of the scale for two or more of three clear attention check items (which have clear ratings of 1 or 7, what Juzek 2016 calls “booby-trap items”) = 24
 - Those who failed two of four ‘instructional manipulation checks’ (i.e., “please select 3”) = 2

After all exclusions, 60 participants remained in the sample. Thirty were female, 27 male, and 3 non-binary or another gender identity. Their mean age was 29.8 years (range: 18-60).

We fit a series of linear mixed-effects models to test for group-level effects.



17

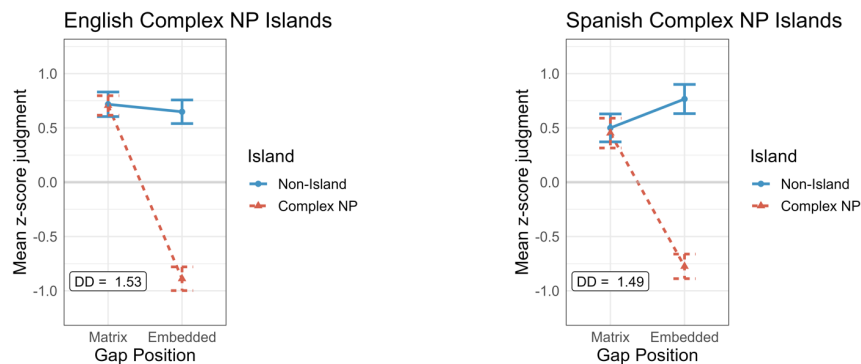
We z-score transformed the raw ratings by participant (Schütze & Sprouse 2013) to address scale compression and skew.

For each island (CNP and *whether*) we examined each 2x2 comparison of Gap x Island separately. Then, to test whether the patterns were different in the two languages, we fit a model with a three-way interaction: Gap x Island x Language.

In each case, we fit the model with a top-down model-fitting procedure that results in the maximal random effects structure that fits the data, removing random effects that explained the least variance until the model converged.

We also calculated a differences-in-differences (DD) score as a measure of effect size (Maxwell & Delaney 2004).

We observe significant interactions between Gap Position and Island and very large effect sizes for complex NP islands in both languages.



18

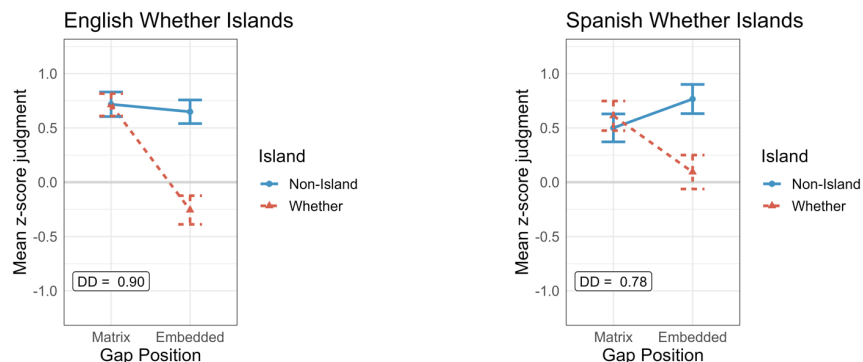
The Complex NP islands are uncontroversially ungrammatical in both languages and serve as a baseline of what an island effect looks like for these participants with these materials. These results demonstrate that participants are sensitive to grammatical restrictions in each language.

Each two-way interaction is statistically significant, suggesting a group-level island effect for each language.

The three-way interaction is not significant, suggesting that the patterns are not different in the two languages.

Kush et al. (2018) characterize DD scores above 1 as 'large' effects, so these are quite large effects.

We observe significant interactions between Gap Position and Island and effect sizes typical of island effects for *whether* islands in both languages.



19

Each two-way interaction is statistically significant, suggesting a group-level island effect for each language with *whether* islands.

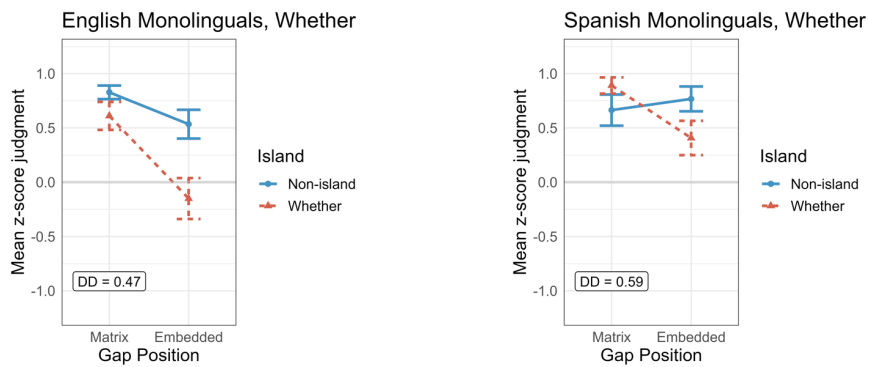
The three-way interaction is not significant, suggesting that the patterns are not different in the two languages.

