SUBLIMINAL WH-ISLANDS IN BRAZILIAN PORTUGUESE AND THE CONSEQUENCES FOR SYNTACTIC THEORY.

Diogo ALMEIDA

Language, Mind and Brain Lab - Division of Science, Psychology, NYU - Abu Dhabi.

ABSTRACT

Categorical acceptability judgments form an important and productive heuristic that provides a substantial body of data to theoretical linguists. Despite its popularity, however, they might not always provide an accurate representation of the acceptability facts, especially when it concerns complex patterns of judgments across a range of different sentences types. In this work, I present evidence that, when categorical acceptability is substituted by a more graded measure of acceptability, one can observe wh-island sensitivity in Brazilian Portuguese in three syntactic phenomena (wh-movement, Topicalization and Left Dislocation), even though the island violating structures are marginally or fully acceptable. I conclude with a discussion about what the existence of such island sensitivity effects in marginally or fully acceptable sentences could mean for theories of syntactic islands, and syntactic theory more broadly construed.

RESUMO

Julgamentos categoriais de aceitabilidade são uma heurística importante e produtiva que provê um corpo empírico substancial à teoria linguística. Entretanto, a despeito de sua popularidade, esses julgamentos nem sempre refletem de maneira correta os fatos relativos à aceitabilidade de sentenças, principalmente no que diz respeito a padrões complexos de julgamento que cruzam uma gama de diferentes tipos de sentenças. Neste estudo, eu apresento evidências que o Português Brasileiro exibe sensibilidade a ilhas de qu- em três fenômenos sintáticos distintos (movimento de qu-, topicalização e deslocamento à esquerda) quando medidas de aceitabilidade mais granulares substituem julgamento categoriais de aceitabilidade. O artigo

conclui com uma discussão a respeito de como a existência de sensibilidade à restrições de ilhas sintáticas em sentenças marginais ou plenamente aceitáveis pode impactar as teorias de ilhas sintáticas e, de maneira mais abrangente, a teoria sintática.

1   Introduction

Informal acceptability judgments have served as one of the primary sources of data for theoretical syntax (CHOMSKY, 1965; SCHÜTZE, 1996; SPROUSE et al., 2013). A recent large scale survey of ten years of the journal Linguistic Inquiry, for instance, has estimated that 48% of the data used in their theoretical syntax papers from 2001 to 2010 came from simple acceptability judgments, more than two times as much as the second most used source of data, judgments about possible interpretations, which were estimated to compose 23% of the data (SPROUSE et al., 2013). However, despite its prevalence and popularity, informal acceptability judgment collection has persistently been a practice surrounded by controversy.

One such controversy, which is the topic of this paper, pertains to how informal acceptability judgments are generally used to inform theory construction and theory evaluation. This is an issue that has received considerable attention in the literature (e.g., BARD et al., 1996; FEATHERSTON, 2005a; SORACE and KELLER, 2005; SPROUSE et al., 2013), and generally surfaces as a debate about how to interpret

gradience in accceptability judgments. While it is relatively uncontroversial that acceptability ratings are gradient in nature (CHOMSKY, 1965), grammars are generally modeled as categorical objects (KELLER, 2000; KELLER and SORACE, 2003; SORACE and KELLER, 2005; ALEXOPOULOU, 2007). Grammars of this sort can then be used to state whether they could have generated any given string.

This mismatch between the nature of the data and the nature of theoretical objects the data helps motivate is not easily solvable. Taking this into consideration, most researchers proceed by making simplifying assumptions about the relationship between the acceptability and the grammatical status of sentences. A common working conjecture is that if after a putative grammatical manipulation a sentence is judged on a binary scale to be "unacceptable", then we have grounds to think that the grammar has something to say - more specifically, to complain - about the resulting structure.

This paper is organized as follows. Section 2 revisits the two basic research heuristics routinely used by syntacticians – namely, the putative isomorphism between acceptability data and grammaticality, and whether categorical grammars necessitate looking at acceptability judgments in a categorical fashion – and explores the consequences of dispensing with one or both heuristics. It will be argued that when acceptability data is not forced into a categorical binary scale, interesting data patterns can be observed that could have consequences for current theories of grammar. In particular, the concept of "subliminal island effects" will be presented in light of a discussion of gradient vs categorical acceptability judgments.

Section 3 will present the case of syntactic island phenomena - more specifically, wh-islands - and use it to illustrate the fact that both English – which is described as obeying wh-islands - and Brazilian Portuguese (BP) – which is generally described as not subject to wh-islands – yield essentially the same *kind* of evidence for wh-island sensitivity when their

pattern of gradient acceptability is analyzed without focusing primarily on whether they cross the tenuous and poorly defined boundary between "acceptable" and "unacceptable" categories.

Section 3 concludes by presenting further evidence that island-sensitivity can be decoupled from binary categorical judgments of acceptability and be observed even in cases where the strings under evaluation are judged to be "acceptable". In particular, evidence will be presented that certain topic constructions in BP, which have been generally analyzed as base-generated exactly on the basis of their insensitivity to island configurations, do in fact show evidence of wh-island sensitivity. Crucially, this is an island constraint that had been posited *not* to operate in BP to begin with.

Section 4 concludes with some considerations about what the existence of these "subliminal islands effects" might mean for a theory of grammar.

2   Acceptability vs Grammaticality

Before proceeding, it is important to offer a working definition of acceptability and grammaticality that will be used throughout this paper. Henceforth, when we refer to the *acceptability* of a sentence, we mean the percept that a native speaker can form when she hears or reads an utterance, much like *timbre* is a perceptual attribute that listeners can experience upon hearing sounds. The *grammaticality* of a sentence, on the other hand, refers to a theoretical claim. More specifically, claiming that a sentence is grammatical is equivalent to claiming that (i) there is a formal object (a grammar) that can generate the specific utterance with its intended meaning and (ii) this formal object is somehow part of/implemented in the mental makeup of the native speaker. It is generally assumed that sentence *acceptability* offers some insight as to its *grammatical* status, under the reasoning that utterances that conform to the mental

grammar of the native speaker would probably sound natural/acceptable as sentences of her language, whereas utterances that violate it would probably sound degraded (see MARANTZ, 2005 and HOJI, 2010 for discussions of this assumption).

The precise nature of the relationship between acceptability and grammaticality is however still very much a mystery. What is clear is that the assumption that sentences should sound acceptable if and only if they are generated by the grammar and should sound degraded otherwise, is falsified by known empirical phenomena. There are sentences that are acceptable and yet considered ungrammatical, like the comparative illusion (cf. PHILLIPS et al., 2011 for review of this and some other cases):

(1)   More people have been to Russia than I have

Conversely, it is also possible to observe sentences that are judged to be unacceptable and yet are considered to be grammatical, like double center embedded structures (CHOMSKY and MILLER, 1963):

(2)   The rat the cat the dog chased killed ate the malt.

These phenomena show that any strong isomorphic thesis between acceptability and grammaticality is empirically untenable. However, a weaker version of the isomorphic thesis could nonetheless be useful as a *heuristic*. In other words, assuming a direct mapping between acceptability and grammaticality is ultimately a faulty strategy, but it can serve, if used discerningly, as a viable working hypothesis for diagnostic purposes, to be revisited and reassessed when the need arises.

