Syntactic islands and learning biases:
Combining experimental syntax and computational modeling to investigate the language acquisition problem

Lisa Pearl & Jon Sprouse
Department of Cognitive Sciences
University of California, Irvine

Abstract

The induction problems facing language learners have played a central role in debates about the types of learning biases that exist in the human brain. Many linguists have argued that some of the learning biases necessary to solve these language induction problems must be both innate and language-specific (i.e., the Universal Grammar (UG) hypothesis). Though there have been several recent high-profile investigations of the necessary learning bias types for different linguistic phenomena, the UG hypothesis is still the dominant assumption for a large segment of linguists due to the lack of studies addressing central phenomena in generative linguistics. To address this, we focus on how to learn constraints on long-distance dependencies, also known as syntactic island constraints. We use formal acceptability judgment data to identify the target state of learning for syntactic island constraints, and conduct a corpus analysis of child-directed data to affirm that there does appear to be an induction problem when learning these constraints. We then create a computational learning model that implements a learning strategy capable of successfully learning the pattern of acceptability judgments observed in formal experiments, based on realistic input. Importantly, this model does not explicitly encode syntactic constraints. We discuss learning biases required by this model in detail as they highlight the potential problems posed by syntactic island effects for any theory of syntactic acquisition. We find that, although the proposed learning strategy requires fewer complex and domain-specific components than previous theories of syntactic island learning, it still raises difficult questions about how the specific biases required by syntactic islands arise in the learner. We discuss the consequences of these results for theories of acquisition and theories of syntax.

1. Introduction

Human learning cannot happen without one or more learning biases. As such, debates in the human learning literature tend to focus on (i) the nature of the evidence available to the learner, and (ii) the nature of those learning biases. Language learning has played a particularly central role in these debates, as the phenomena of language tend to be relatively complex, suggesting that either the evidence available to children must be relatively rich, or that the learning biases available to children must themselves be relatively complex (e.g., Chomsky 1965, 1980). The problem posed by language learning is that neither of these conclusions is particularly appealing. The first conclusion appears to be empirically false: The input available to children appears to be impoverished relative to the complexity of syntactic phenomena, and thus compatible with multiple hypotheses about the adult target state. The second conclusion appears to be theoretically unappealing: The complex learning biases necessary to overcome this induction problem appear to be an order (or orders) of magnitude more complex than learning biases in any other domain of cognition. Our goal in this article is to investigate this tension between the empirical evidence available to children and the complexity of the learning biases necessary to learn from that evidence.

To make this study as relevant as possible to the learning debates, we focus on a syntactic phenomenon that is simultaneously central to modern syntactic theories, and central to proposals for complex learning biases: syntactic island effects. Our methodology is straightforward. First, we will present experimental evidence from formal acceptability judgments that provides a quantitative description of the target state for acquisition, which is adult knowledge of syntactic islands. Then, we will present a quantitative assessment of the evidence available to children based on both automated and manual structural annotation of 148,784 utterances of realistic child-directed speech from the CHILDES corpus (MacWhinney 2000). We will subsequently present a computational model of a statistical learning strategy that can accurately learn the behavior of adult speakers with respect to syntactic island effects using the simplest set of learning biases that we could uncover. We will then discuss each of the biases required by the learning strategy to determine both the type of biases required (e.g., innate versus derived, domain-specific versus domain-general), and the nature of those biases. The results suggest a complicated picture: On the one hand, it is possible in principle to learn syntactic island effects with few, if any, innate, domain-specific biases, and crucially without any biases that specifically instantiate syntactic theories (e.g., the Subjacency Condition); on the other hand, the biases that still appear to be necessary (e.g., tracking trigrams of phrase structure nodes that are part of the syntactic dependency) raise difficult questions about why these particular biases (as opposed to other logically possible biases) are the ones that are part of the successful language learning strategy. Nonetheless, computational models developed using realistic child-directed input allow us to make progress on two fronts. First, they provide a formal mechanism for exploring biases that do not specifically instantiate syntactic theories. Second, they highlight the difficult questions that remain for future research even when a successful learning strategy is found, such as how the remaining biases arise in the learner.