Kush et al. (2019:401) note that “DD scores for island effects typically fall within the range of 0.75–1.25,” as these effects do.

We thus observe an island effect in English and in Spanish for *whether* islands.

Yet it is noteworthy that in both languages these effects are weaker than the CNP islands, with ratings of the violation sentences hovering around the midpoint of the scale. These resemble “subliminal islands” in Brazilian Portuguese (Almeida 2014) with ratings that suggest mild degradation rather than stark rejection.

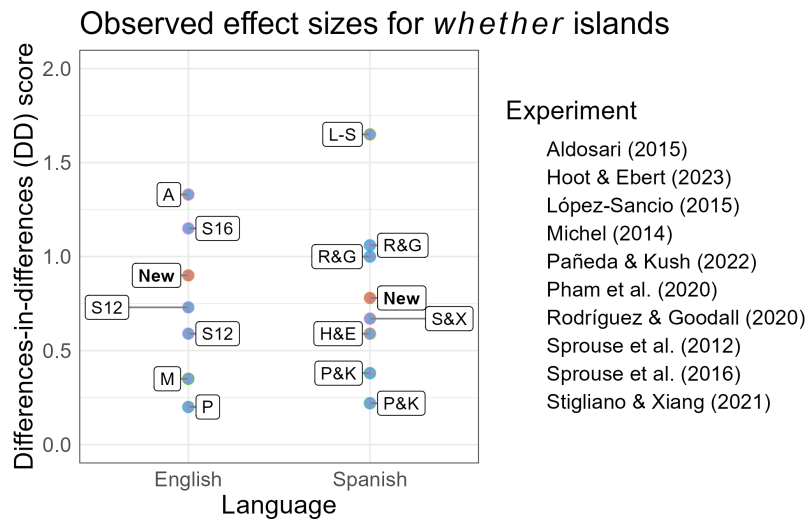
To rule out effects of bilingualism, we compared our speakers to monolingual speakers; the patterns are largely the same.



20

We compared our participants to monolingual speakers of each language and found basically the same pattern, which we can explore in more detail in the Q&A.

Our results add another data point to the picture of Spanish and English *whether* islands.



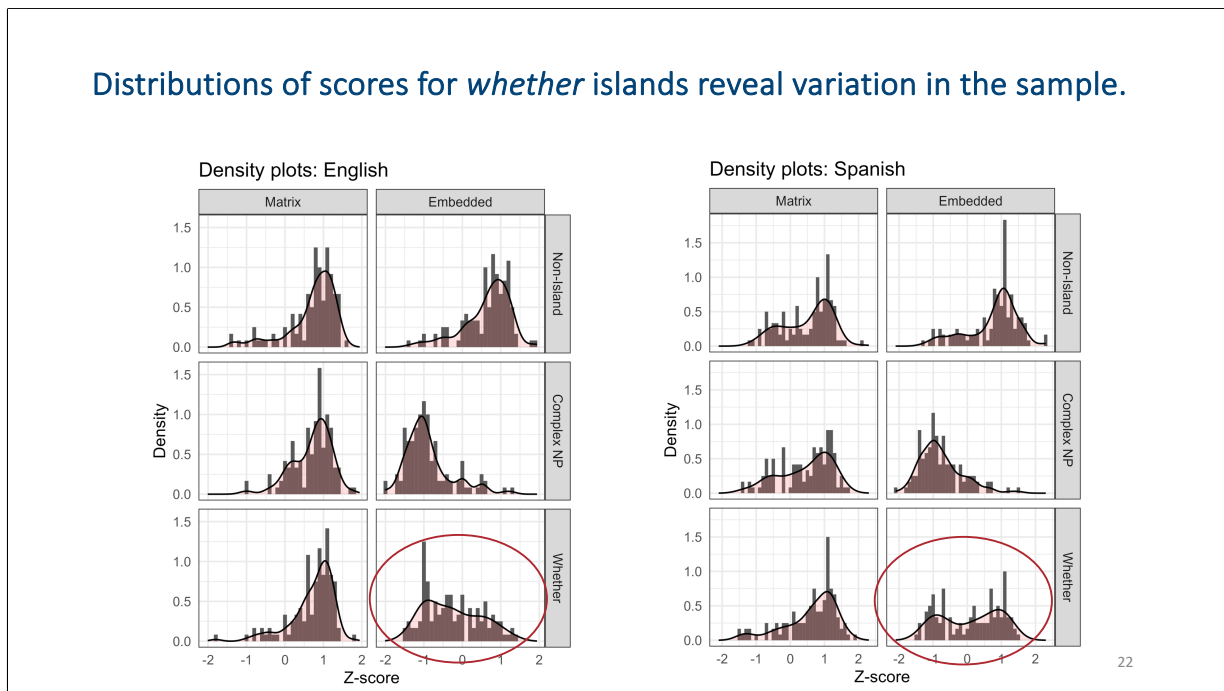
21

Our results fit right in the middle of the previous ranges of results. These findings have the advantage of comparing English and Spanish directly with the same materials and with the same speakers, controlling for some sources of inter-experiment variation and allowing for direct comparisons.