2.1 Categorical acceptability patterns vs Gradient acceptability patterns

Given the fact that sentence acceptability is generally used *heuristically* to shed light on questions pertaining to sentence grammaticality, we can ask how the different options of reporting acceptability usually given to native speakers might impact linguistic research.

While it is widely acknowledged that sentence acceptability is a gradient psychological quantity (CHOMSKY, 1965; PHILLIPS, 2009), linguists often prefer to abstract away from that gradience and ask their informants to categorize sentences as "good" or "bad", or as "acceptable" and "unacceptable". This forced categorization onto a binary scale makes the direct mapping easier between data and theory. However, it is important to stress that this practice is yet another *heuristic*, and that this accumulation of heuristics is justified not by a solid theory of acceptability judgments (which does not currently exist; see HOFMEISTER et al., 2013 for discussion) but rather by the amount of progress one can make by judiciously invoking them as working hypotheses.

Therefore, it is an open question whether acceptability facts would change if one were not to rely on either one or both of these heuristics. Crucially, if the description of the syntactic phenomenon changes according to the choice of the heuristics used in data collection and data interpretation, it is important to consider the potential implications for syntactic theory.

2.2 Syntactic islands and graded acceptability judgments

Syntactic islands are a fertile ground for this sort of inquiry, because (a) they involve the interaction of different syntactic mechanisms, (b) they elicit varying degrees of acceptability within languages – to the point that there have been proposals to reduce them to extra-grammatical factors, and (c) they are cross-linguistically diverse. We turn to these properties next.

2.2.1 Syntactic accounts of islands require two ingredients

Traditional syntactic accounts of island effects generally require two elements: a long distance dependency formation mechanism (e.g., Movement) and a barrier to this mechanism (e.g., an embedded clause headed by a wh- word). The presence of either in a sentence by itself generates no problems. For example, objects can be moved to the front of the clause, and embedded clauses can be headed by a wh- word, as shown in the following examples:

(3)   Who did Mary see?

(4)   Mary wondered whether Bill saw Jane.

However, trying to extract an object out of an embedded clause headed by a wh- word results in a string deemed unacceptable by many native speakers of English:

(5)   *Who did Mary wonder whether Bill saw?

Notice the contrast in the case where the object is extracted out of an embedded clause headed by the complementizer *that*:

(6)   Who did Mary think that Bill saw?

2.2.2 Syntactic islands elicit varying degrees of acceptability

The schema sketched above forms the basic template for several other so-called syntactic island effects. These can be stated as constraints on which kind of constituents can be engaged by a particular long-distance dependency formation mechanism (such as Movement). Examples include the apparent bans on the extraction of an NP (a) out of another NP in subject position, (b) out of a coordinated NP or (c) out of an adjunct clause.

However, it is generally accepted that such restrictions on long-distance dependencies exhibit variations in how degraded the outcomes of their violations are judged to be. For instance, adjunct island effects are reported to sound more degraded than wh-island effects in English, even though both cases generate low acceptability ratings and are generally classified as "unacceptable" when forced onto a binary scale:

Adjunct island:

(7)   *How$_i$ does Peter wonder whether Mary fixed the car t_i?

Wh-island:

(8)   ?What$_i$ does Peter wonder whether Mary fixed t_i?

2.2.3 Syntactic islands vary cross-linguistically

Syntactic islands have also been reported to exhibit substantial cross-linguistic variation. For example, RIZZI (1982) followed by Torrego, (1984), presented data suggesting that Italian and Iberian Spanish are not subject to the same wh-island constraints as English. This cross-linguistic contrast has been captured theoretically by proposing that the inventory of the categories that can act as a barrier to movement can vary across languages. In some, like English, the functional projection IP blocks movement, whereas in other languages, like Italian (RIZZI, 1982), Iberian Spanish (TORREGO, 1984) and BP (MIOTO et al., 2000; MIOTO e KATO, 2005), it is CP that acts as a barrier. This has been a tremendously influential proposal, and it is predicated on contrasts like the one below (cf. MIOTO and KATO, 2005):

English

(9)   What does Peter think that Mary bought t_i?

(10)*What does Peter wonder whether Mary bought t_i?

BP

(11) O que$_i$ o   Pedro achou que   a   Maria comprou t_i?
     What   the Pedro thought that the Mary bought
     What did Peter think that Mary bought?

(12) O que$_i$ o Pedro perguntou se   a   Maria   comprou t_i?
     What   the Pedro asked whether the Mary   bought
     What did Peter ask whether Mary bought?

2.2.4 Different possible diagnoses of syntactic islands.

As an example of how looking at patterns of gradient acceptability might change the description of the basic cross-linguistic facts, we can turn to wh-islands in English and BP. Despite being described as generally acceptable – and therefore grammatical – in BP, sentences that violate the wh-island constraint do not sound *perfectly* acceptable, and a range of acceptability judgments can be elicited. It might be the case that, if forced to use a binary scale, sentences that violate wh-islands are judged just above the tenuous and ill-defined threshold of the binary category "acceptable", while their English counterparts are judged to be below the same threshold. This forces us to consider an important question: How exactly does one define what counts as syntactic island *sensitivity*?

The usual strategy for defining island sensitivity is whether the offending cases are categorized as "bad" or "unacceptable" on a binary scale. As we have argued above, however, this type of reasoning is better defined as a research *heuristic*, and therefore it is ultimately unsuitable as a definitive diagnostic of islandhood. Once this way of defining islands is called into question, one can devise alternative means of diagnosing the conditions under which certain kinds of long distance dependency formation mechanisms (like Movement) are blocked.

The factorial definition of islands, and the challenges from reductionist processing theories of islands

An alternative way of diagnosing syntactic islands has been proposed in the context of a debate about theories of island effects that sought to reduce them to extra-grammatical factors.

The simplest theory of this kind would propose that at least some island sensitivity effects might be better understood as the result of a conspiracy of independent constraints on parsing, rather than a grammatical constraint. Under such a proposal, the two ingredients of island effects - the presence of movement and the presence of an island configuration (for example, an embedded clause headed by a wh-word) - might have independent *parsing* (or *processing*) costs. The sentences below illustrate the paradigm for wh-island in English (diacritics indicating acceptability intentionally left out):

Movement out of matrix clause (no island structure/island structure)

(13) Who (thinks that/wonders whether) Mary read the book?

Movement out of embedded clause, (no island structure/island structure)

(14) What does John (think that/wonder whether) Mary read?

Figure 1 illustrates the different scenarios under which the length of movement and the presence of an island structure both have verifiable independent costs. The gray line in the middle of the plot illustrates the hypothetical boundary between "acceptable" and "unacceptable" categories in the traditional binary scale, and the lighter-gray area around it represents the area in which judgments might be categorized as either acceptable or unacceptable, depending on the person, and test situations.