Overall, we find significant effects of similar magnitude for Spanish and English, suggesting that, whatever the source of *whether* island effects, they are not different between the two languages.

We also contribute a new data point to the range provided by existing studies. As we as a field continue to refine our estimates of these effects in different samples, each new data point contributes to a clearer picture of what the underlying population effect is.

Distributions of scores for *whether* islands reveal variation in the sample.



We could stop there, but given the variation we have observed previously, we also asked whether we would see variation in our current sample. To this end, we can examine “second-order acceptability effects” (Kush et al. 2019), including the scores’ distribution by group and individual.

First, we examine the distribution of the scores at the group level by examining histograms (with overlaid density plots) of the z-scores for each condition.

We observe substantial variation for the *whether* islands in both languages, but less for any other condition. Therefore, although our group-level results for each language show mean ratings around the midpoint of the scale, these ratings do not reflect participants agreeing upon generally middling acceptability; rather, they reflect substantial *variation* in the judgments.

Also, it is notable that in the Spanish data we observe some variation in the always grammatical matrix extraction conditions, with more rejections than we expected for those items.

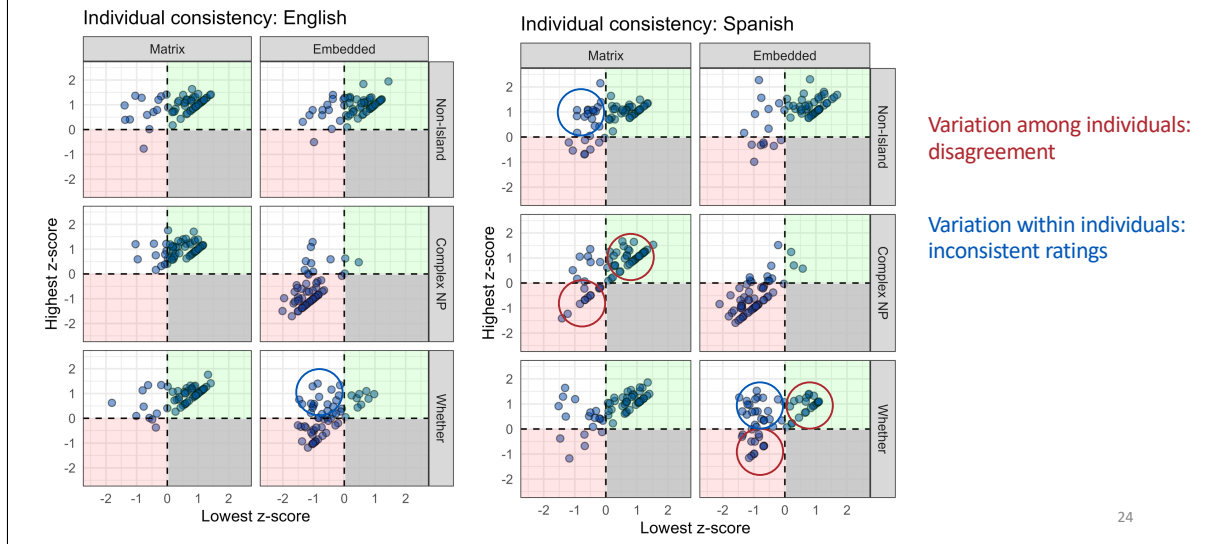
Variation has at least two sources.



23

Variation in judgments could be between individuals – some people accept the islands and some don't – or within individuals – people rate them inconsistently.

Individual consistency plots reveal both inter- and intra-individual variation.



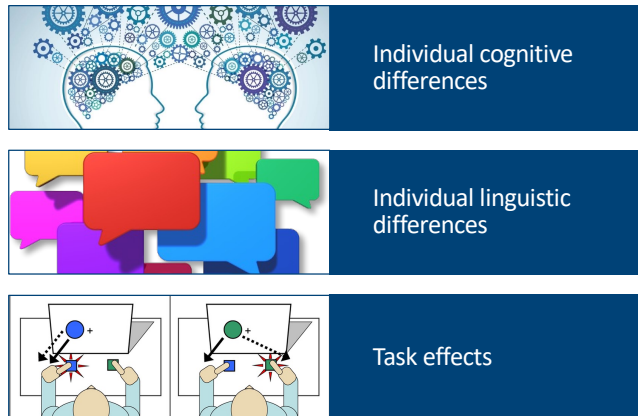
Variation in judgments could be between individuals – some people accept the islands and some don't – or within individuals – people rate them inconsistently. To identify the source of variation, we examine individual consistency by plotting each person's highest rating against their lowest rating (following Pañeda & Kush 2022).

Each person gave two ratings for each condition. Those with z-scores above 0 for both sentences in a condition—consistent acceptors—appear in the upper right (green) quadrant. Those with z-scores below 0 for both sentences in a condition—consistent rejectors—appear in the lower left (red) quadrant. Those who split their ratings—inconsistent raters—appear in the upper left (white) quadrant.

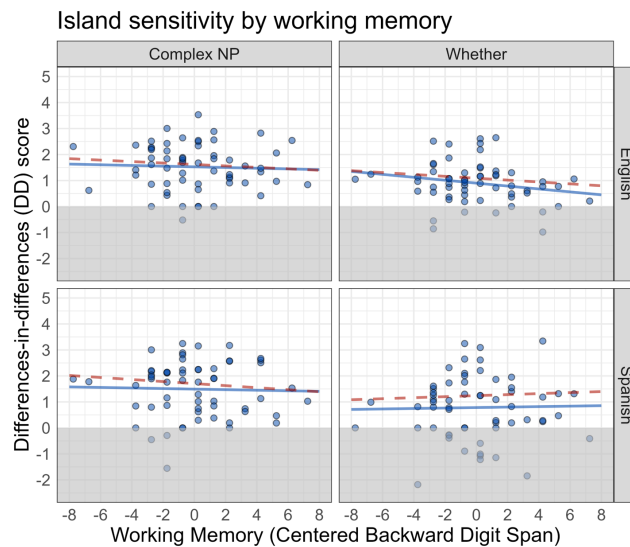
Note that we observe variation both between individuals and within individuals.

It is also noteworthy that some participants are inconsistent in their ratings of non-island structures as well.

We examined possible causes of variation.



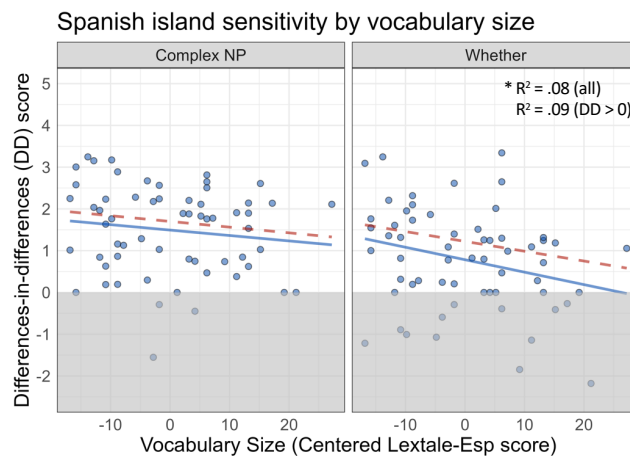
Regressions reveal no relationship between individual working memory and sensitivity to either island effect in either language.



To investigate whether individual differences correlated to working memory scores, we carried out a linear regression for each condition. We ran it twice, once with all the scores (blue solid line) and once with scores below 0 removed (red dashed line), following Pham et al. (2020).

None of these regressions are statistically significant.

Regressions reveal a small effect of individual vocabulary size on sensitivity to *whether* islands, but no relationship for CNP islands.



27

We also examined the effect of vocabulary size in Spanish as a proxy for global proficiency. We only collected a vocabulary size measure for Spanish.

To investigate whether individual differences correlated to vocabulary size scores, we carried out a linear regression in each condition. We ran it twice, once with all the scores (blue solid line) and once with scores below 0 removed (red dashed line), following Pham et al. (2020).

We observe no significant effect of vocabulary size on CNP island sensitivity.

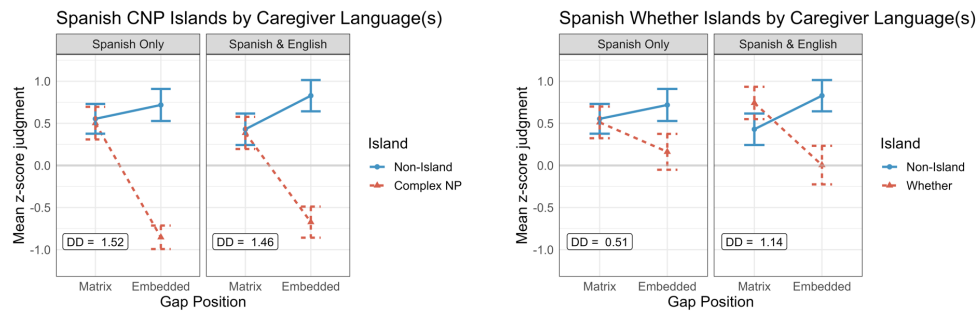
The regression for the *whether* islands is statistically significant ($p = .04$) and produces a medium effect ($R^2 = .08$). The version with only $DD > 0$ is marginally significant ($p = .054$) but slightly larger ($R^2 = 0.09$).

(For reference, Cohen (1992, cited in Larson-Hall 2010) defined effect sizes for R^2 as small for $R^2 = .01$, medium for $R^2 = .09$, and large for $R^2 = .25$. In the context of L2 research, Larson-Hall (2010) calls $R^2 \geq .25$ large and $R^2 = .01 - .05$ small.)

The possibility that *whether* island sensitivity in Spanish varies across individuals as a

function of vocabulary size or global proficiency is suggestive but far from conclusive with this level of evidence.

Language(s) spoken by caregivers may affect the sensitivity to Spanish *whether* island effects, with weaker effects for Spanish only.