The offending sentence (ie, the one that incurs in both independent costs simultaneously) is marked in red. The top row of Figure 1 illustrates the cases in which the two independent costs add up linearly. The top left plot shows what the pattern would be in case the two independent processing costs existed, but would still generate sentences that would be rated as "acceptable" in the traditional binary scale. The top middle plot displays the pattern that would be expected if when the two independent costs combined, they would lower the acceptability of the offending sentence just enough for it to be judged "acceptable" or "unacceptable" in the traditional binary scale, although only marginally so. Finally, the top right plot illustrates the case in which the two costs combine linearly and produce an offending sentence that clearly cross the "acceptability" boundary in the traditional binary scale. It is this plot that corresponds to the simplest processing-based account of island effects, since the entirety of the island effect is accounted for by the simple addition of these two independently motivated costs incurred by the parser.

FIGURE 1. Hypothetical scenarios relating the factors of Structure (what kind of embedded clause is present in the sentence) and the site of origin of the moved element (matrix or embedded clause) that would generate an island sensitivity effect.
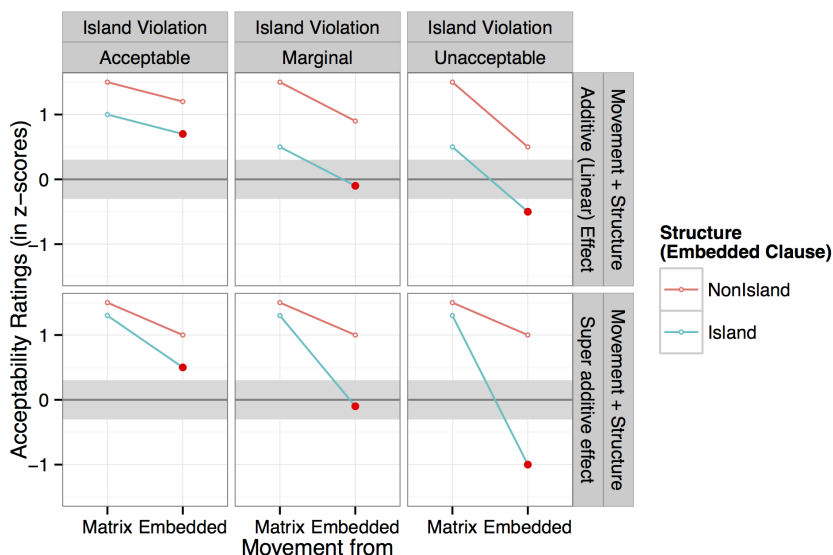
If the pattern of acceptability in English could be described by the top right plot in Figure 1, then the simple processing-based account of island might be a tenable model. However, a large corpus of acceptability judgment experiments has already demonstrated that, by and large, island effects are better described by the bottom right plot in Figure 1 (SPROUSE et al., 2012). In this plot, we see that the two processing costs do exist —although in reality, it is not the case that they do for every island under consideration (see SPROUSE et al., 2012) — but they do not add up linearly. In other words, the severity of the acceptability penalty incurred by the offending sentence is larger than what would have been predicted by each cost separately. This fact is not

predicted by the simple processing-based theory considered so far, but it is compatible with a grammar-based account of islands, in which that particular structure (the one marked in red in Figure 1) is the target of a structural well-formedness constraint.

However, there are other processing-based theories of island effects that can account for the same super additive effect of length of movement (from matrix or embedded clause) and presence of an island structure. For instance, KLUENDER and KUTAS (1993) have argued that some island constraints — the ones subsumed by Subjacency — can be reduced to the conspiracy of these independent processing facts, and no grammatical constraint need be posited.

KLUENDER and KUTAS (1998)'s account starts by assuming that the two ingredients for island effects have independent processing costs. More precisely, these factors are costly to the parser because they tax its working memory resources. If these resources are exceeded, then the parser simply fails to process the sentence, resulting in a sentence of much lower acceptability than it would have been predicted by the encumbrance of each independent cost by itself. This kind of model has been revisited recently by HOFMEISTER and SAG (2010), who explicitly reaffirm that limits on working memory capacity, when exceeded, might lead to the breakdown of the parsing process.

Under this account, however, the same pattern of acceptability as the one posited by grammatical accounts of syntactic islands is predicted (ie., the lower bottom right plot in Figure 1). Finding the test cases that could distinguish one account from the other is not trivial, and the interested reader is referred to the recent series of articles and responses between (SPROUSE et al., 2012a, 2012b) on one side and (HOFMEISTER et al., 2012a, 2012b) on the other for further discussion.

It is important to notice, however, that while both models have the requisite mechanisms that allow them to naturally predict the pattern in the bottom right plot of Figure 1, neither model can readily account for the hypothetical patterns of the bottom left and bottom middle plots.

Under these two scenarios, the ingredients of island effects still interact super additively, but the offending sentence is still fully or marginally acceptable.

These patterns would be problematic for traditional grammatical accounts of islands because these are generally stated as constraints on structural well-formedness. The patterns in the bottom left and middle plots, however, would imply that perhaps some sort of constraint is being applied, but its result is nonetheless judged to be well-formed enough by native speakers.

In the same vein, the patterns in the bottom left and middle plots of Figure 1 would also be problematic for KLUENDER and KUTAS (1993)'s processing-based account. This is because the way the super-additive effect is derived under this kind of model is via a breakdown of the parsing process, a direct result of its working memory resources being exceeded. The effects in the bottom left and middle plots are still super-additive, but they do not result in particularly low acceptability ratings, indicating that these sentences would have probably been parsed well enough to generate natural sounding sentences – especially since under KLUENDER and KUTAS (1993)'s account, there is nothing *structurally ill-formed* about these sentences.

2.2.5 Looking at the gradient acceptability of island phenomena: The concept of *subliminal* island effects.

Once a factorial definition of what constitute an island effect is taken into consideration, and the heuristic adherence to categorical judgments is dispensed with, it is at least logically possible to conceive of the scenarios depicted in the bottom left and middle plots of Figure 1. These are both cases in which length of movement and the presence of an island structure combine super-additively, and yet the final result is still acceptable or marginally so. We will refer to this kind of island sensitivity as *subliminal island effects*, and contrast them with the traditional island effects, which we could refer to as *supraliminal island effects*, as the other side of the same phenomenon.

If island effects can be detected even within structures that are generally categorized as acceptable in the traditional binary scale, this could have consequences for how island-constraints are discussed in the theoretical literature. One of the goals of this paper is to explore the hypothesis that at least some of the cross-linguistic variation observed in the syntactic island literature is only superficial in nature. Put differently, it may be the case that island constraints are indeed universal, and the reason for the apparent cross-linguistic variation is not due to variations in the inventory of well-formedness constraints any given grammar might possess. The apparent variation would rather be due to fact that the acceptability penalties that are incurred when these putative universal constraints are violated simply vary across languages.

If this kind of explanation is on the right track, then it would have two immediate consequences: First, for grammatical accounts of syntactic islands, it would invite significant simplification in the basic description of what an island effect is. Second, it would push the object of explanation away from ontological considerations about grammars, and instead force researchers to consider how a singular set of universal constraints ends up having different acceptability costs across different languages.