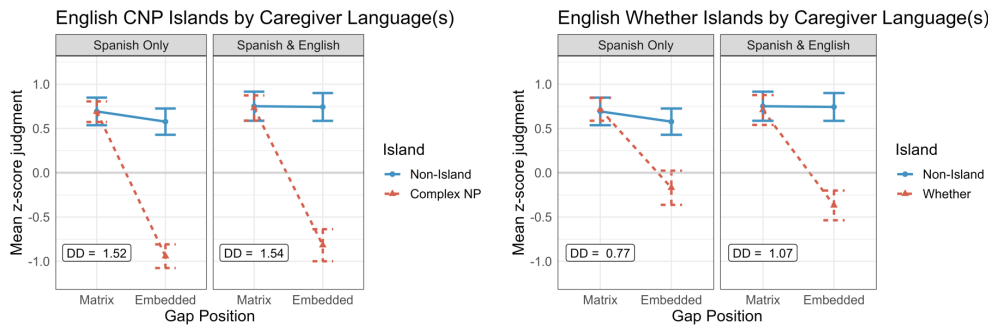
28

The only other individual background characteristic that appeared to have an effect on island judgments was the language spoken by caregivers.

Comparing the interactions for those whose caregivers spoke only Spanish and those whose caregivers spoke both languages, we found that the three-way interaction is not significant for Complex NP islands but it is for *whether* ($p = .03$). Post hoc tests suggest that all four points differ from one another for the Spanish & English group, but the Spanish-only group does not differ on matrix extraction or between the two non-island cases.

Nevertheless, both groups show the same pattern. It may be that difference in the matrix extractions for the Spanish & English group is partially driving this effect, because both groups have significant differences between matrix and embedded extraction in island cases, and between non-island and islands within the embedded cases.

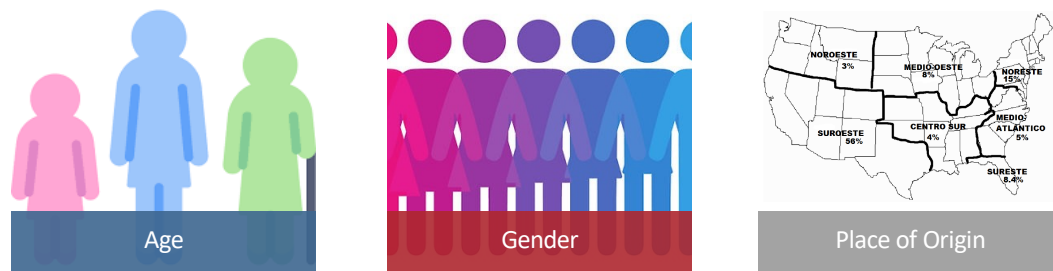
Language(s) spoken by caregivers did not affect the sensitivity to English island effects.



29

On the other hand, the language(s) of caregivers did not affect English judgments. Three-way interactions here were not significant.

Variation is not correlated to any other linguistic background factors we collected, although detailed dialectal comparisons were not possible.

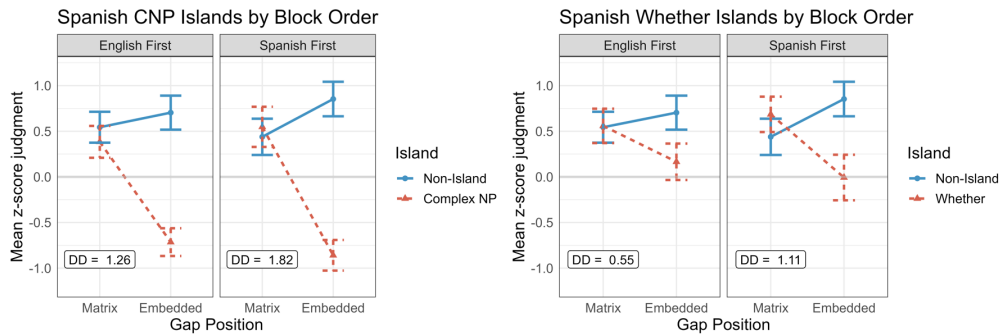


30

We examined DD scores by age, gender, and self-reported answer to “Where did you grow up?” found no relationships. We collected some data about participants’ dialects by asking about caregiver birthplaces, but the range of answers did not permit for a detailed comparison by dialect. Nevertheless, we are aware of no evidence that island effects vary by dialect in Spanish.

For the place of origin, we grouped participants by region (using Escobar & Potowski’s 2015 regional map, displayed here). About half the participants came from the Southwest, while the rest were distributed among several regions, so we could only compare Southwest vs. Not Southwest. We observed no apparent differences.

Presentation order may affect sensitivity to Spanish island effects.



31

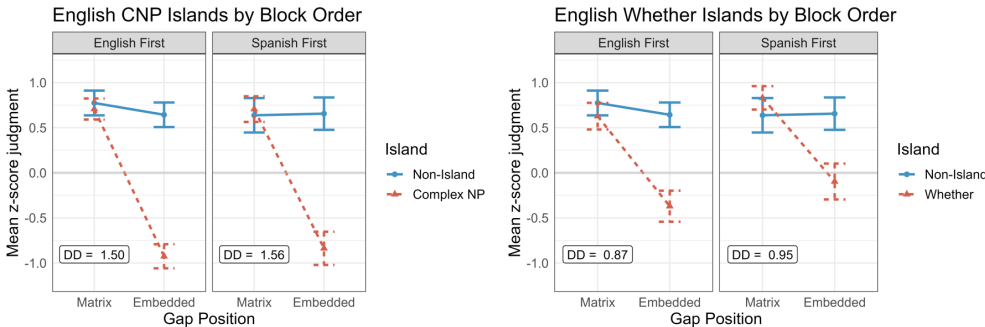
Finally, we can consider task effects as a possible source of variation, comparing those who saw the English block first to those who saw the Spanish block first.

For the two island types in Spanish, the three-way interaction is significant for Complex NP islands ($p = .04$). Post hoc tests suggest those who got English first did not differ in their judgments of non-island cases (blue line), whereas those who got Spanish first judged non-island extraction from embedded clauses significantly better than non-island extraction from matrix clauses, contrary to any expectations. The distributions are otherwise the same, suggesting that this difference is driven by the control cases, not by different ratings of the island violation.

The three-way interaction is not significant for *whether* but is marginal ($p = .06$). Examining the graph suggests that this could be driven by the control cases as well.

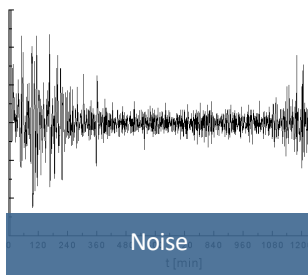
The DD scores are larger for those who saw the Spanish block first, but we should be cautious in interpreting these effects.

Presentation order did not affect the sensitivity to English island effects.



The three-way interaction is not significant for either island type.

Variation raises questions about the nature of variation in judgment experiments that the field has not resolved.



33

Schütze & Sprouse (2013) identify three possible interpretations of individual variation: variation as noise, the baseline assumption in experimental psychology and position taken by Cowart (1997) for judgment studies; variation as dialect or idiolect, reflecting different grammars among individuals; and variation as disconfirmation, taking the strong view that true grammatical effects should be unambiguous.

Experimental syntax often implicitly assumes the first option, but the field has not arrived at a consensus.

Note also that there is likely a difference between within- and between-individual variation, which may need to be approached in a different way. We may need to assign different sources to different types of variation.

Our results here bear on that discussion because we see that much of what we observe when it comes to whether islands appears to be driven by both inter- and intra-individual variation, rather than a consensus among participants that these items are rated somewhere in the middle. Yet most work on whether islands (and other syntactic structures) does not report the individual variation.

Considering the variation we observe overall, we should also note that the Spanish data we've collected here is rather noisier than our previous results with monolinguals, which could be a result of when testing bilinguals in what may be their weaker language, entirely separate from the issue of cross-linguistic influence from English. Furthermore, there is also more noise in English, which is presumably their stronger language. Could this be a result of being in bilingual mode? Or is it merely the result of the poor quality of crowd-sourced data collected via the Internet, despite our extensive controls to weed out non-cooperative participants?

The point is that we don't know, because this is an aspect of experimental syntax that hasn't received enough attention. While we cannot offer a solution today, our intention in highlighting this issue is to make it salient as a way to work toward eventual resolution.

In summary, holding constant materials and participants, *whether* islands produce island effects in both Spanish and English, but with variation.

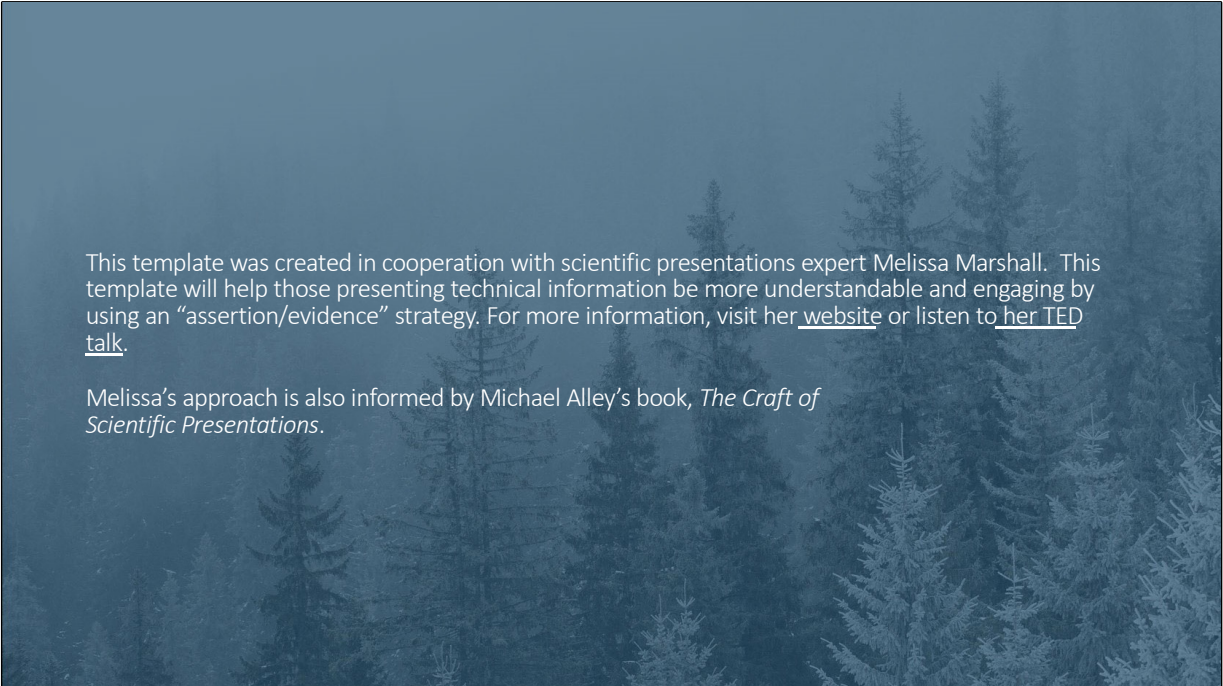
- Effects are of similar magnitude and the same pattern across languages.
- Cross-linguistic influence cannot be ruled out, but monolinguals pattern the same.
- Two proxies for Spanish proficiency (vocabulary size and caregivers only using Spanish) tentatively correlate with lower effect sizes for *whether* islands.
- The variation we observe raises questions about the interpretation of judgment experiments that the field has not resolved.