For processing-based theories of islands, the existence of subliminal island effects would have even more far-reaching consequences. Fully reductionist theories such as the ones proposed by KLUENDER and KUTAS (1993) and HOFMEISTER and SAG (2010) would be forced to abandon the goal of deriving island effects purely via extra-grammatical factors. This is because the facts of cross-linguistic variation would still point to super-additive effects that simply do not cause the parsing process to break down. If that is the case, then the linking between the two in the traditional *supraliminal island* cases might be simply fortuitous, and not the defining characteristic of what an island effect is. While processing-based accounts would still have plenty of room to posit a significant role for processing effects in the derivation of specific island

effects, if *subliminal island effects* can be documented, they would almost certainly necessitate language-specific, and therefore grammar-specific, constraints to be in place.

3 Wh-Island sensitivity in wh-movement, topicalization and left dislocation in English and BP: Do *whether* clauses in BP create subliminal islands?

As an example of how this hypothesis would work, let us consider the case of wh-island constraints. In some languages, like English, the violation of this constraint generates a large enough acceptability penalty that makes it visible through the lenses of categorical acceptability judgments. Because these cases are visible to the "naked eye" and are more easily categorizable by native speakers, we have proposed referring to them as *supraliminal islands*.

The first hypothesis to be considered in this study is whether in some languages, like BP, the violation of a wh-island constraint would simply be more difficult to observe, because the resulting structure might never receive a large enough acceptability penalty to be easily categorizable as "unacceptable" in the traditional binary categorical scale.

In order to test the viability of such hypothesis, we will compare cases of whether-island structures in English and BP under this new factorial definition and see whether similar acceptability patterns are observed with different end results for the offending sentences ("unacceptable" for English, "acceptable" or "marginally acceptable" in BP).

If subliminal island effects can be detected, then this calls into questions all the cases in which the structural representation of a given phenomenon is decided on the basis of their putative sensitivity to island effects. The second test case evaluated in this study is the distinction between Topicalization (CHOMSKY, 1977) and Left Dislocation (ROSS, 1967), two types of topic constructions that are superficially

very similar, in that both are phenomena in which a phrasal unit (for eg. an NP) occupies the left edge of the clause, and stands in a long distance dependency relation with an element inside of it:

Topicalization/Left Dislocation

(15) That bus$_i$, the professor thinks that Matt missed (t_i/it$_i$).

In the case of Topicalization, it is normally assumed that the element involved in the long distance dependency is the trace of the moved phrase. In Left Dislocation, the element inside the clause is thought to be a co-referential pronoun. Therefore, despite their superficial similarities, it has long been proposed that Topicalization originates from movement, whereas Left Dislocation is a base generated structure. One of the empirical reasons to carve out a different genesis for these two phenomena is their divergent behavior when it comes to island sensitivity. Namely, Topicalization seems to be sensitive to syntactic islands, whereas Left Dislocation seems not to be:

Topicalization

(16) That bus, the professor (thinks that/*wonders whether) Matt missed.

Left dislocation

(17) That bus, the professor (thinks that/wonders whether) Matt missed it.

These facts are compatible with the view according to which movement is constrained by certain barriers (like embedded clauses headed by wh-words), whereas simple anaphoric relationships (which are taken to be at the heart of Left Dislocation) are not.

3.1 Experimental design.

Three subexperiments were embedded into a single linguistic survey, where participants were asked to rate the acceptability of test items on 7-point scale. The first subexperiment compared wh-island cases in English and BP. The second subexperiment compared Left Dislocation in English and BP, and the third subexperiment compared Topicalization in English and BP. All subexperiments had four conditions in order to explore how the factorial decomposition of wh-islands is reflected in gradient acceptability data. The only exception was in English version of the third subexperiment (Topicalization), where only extraction out of object positions were possible, due to English not allowing null subjects. This yielded 10 experimental conditions in total.

3.2 Materials

Ten base lexicalizations were created and modified to yield ten versions of each experimental condition. These were distributed across ten different lists, following a Latin square design. Therefore, every condition in each list contained an experimental item derived from a different base lexicalization, and each participant was only presented with one item per condition, never from the same base lexicalization.

In addition, thirty-two filler sentences spanning the full acceptability spectrum observed in a large scale survey of English data judgments (SPROUSE et al., 2013) were selected as to yield an approximate 1:1 ratio of acceptable vs unacceptable sentences, as in SPROUSE et al., (In Press). The English fillers were used without any modifications

for the English experiment, and were translated into BP for the BP experiment. The author used his native speaker intuition to verify that the approximate 1:1 acceptable/unacceptable ratio was maintained in the translations.

The use of 1:1 acceptable to unacceptable ratio that nonetheless spans the full range of the acceptability spectrum is not strictly necessary from an experimental perspective, but it does introduce a set of desirable properties from a methodological standpoint. The first is that participants are encouraged to use the full range of whatever acceptability scale they are presented with (a 7-point scale in our case), which helps minimizing scale-biases effects. The second desirable property is that a 1:1 acceptable-to-unacceptable ratio of sentences in the experiment provides for a more direct translation between the z-score transformed results (see Analysis below) and their potential binary counterparts, since 0 on the z-score scale would map rather naturally to an area closer to the boundary between "acceptable" and "unacceptable" categories.

Subexperiment 1: English (A) and BP (B) wh-island cases

A full paradigm is shown below for one base lexicalization in English and one in BP (diacritics indicating categorical acceptability omitted):

Movement out of matrix clause; (no island structure/island structure)

(18) Who (thinks that/wonders whether) Matt missed the bus?

(19) Quem (achou que/perguntou se)   o Marcos perdeu o ônibus?
Who (thought that/asked whether) the Marcus missed the bus
Who (thought that/asked whether) Marcus missed the bus?

Movement out of embedded clause; (no island structure/island structure)

(20) What does the professor (think that/wonder whether) Matt missed?

(21) O que que a professora (achou que/perguntou se) o Marcos perdeu?
What that the professor.FEM (thought that/asked whether) the
Marcus missed
What did the professor (think that/ask whether) Marcus missed?

Subexperiment 2: Left Dislocation in English (A) and BP (B).

Both English and BP allow us to test the full factorial paradigm for islands for cases of left dislocation (LD), as shown below (diacritics indicating categorical acceptability omitted):

Pronoun in matrix clause; (no island structure/island structure)

(22) That professor, she (thinks that/wonders whether) Matt missed the bus.

(23) Aquela professora, ela (disse que/perguntou se) o Marcos perdeu o ônibus.
That professor.FEM, she (said that/asked whether) the Marcus missed the bus
That professor, she (said that/asked whether) Marcus missed the bus.

Pronoun in embedded clause; (no island structure/island structure)

(24) That bus, the professor (thinks that/wonders whether) Matt missed it.

(25) Aquele ônibus, a professora (disse que/perguntou se) o Marcos perdeu ele.
That bus, the professor.FEM (said that/asked whether) the Marcus

missed him
That bus, the professor (said that/asked whether) Marcus missed it.

Subexperiment 3: Topicalization in English (A) and BP (B)

When it comes to topicalization, unfortunately English does not allow the full factorial paradigm, because the language does not accept null subjects:

(26) *That professor, *t thinks* that Matt missed the bus.

Because of this, only topicalization of objects will be tested, as in the paradigm below:

Movement out of embedded clause, (no island structure/island structure)

(27) That bus, the professor (thinks that/wonders whether) Matt missed.