tinyurl.com/LSRL2023



References

- Aldosari, Saad. 2015. The Role of Individual Differences in the Acceptability of Island Violations in Native and Non-native Speakers. University of Kansas Dissertation.
- Almeida, Diogo. 2014. Subliminal wh-islands in Brazilian Portuguese and the consequences for syntactic theory. *Revista da ABRALIN* 13(2). 55–93. doi:10.5380/rabi.v13i2.39611.
- Boeckx, Cedric. 2012. *Syntactic islands*. Cambridge: Cambridge University Press.
- Bresnan, Joan. 1976. On the form and functioning of transformations. *Linguistic Inquiry* 7. 3–40.
- Chomsky, Noam. 1973. Conditions on transformations. In Stephen R. Anderson & Paul Kiparsky (eds.), *A Festschrift for Morris Halle*, 232–386. New York: Holt, Rinehart and Winston.
- Chomsky, Noam. 1986. *Barriers*. (Linguistic Inquiry Monographs 13). Cambridge, Mass: MIT Press.
- Chomsky, Noam, Ángel J. Gallego & Dennis Ott. 2019. Generative Grammar and the Faculty of Language: Insights, Questions, and Challenges. *Catalan Journal of Linguistics*. 229. doi:10.5565/rev/catjl.288.
- Citko, Barbara. 2016. *Islands*. Oxford: Oxford University Press. doi:10.1093/obo/9780199772810-0101.
- Cowart, Wayne. 1997. *Experimental syntax: Applying objective methods to sentence judgements*. Thousand Oaks, Calif.: Sage.
- Escobar, Anna Maria & Kim Potowski. 2015. *El español de los Estados Unidos*. Cambridge: Cambridge University Press.
- Francom, Jerid. 2012. Wh-movement: Interrogatives, exclamatives, and relatives. In José Ignacio Hualde, Antxon Olarrea & Erin O'Rourke (eds.), *The handbook of Hispanic linguistics*, 533–556. West Sussex: Wiley-Blackwell.
- Grosjean, François. 1998. Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition* 1(2). 131–149. doi:10.1017/S136672899800025X.
- Häussler, Jana & Tom S. Jurek. 2021. Variation in participants and stimuli in acceptability experiments. In Grant Goodall (ed.), *The Cambridge Handbook of Experimental Syntax*, 97–117. Cambridge University Press. doi:10.1017/9781108569620.005.
- Hoot, Bradley & Shane Ebert. 2023. Individual variation in locality effects: Experimental evidence from Spanish wh-islands. Paper presented at the Locality in Theory, Processing, and Acquisition, Philadelphia. <https://tinyurl.com/Locality23>.
- Huang, C. T. James. 1982. Move WH in a language without WH movement. *The Linguistic Review* 1(4). doi:10.1515/tlir.1982.1.4.369. <https://www.degruyter.com/document/doi/10.1515/tlir.1982.1.4.369/html>.
- Izura, Cristina, Fernando Cueto & Marc Brysbaert. 2014. Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicologica: International Journal of Methodology and Experimental Psychology* 35(1). 49–66.
- Jurek, Tom. 2016. *Acceptability Judgement Tasks and Grammatical Theory*. Oxford: University of Oxford Dissertation.
- Kayne, Richard S. 1984. *Connectedness and binary branching*. (Studies in Generative Grammar 16). Dordrecht, Holland; Cinnaminson, N.J., U.S.A.: Foris Publications.
- Kush, Dave, Terje Lohndal & Jon Sprouse. 2018. Investigating variation in island effects: A case study of Norwegian wh-extraction. *Natural Language & Linguistic Theory* 36(3). 743–779. doi:10.1007/s11049-017-9390-z.
- Kush, Dave, Terje Lohndal & Jon Sprouse. 2019. On the island sensitivity of topicalization in Norwegian: An experimental investigation. *Language* 95(3). 393–420. doi:10.1353/lan.2019.0051.
- Larson-Hall, Jennifer. 2010. *A guide to doing statistics in second language research using SPSS*. New York/London: Routledge.
- López-Sancio, Sergio. 2015. Testing syntactic islands in Spanish. Vitoria-Gasteiz: Universidad del País Vasco/Euskal Herriko Unibertsitatea MA Thesis.