Brazilian Portuguese, in contrast with English, allows for null subjects in some circumstances. In general, subjectless matrix clauses are not allowed, unless there is a prominent element in the discourse that it can be coreferent with:

(28) * ø   disse que o   Marcos perdeu o   ônibus.
      NULL said   that the Marcus missed the bus
      Said that Marcus missed the bus.

(29) Aquela professora? Então, ø disse que o Marcos perdeu o ônibus.
     That professor.FEM? So, NULL said that the Marcus missed the bus.
     That professor? Well, she said that Marcus missed the bus.

This allows for an attempt to have the full factorial paradigm in Brazilian Portuguese:

Movement out of matrix clause; (no island structure/island structure)

(30) Aquela professora, ø (disse que/perguntou se) o Marcos perdeu o ônibus.
That professor.FEM, NULL (said that/asked whether) the Marcus missed the bus
That professor, ø (said that/asked whether) Marcus missed the bus.

Movement out of embedded clause; (no island structure/island structure)

(31) Aquele ônibus, a professora (disse que/perguntou se) o Marcos perdeu ø.
That bus, the professor.FEM (said that/asked whether) the Marcus missed NULL.
That bus, the professor (said that/asked whether) Marcus missed ø.

3.3 Task

Participants were instructed to judge the acceptability of sentences on a 7-point scale, ranging from 1 ("very unacceptable") to 7 ("very acceptable"). The 7-point Likert scale judgment task has been shown to be at least as powerful as Magnitude Estimation (WESKOTT e FANSELOW, 2011; SPROUSE et al., 2013), but it is easier to implement and for participants to understand.

3.4   Procedure

The experiment was implemented using the Qualtrics software (Qualtrics, Provo, UT; version: 57336). The experiment began with a 6 sentence practice phase (3 acceptable and 3 unacceptable sentences each), followed the experimental phase. In the experimental phase, there was a 10-sentence adaptation period, in which fillers spanning the full spectrum of acceptability were presented (5 acceptable, 5 unacceptable). This was done to maximize the chance that participants would have used the full range of the 7 point scale before they encountered any of the experimental items. The items in this adaptation phase were always the same across participants, but were presented in random order for each participant. They were not distinguished in any way from the other items in the experimental phase. Following the adaptation phase, the rest of the materials were presented in randomized order for each participant.

Participants completed the survey on their own pace, but were instructed to not overthink their judgments. Participants were asked to judge 44 experimental sentences, in addition to the six practice trials.

3.5   Participants

60 self-reported native speakers of American English were recruited via Amazon Mechanical Turk. 38 self-reported native speakers of BP were recruited via different social media websites and word of mouth.

3.6 Analysis

The 7-point acceptability ratings of the 44 experimental items from each participant were scaled to a standard deviation unit. This procedure, called z-score transformation, is carried out by first subtracting the mean rating of each participant from every item they rated and then subsequently dividing these values by the standard deviation of their raw ratings. The z-score transformation helps to mitigate potential scale biases arising from inconsistencies across participants regarding their use of the 7 point scale. For instance, some participants might have a bias towards using the lower or the upper end of the scale, or to use a smaller or larger range of values. Transforming the data into z-scores, these potential issues with the raw scores are minimized, and the data can be more meaningfully compared across participants. Visual inspection of the z-score transformed ratings confirmed the emerging consensus (FEATHERSTON, 2005a; SPROUSE, 2007; FEATHERSTON, 2009) that acceptability ratings do not benefit from log-transformation, contrary to some suggestions in the early experimental syntax literature (BARD et al., 1996), and therefore only the z-score transformed data was used in the statistical analyses.

3.7 Results

The results of each subexperiment are summarized in Figure 2, and discussed below.

FIGURE 2: Results of the three subexperiments.
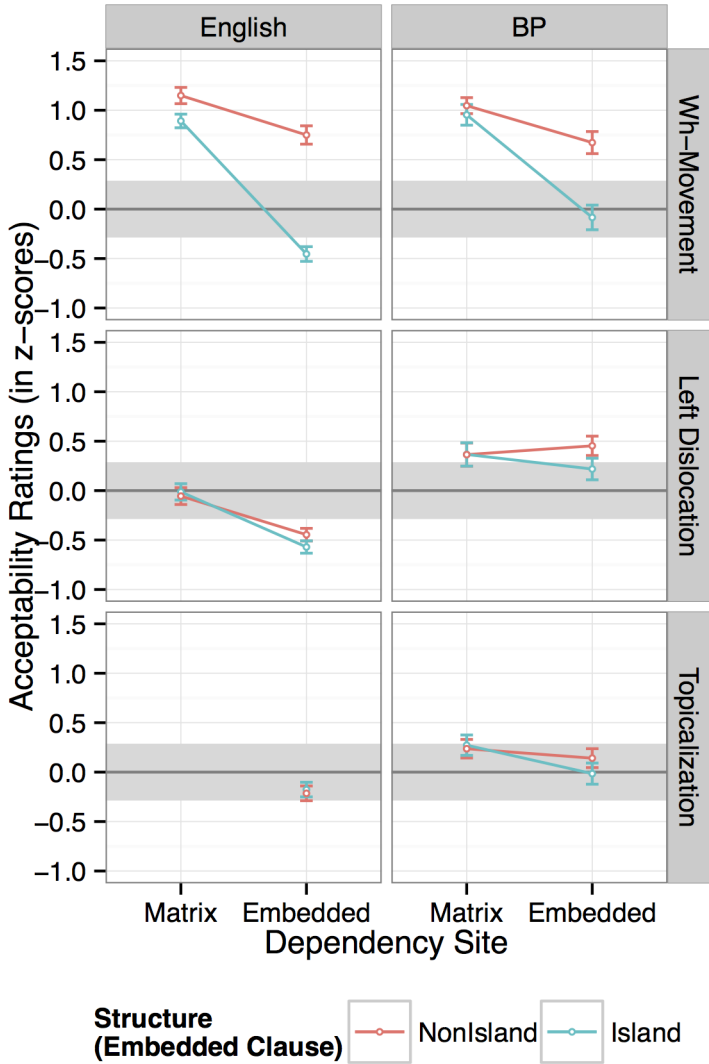
Subexperiment 1A: Wh-islands in English

In English, the interaction between Origin of Movement and Type of Embedded Clause was significant ($F(1, 61) = 42.79$, $p < .0001$), confirming the predicted super-additive effect that characterizes islandhood under its factorial definition.

Both main effects were also significant (Origin of Movement: $F(1, 61) = 218.5$, $p < .0001$; Type of Embedded Clause: $F(1, 61) = 42.79$, $p < .0001$). Movement out of embedded clauses yielded generally lower acceptability ratings than sentences containing movement out of the matrix clauses (mean_Embedded = 0.15, mean_Matrix = 1.02). Sentences containing an embedded island structure received lower acceptability ratings than sentences with regular embedded clauses (mean_Island = 0.22, mean_NonIsland = 0.95).

Planned comparisons (via paired t-tests) within each factor of the 2 x 2 design confirmed that the effect of Origin of Movement was significant both in the Island ($t(61) = 13.3634$, $p < .0001$) and NonIsland structures ($t(61) = 4.6712$, $p < .0001$). The effect of Type of Embedded Clause was significant for sentences with movement from the matrix clause ($t(61) = 2.515$, $p = .0015$), as well as for sentences with movement from the embedded clause ($t(61) = 8.7655$, $p < .0001$).

Subexperiment 1B: Wh-islands in BP

In BP, the interaction between Origin of Movement and Type of Embedded Clause was significant for wh-island cases ($F(1, 37) = 12.32$, $p = .0012$), showing a similar super-additive effect that characterizes islandhood in its factorial definition in English.

Both main effects were also significant (Origin of Movement: $F(1, 37) = 39.22$, $p < .0001$; Type of Embedded Clause: $F(1, 37) = 23.34$, $p < .0001$). Movement out of embedded clauses yielded generally lower acceptability ratings than sentences containing movement out of the matrix clauses (mean_Embedded = 0.29, mean_Matrix = 1). Sentences

containing an embedded island structure received lower acceptability ratings than sentences with regular embedded clauses (mean_Island = 0.44, mean_NonIsland = 0.86).

Planned comparisons (via paired t-tests) within each factor of the 2 x 2 design further demonstrated that the effect of Origin of Movement was significant both in the Island ($t(37) = 5.5978, p < .0001$) and NonIsland structures ($t(37) = 3.9566, p = .0003$). The effect of Type of Embedded Clause was not significant for sentences with movement from the matrix clause ($t(37) = 1.1517, p = .2568$), but was significant for sentences with movement from the embedded clause ($t(37) = 4.6199, p < .0001$).

Subexperiment 2A: Left Dislocation in English

In English, the interaction between Position of Resumptive and Type of Embedded Clause was not significant ($F(1, 61) = 0.001, p = .972$). The super-additive effect that characterizes islandhood in its factorial definition was not observed in this manipulation.

Only the main effect of Position of Resumptive was significant (Position of Resumptive: $F(1, 61) = 5.788, p = .019$; Type of Embedded Clause: $F(1, 61) = 0.396, p = .532$). Resumptives in embedded clauses yielded generally lower acceptability ratings than sentences containing resumptive pronouns in matrix clauses (mean_Embedded = -0.195, mean_Matrix = -0.03).

Planned comparisons (via paired t-tests) within each factor of the 2 x 2 design showed that the effect of Position of Resumptive was significant in the Island ($t(61) = 2.0524, p = .044$) but not in the NonIsland structures ($t(61) = 1.5812, p = .119$). The effect of Type of Embedded Clause was not significant neither for sentences with resumptives in the matrix clause ($t(61) = -0.4517, p = .653$), nor for sentences with resumptives in the embedded clause ($t(61) = -0.4631, p = .645$).

Subexperiment 2B: Left Dislocation in BP

In BP, the interaction between Position of Resumptive and Type of Embedded Clause was not significant ($F(1, 37) = 1.517$, $p = .226$). The super-additive effect that characterizes islandhood in its factorial definition was not observed in this manipulation.

Only the main effect of Position of Resumptive was significant (Position of Resumptive: $F(1, 37) = 5.284$, $p = .0273$; Type of Embedded Clause: $F(1, 37) = 0.608$, $p = .441$). Resumptives in embedded clauses yielded generally lower acceptability ratings than sentences containing resumptive pronouns in matrix clauses (mean_Embedded = 0.06, mean_Matrix = 0.25).

Planned comparisons (via paired t-tests) within each factor of the 2 x 2 design showed that the effect of Position of Resumptive was significant in the Island ($t(37) = 2.2124$, $p = .003$) but not in the NonIsland structures ($t(37) = 0.9992$, $p = .3242$). The effect of Type of Embedded Clause was not significant for sentences with resumptives in the matrix clause ($t(37) = -0.359$, $p = .7216$), nor for sentences with resumptives in the embedded clause ($t(37) = 1.3083$, $p = .1988$).

Subexperiment 3A: Topicalization in English

In English, the topicalization paradigm had to be restricted to include only topicalization from object positions, due to the fact that English does not allow null subject sentences. The only manipulation used in this experiment was the Type of Embedded Clause. The results show that higher ratings were given in average to sentences containing topics extracted from regular complement clauses than to sentences containing topics extracted from interrogative complement clause (wh-islands), but this numeric difference is not statistically significant ($t(62) = 1.4364$, $p = 0.156$)

Subexperiment 3B: Topicalization in BP

In BP, the interaction between Position of Resumptive and Type of Embedded Clause was marginally significant ($F(1, 37) = 2.569$, $p = .1$). This finding is compatible with the predicted super-additive effect that characterizes islandhood under its factorial definition, although the evidence is weak. No other main effect was significant.

Planned comparisons (via paired t-tests) within each factor of the 2 x 2 design showed that the effect of Position of Extraction was not significant neither in the Island ($t(37) = .8044$, $p = .426$) nor in the NonIsland structures ($t(37) = -0.5097$, $p = .6133$). The effect of Type of Embedded Clause was not significant for sentences with topics extracted out of the matrix clauses ($t(37) = -0.0166$, $p = .9868$), but was marginal for sentences with topics extracted from the embedded clause ($t(37) = 1.8772$, $p = .0684$).

3.8 Discussion

Subexperiment 1 demonstrates that both English and BP show wh-island sensitivity. In fact, the average rating of the BP island-violating structure is barely below 0, an indication perhaps that this structure would not have been judged categorically unacceptable if the traditional binary scale had been used. This corresponds quite closely to the hypothetical scenario illustrated in the bottom middle plot of Figure 1.

Subexperiments 2 and 3 investigated two different kinds of topic constructions: Left Dislocation, a construction that is generally described as being insensitive to islands, and Topicalization, which is generally assumed to be an island-sensitive construction. The results from the English part of these subexperiments (2A and 3A) shows that the Left Dislocation paradigm is judged as relatively unacceptable, and shows that the site of the resumption seems to affect the acceptability of the construction: When resumptive pronouns occur in an embedded

clause, the sentence is judged to be less acceptable than the cases where the resumptive pronouns appear in the matrix clause. Despite this linear distance effect, however, the English left dislocation data do not show any evidence of island-sensitivity.

When it comes to Topicalization, we unfortunately could not test the full factorial paradigm in English, due to the restriction against null subjects in the language. The only conditions that could be tested were the extractions out of the embedded clauses. In the case of English, the island-status of the embedded clause did not seem to modulate the acceptability of the topicalized sentences. The Topicalization data shows however that English-speaking participants seem to exhibit a small preference for object topicalization from embedded clauses compared to object left dislocation from embedded clauses.

Unlike the case of the wh-island in regular wh-movement, where both languages show similar acceptability patterns, the data from Left Dislocation and Topicalization showed a divergence between the two languages. Both Left Dislocation and Topicalization in BP demonstrated a small but apparently reliable sensitivity to wh-islands, something that is not found in the English data. Moreover, the acceptability pattern is the same across the two constructions, and shows that, at least qualitatively, the two constructions display the super-additive effect that potentially identifies islandhood under its factorial definition. In addition to the numeric results trending in the direction of island-sensitivity for both topic constructions, Topicalization shows marginal statistical evidence of super-additivity between the two pieces of information (structure of embedded clause and site of the dependency) that define islandhood factorially.