- Maxwell, Scott E. & Harold D. Delaney. 2004. *Designing experiments and analyzing data: A model comparison perspective*. Second edition. Mahwah, New Jersey: Lawrence Erlbaum.
- Michel, Daniel. 2014. Individual Cognitive Measures and Working Memory Accounts of Syntactic Island Phenomena. San Diego, Calif.: University of California, San Diego Dissertation.
- Ortega-Santos, Iván, Lara Reglero & Jon A. Franco. 2018. Wh-islands in L2 Spanish and L2 English: A Poverty of the Stimulus and Data Assessment. *Fontes Linguae Vasconum* 126. 435–471. doi:10.35462/FLV126.7.
- Pañeda, Claudia & Dave Kush. 2022. Spanish embedded question island effects revisited: an experimental study. *Linguistics* 60(2). 463–504. doi:10.1515/ling-2020-0110.
- Pañeda, Claudia, Sol Lago, Elena Vares, João Veríssimo & Claudia Felser. 2020. Island effects in Spanish comprehension. *Glossa: a journal of general linguistics* 5(1). 21. doi:10.5334/gigl.1058.
- Pham, Catherine, Lauren Covey, Alison Gabriele, Saad Aldosari & Robert Fiorentino. 2020. Investigating the relationship between individual differences and island sensitivity. *Glossa: a journal of general linguistics* 5(1). 94. doi:10.5334/gigl.1199.
- Rizzi, Luigi. 1990. *Relativized minimality*. (Linguistic Inquiry Monographs 16). Cambridge, Mass: MIT Press.
- Rodríguez, Alejandro & Grant Goodall. 2020. On the universality of wh-islands: Experimental evidence from Spanish. Poster. Paper presented at the Linguistic Symposium on Romance Languages, Austin.
- Ross, John Robert. 1967. *Constraints on variables in syntax*. Massachusetts Institute of Technology Dissertation. <https://dspace.mit.edu/handle/1721.1/151366>.
- Schütze, Carson T. & Jon Sprouse. 2013. Judgment data. In Robert J. Podesva & Devyani Sharma (eds.), *Research methods in linguistics*, 27–50. Cambridge: Cambridge University Press.
- Sprouse, Jon, Ivano Caponigro, Ciro Greco & Carlo Cecchetto. 2016. Experimental syntax and the variation of island effects in English and Italian. *Natural Language & Linguistic Theory* 34(1). 307–344. doi:10.1007/s11049-015-9286-8.
- Sprouse, Jon & Sandra Villata. 2021. Island Effects. In Grant Goodall (ed.), *The Cambridge Handbook of Experimental Syntax*, 227–257. Cambridge: Cambridge University Press. doi:10.1017/9781108569620.010.
- Sprouse, Jon, Matt Wagers & Colin Phillips. 2012. A test of the relation between working-memory capacity and syntactic island effects. *Language* 88(1). 82–123. doi:10.1353/lan.2012.0004.
- Stigliano, Laura & Ming Xiang. 2021. Experimental Evidence on Island Effects in Spanish Relative Clauses. *Probus* 33(2). 271–296. doi:10.1515/prbs-2021-0008.
- Suñer, Margarita. 1991. Indirect questions and the structure of CP: Some consequences. In Hector Campos & Fernando Martínez-Gil (eds.), *Current studies in Spanish linguistics*, 283–312. Washington, DC: Georgetown University Press.
- Szabolcsi, Anna & Terje Lohndal. 2017. Strong vs. Weak Islands. In Martin Everaert & Henk C. van Riemsdijk (eds.), *The Wiley Blackwell Companion to Syntax, Second Edition*, 1–51. Hoboken, NJ, USA: John Wiley & Sons, Inc. doi:10.1002/9781118358733.wbcsyncom008.
- Torrego, Esther. 1984. On inversion in Spanish and Some of Its Effects. *Linguistic Inquiry* 15(1). 103–129. doi:10.2307/4178369.
- Wechsler, David. 1997. *Wechsler Memory Scale*. 3rd ed. San Antonio, Tex.: Psychological Corporation.



This template was created in cooperation with scientific presentations expert Melissa Marshall. This template will help those presenting technical information be more understandable and engaging by using an “assertion/evidence” strategy. For more information, visit her [website](#) or listen to [her TED talk](#).

Melissa’s approach is also informed by Michael Alley’s book, *The Craft of Scientific Presentations*.