This result is not exactly replicated for Left Dislocation, where the predicted statistical interaction between the two factors is not observed. However, when both Topicalization and Left Dislocation are entered into a three-way 2 x 2 x 2 repeated measures ANOVA, coding Dependency Site

(matrix, embedded), Type of Embedded Clause (island, non-island) and Type of Topic construction (topicalization, left-dislocation) as factors, the interaction between the two relevant factors (Dependency Site and Type of Embedded Clause) result in marginal statistical significance ($F(1,37) = 3.276, p = .07$). This suggests that there is a real, reliable, but perhaps small effect of island-sensitivity in both Topicalization and Left Dislocation in BP. With only 38 participants, each only rating one token from each experimental condition, it is conceivable that our sample may not be large enough to reliably detect a small island sensitivity effect across two different structures (Topicalization and Left Dislocation).

The admittedly still tentative evidence of island-sensitivity for both types of topic constructions in BP has another interesting feature to it: All the items in the factorial paradigm are rated as relatively acceptable sentences, and yet there is some evidence that the acceptability ratings are sensitive to syntactic islands. If this is true, this is a good illustration of the hypothetical scenario depicted in the bottom left plot in Figure 1, and thus would be an example of a *subliminal island effect*.

4   Conclusion

The goal of this study was to try to explore the consequences of refusing to grant epistemological priority to binary categorical judgments of sentence acceptability in linguistic theory. As a test case, we decided to focus on syntactic islands, since they have been a very important class of phenomena within linguistic theory. Because syntactic islands seem to provide a window into the inner workings of long-distance dependency formation in the linguistic computational system, they are both the focus of intense theoretical investigation. They are also used as an important diagnostic tool that theoreticians exploit to adjudicate between competing linguistic analyses. For instance, a common inference that is drawn is that if a long-distance dependency is somehow

insensitive to syntactic islands, then it was probably not generated by the syntactic operation Movement (or whatever ultimately generates the natural class of long-distance dependencies that motivates the existence of this syntactic operation in the first place).

In this study, both of these aspects of syntactic islands were investigated. The first hypothesis that we set out to explore was whether at least some of the within- and across- language variation surrounding syntactic islands was artifactual, more of a consequence of the tools routinely used by syntacticians than a property of the acceptability judgment data they investigate. The test case used in this work was the cross-linguistic variation of wh-islands. It has long been observed that some languages, like English, seem to impose a stronger acceptability penalty on extractions out of specific structural configurations, like an embedded clause headed by a wh-word, when compared to other languages, like Italian or BP. This basic cross-linguistic variation is then generally hypothesized to follow from differences in the grammar of these two sets of languages.

However, as discussed in the introduction, no theory of acceptability judgments and their relationship with grammaticality exists that would license this kind of inference with any degree of certainty. While it is possible that the difference between English and BP when it comes to wh-islands does follow from differences in the two grammars, it is also conceivable that these differences are rather superficial, and that the grammars of the two languages are identical with respect to the type of restrictions they place on long-distance dependencies of the same kind.

The experiment presented in this paper tested this hypothesis and showed that, contrary to traditional descriptions, both English and BP show evidence of wh-island sensitivity. More interestingly, the results also suggest that the reason for the traditional description of the lack of wh-island sensitivity in languages like BP may lie in the simple fact that the structures in BP that violate wh-islands sound more natural

than their English counterparts. This is a pattern of results that we have been referring to throughout this work as *subliminal islands*: cases where measurable island sensitivity effects are observed, and yet do not lead to gross sentence unacceptability.

To provide a stronger case for the existence of subliminal islands, we turned to topic constructions in English and BP. One of them, Topicalization, has generally been described as being island-sensitive, contrary to the other, Left Dislocation. The inference that is generally drawn from these acceptability judgment facts is that Topicalization is generated by movement, while Left Dislocation is not. In keeping with the idea that syntactic island sensitivity may or may not induce gross sentence unacceptability, we have presented data suggesting that, at least in BP, both Topicalization and Left Dislocation show some degree of wh-island sensitivity.

These results are noteworthy for two reasons: First, BP is generally described as not subject to wh-islands. Second, Left Dislocation is generally described as not subject to syntactic islands in general. If the results of our experiment are correct (and the evidence is still tentative at this point), then a profoundly different view of a phenomenon like Left Dislocation emerges. For instance, KATO (1998) has an interesting theory according to which left-dislocated NPs in BP are in fact generated by movement of secondary predicates. However, given the traditional description of the Left Dislocation facts, KATO (1998) has to posit that this particular kind of movement is not subject to island constraints. If Left Dislocation is indeed subliminally sensitive to islands, then Kato's proposal is simplified, and her analysis based on Movement would naturally predict these subliminal island effects. Conversely, all the arguments for a base-generation approach to Left Dislocation that were made primarily on the force of its supposedly island-insensitivity properties would need to be re-evaluated.

However, for all its provocative nature, it is important to be very clear about what the data presented in this paper does not show: First, the data presented in this paper does not challenge in any way the traditional description of the acceptability facts reported in the literature. To native speakers (the author included), wh-island violations in BP do not sound perfect, but neither do they sound utterly degraded. They occupy a position somewhere in the middle range of acceptability. The same pattern is borne out in the data presented here, where violations of wh-islands in BP were judged to be slightly below average acceptability. In addition, when it comes to Left Dislocation, the traditional description is that even in cases where the pronominal element is inside an island, the sentence still sounds acceptable. This is exactly the pattern observed in the data. The major contribution of the results presented here is that, despite the correctness of the data description in the theoretical literature, it is still possible to dissociate *island sensitivity effects* from *categorical unacceptability*.

4.1   Consequences for reductionist theories of islands.

What about reductionist theories of island effects? If the data presented in this paper is on the right track, it nevertheless still presents a challenge to purely reductionist theories of islands. First of all, the very definition of an island effect used in this proposal is that the lowering in acceptability not be reducible to the linear addition of the two constituent parts of islands. In other words, what we consider to be the hallmark of an island effect is that the whole (ie, a long distance dependency inside an island) is more (in this case worse) than the sum of its parts (ie, the effect of having either a long distance dependency or a potentially island-inducing embedded clause), which runs counter reductionist argumentation. Nonetheless, let us assume that the super-additive effect observed in island effects is somehow compatible with the claims of reductionist theories. For instance, let us imagine

a reductionist theory based on working memory limitations, such as KLUENDER and KUTAS (1993). In such a theory, one might propose that when the resources available to correctly parse the sentence are exhausted, a catastrophic failure occurs, leading to the super-additivity effects observed in island violations.

The first problem with a theory like this is that it is virtually indistinguishable from the claims of grammatical theories. That in itself should certainly not count as an argument *against* such a theory, but it certainly does not provides any independently testable predictions on its own. However, if a proposal like the one sketched above is correct, then reductionist theories like the one just mentioned make exactly the *wrong kind* of prediction if *subliminal island effects* are a real phenomenon. This is because now we have cases where we observe a super-additive effect, seemingly created by an island-violation, that nonetheless does not result in a failure to parse the sentence, but rather in a final string that actually sounds marginally acceptable, like the case of wh-movement out of wh-islands in BP, or virtually fully acceptable, like Topicalization and Left Dislocation in BP. The cross-linguistic variation of island-phenomena has always been a challenge for reductionist theories, but if island-effects are indeed universal, but subject to cross-linguistic variation in terms of the categorical acceptability of the island-violating structures, it becomes extremely hard to see how invoking universal costs conspiring with language-independent mechanisms could help explain the complex cross-linguistic island sensitivity facts. It seems that, if anything, this proposal would force us into considering the mirror image of a reductionist theory, one in which there are universal processing costs (which may or may not be language specific), but that conspire with language-dependent structures/mechanisms to give rise to the complex cross-linguistic pattern of acceptability judgments surrounding island phenomena.

References

ARNON, I. et al. **Cross-linguistic variation in a processing account:** The case of multiple wh-questions. Proceedings of BLS. Anais, 2006

BARD, E. G.; ROBERTSON, D.; SORACE, **A. Magnitude estimation of linguistic acceptability.** Language, v. 72, n. 1, p. 32–68, 1996.

CHOMSKY, N. **Aspects of the theory of syntax**. MIT press, 1965.
_____. On wh-movement. In: CULICOVER, P.; WASOW, T.; AKMAJIAN, A. (Eds.). Formal syntax. New York: Academic Press, 1977. p. 71–132.

CHOMSKY, N.; MILLER, G. A. **Introduction to the formal analysis of natural languages**. In: LUCE, R. D.; BUSH, R. R.; GALANTER, E. (Eds.). Handbook of mathematical psychology. Wiley, 1963. v. 2p. 269–321.

FEATHERSTON, S. **Magnitude estimation and what it can do for your syntax:** Some wh-constraints in german. Lingua, v. 115, n. 11, p. 1525–1550, 2005a.
_____. **Universals and grammaticality:** Wh-constraints in german and english. Linguistics, v. 43, n. 4, p. 667–711, 2005b.
_____. **Relax, lean back, and be a linguist**. Zeitschrift für Sprachwissenschaft, v. 28, n. 1, p. 127–132, 2009.

GIBSON, E.; FEDORENKO, E. **Weak quantitative standards in linguistics research**. Trends in cognitive sciences, v. 14, n. 6, p. 233–234, 2010.

GREWENDORF, G. **Empirical evidence and theoretical reasoning in generative grammar**. Theoretical Linguistics, v. 33, n. 3, p. 369–380, 2007.

HAIDER, H. **As a matter of facts–comments on Featherston's sticks and carrots**. Theoretical Linguistics, v. 33, n. 3, p. 381–394, 2007.

HOFMEISTER, P.; CASASANTO, L. S.; SAG, I. **A. How do individual cognitive differences relate to acceptability judgments?:** a reply to Sprouse, Wagers, and Phillips. Language, v. 88, n. 2, p. 390–400, 2012a.
_____. **Misapplying working-memory tests:** A reductio ad absurdum. Language, v. 88, n. 2, p. 408–409, 2012b.

HOFMEISTER, P. et al. **The source ambiguity problem:** Distinguishing the effects of grammar and processing on acceptability judgments. Language and cognitive processes, v. 28, n. 1-2, p. 48–87, 2013.

HOFMEISTER, P.; SAG, I. A. **Cognitive constraints and island effects.** Language, v. 86, n. 2, p. 366–415, 2010.

HOJI, H. **Hypothesis testing in generative grammar:** Evaluation of predicted schematic asymmetries. Journal of Japanese Linguistics, v. 26, p. 25–52, 2010.

KATO, M. A. **Tópicos como alçamento de predicados secundários.** Cadernos de estudos linguisticos, n. 34, p. 67–76, 1998.

KELLER, F. **Gradience in grammar:** Experimental and computational aspects of degrees of grammaticality. University of Edinburgh, 2000.

KELLER, F.; SORACE, A. **Gradient auxiliary selection and impersonal passivization in german:** an experimental investigation. Journal of Linguistics, v. 39, n. 01, p. 57–108, 2003.

KLUENDER, R.; KUTAS, M. **Subjacency as a processing phenomenon**. Language and cognitive processes, v. 8, n. 4, p. 573–633, 1993.

MARANTZ, A. **Generative linguistics within the cognitive neuroscience of language.** The Linguistic Review, v. 22, n. 2-4, p. 429–445, 2005.

MERCHANT, J. **The syntax of silence:** Sluicing, islands, and the theory of ellipsis. Oxford University Press, 2005.

MIOTO, C.; KATO, M. A. **As interrogativas q do português europeu e do português brasileiro atuais**. Revista da ABRALIN, v. 4, n. 1, p. 171–196, 2005.

MIOTO, C.; SILVA, M. C. F.; LOPES, R. E. V. **Manual de sintaxe.** Insular, 2000.

MYERS, J. **Syntactic judgment experiments**. Language and Linguistics Compass, v. 3, n. 1, p. 406–423, 2009.

PHILLIPS, C. **Should we impeach armchair linguists**. Japanese/Korean Linguistics, v. 17, p. 49–64, 2009.

PHILLIPS, C.; WAGERS, M. W.; LAU, E. F. **5 grammatical illusions and selective fallibility in real-time language comprehension**. In: RUNNER, J. (Ed.). Syntax and semantics. Emerald Group Publishing Limited, 2011. v. 37p. 147–180.

RIZZI, L. **Violations of the wh-island constraint and the subjacency condition**. In: Issues in italian syntax. Walter de Gruyter, 1982. v. 11.

ROSS, J. **Constraints on variables in syntax.** Masschusetts Institute of Technology, 1967.

SCHUTZE, C. T. **The empirical base of linguistics:** Grammaticality judgments and linguistic methodology. University of Chicago Press, 1996.

SORACE, A.; KELLER, F. **Gradience in linguistic data**. Lingua, v. 115, n. 11, p. 1497–1524, 2005.

SPROUSE, J. **Continuous acceptability, categorical grammaticality, and experimental syntax**. Biolinguistics, v. 1, p. 123–134, 2007.
_____. **A validation of amazon mechanical turk for the collection of acceptability judgments in linguistic theory.** Behavior and Research Methods, v. 43, n. 1, p. 155–167, 2011.

SPROUSE, J.; ALMEIDA, D. **The empirical status of data in syntax:** A reply to gibson and fedorenko. Language and Cognitive Processes, v. 28, n. 3, p. 222–228, 2013.

SPROUSE, J. et al. **Experimental syntax and the cross-linguistic variation of island effects in English and Italian.** Natural Language and Linguistic Theory. In press.

SPROUSE, J.; SCHÜTZE, C.; ALMEIDA, D. **A comparison of informal and formal acceptability judgments using a random sample from linguistic inquiry:** 2001–2010. Lingua, v. 134, p. 219–248, 2013.

SPROUSE, J.; WAGERS, M.; PHILLIPS, C. **A test of the relation between working-memory capacity and syntactic island effects**. Language, v. 88, n. 1, p. 82–123, 2012a.
_____. **Working-memory capacity and island effects:** A reminder of the issues and the facts. Language, v. 88, n. 2, p. 401–407, 2012b.

TORREGO, E. **On inversion in spanish and some of its effects.** Linguistic inquiry, p. 103–129, 1984.

WESKOTT, T.; FANSELOW, G. **On the informativity of different measures of linguistic acceptability.** Language, v. 87, n. 2, p. 249–273, 2011.