1.1 Categorizing learning biases

Debates about language learning are often framed as a comparison of the *Universal Grammar Hypothesis* (UG) versus non-UG hypotheses. The UG hypothesis takes as its starting point the

assumption that the data available to young children during the language learning process are compatible with multiple hypotheses about linguistic knowledge, resulting in an induction problem known variously as the "Poverty of the Stimulus" (e.g., Chomsky 1980, Lightfoot 1989, Crain 1991), the "Logical Problem of Language Acquisition" (e.g., Baker 1981, Hornstein & Lightfoot 1981), and "Plato's Problem" (e.g., Chomsky 1988, Dresher 2003). The UG hypothesis argues that at least some of the learning biases necessary to solve this induction problem take the form of innately specified, language-specific constraints (Chomsky 1965), which often correspond to specific linguistic phenomena (e.g., anaphoric *one*: Baker 1978, Lidz, Waxman, & Freedman 2003; interpretation of disjunctives: Crain & Pietroski 2002; structure dependence: Chomsky 1965). Non-UG hypotheses, in contrast, attempt to solve the induction problem without postulating any innate, domain-specific constraints. Even a cursory review of the language learning literature reveals that the hypothesis space of non-UG learning biases is potentially very large (some examples are below):

(i) a sensitivity to the distributional data in the available input
[Sakas & Fodor 2001, Pullum & Scholz 2002, Scholz & Pullum 2002, Yang 2002, Regier & Gahl 2004, Yang 2004, Legate & Yang 2007, Pearl & Weinberg 2007, Foraker, Regier, Kheterpal, Perfors, & Tenenbaum 2009, McMurray & Hollich 2009, Pearl & Lidz 2009, Mitchener & Becker 2011, Pearl 2011, Pearl & Mis 2011, Perfors, Tenenbaum, & Regier 2011]

(ii) a preference for simpler/smaller/narrower hypotheses
[Regier & Gahl 2004, Foraker et al. 2009, Pearl & Lidz 2009, Mitchener & Becker 2011, Pearl & Mis 2011, Perfors et al. 2011]

(iii) a preference for highly informative data
[Fodor 1998b, Pearl & Weinberg 2007, Pearl 2008]

(iv) a preference for learning in cases of local uncertainty
[Pearl & Lidz 2009]

(v) a preference for data with multiple correlated cues
[Soderstrom, Conwell, Feldman, & Morgan 2009]

The size and diversity of this hypothesis space of learning biases suggests that a finer-grained framework may be more informative than the traditional binary framework (UG versus non-UG). For the present study, we suggest that learning biases may be categorized along (at least) three dimensions:

a) Are they *domain-specific* or *domain-general*?
b) Are they *innate* or *derived* from prior experience?
c) Are they a constraint on the *hypothesis space*, or a constraint on the *learning mechanism*?

Under this system, the UG hypothesis simply holds that there is at least one innate, domain-specific learning bias (either on the hypothesis space or on the learning mechanism).[1] Similarly a non-UG approach would be one that contains no innate, domain-specific biases: Only innate, domain-general biases, derived, domain-general biases, and derived, domain-specific biases are allowed. For example, all of the learning biases listed in (i-v) above are likely either innate and domain-general (i-iv), or derived and domain-general (v), and therefore would not qualify as UG-biases. However, a sensitivity to linguistic representations that are innately specified (and their distributions in the input) would be an innate and domain-specific bias, and therefore qualify as a UG-bias (e.g., Sakas & Fodor 2001, Yang 2002, Yang 2004, Legate & Yang 2007, Pearl & Lidz 2009, Mitchener & Becker 2011, Pearl 2011, Pearl & Mis 2011, Pearl & Mis 2012).

1.2 Previous investigations of learning biases in syntax

There have been several recent high-profile investigations of the types of learning biases required to learn various aspects of the syntax of human languages. For example, Perfors et al. (2011) have shown that an ideal learner using Bayesian inference will choose hierarchical representations over other kinds of possible representations, given child-directed speech data. This then shows that children do not necessarily need to know beforehand that language uses hierarchical representations; instead, this knowledge can be derived from a domain-general sensitivity to the distributional properties of the data. Importantly, children must still know that hierarchical representations are one possible hypothesis – but they do not need to have competing representations ruled out a priori.[2]

As another example, a number of researchers have recently conducted computational investigations of the acquisition of English anaphoric *one* (e.g., "Look, a red bottle! Oh look, another *one*."). Regier & Gahl (2004) demonstrated that a learner using online Bayesian inference can learn the correct syntactic representation and semantic interpretation of *one* from child-directed speech, provided that the child expands the range of informative data beyond the traditional data set of unambiguous data. Their model highlights the utility of a bias to use statistical distribution information in the data and a bias to prefer simpler/smaller/narrower hypotheses when encountering ambiguous data. Pearl & Lidz (2009) discovered this was an effective strategy only as long as the child knew to ignore certain kinds of ambiguous data;

---

[1] Since the distinction between *hypothesis space* and *learning mechanism* does not impact a bias's status as UG or not, we will not discuss it further here. However, it is worth noting this distinction because many UG proposals tend to involve explicit constraints on the hypothesis space (e.g., certain hypotheses are not available to the child *a priori*), while many non-UG proposals tend to involve implicit constraints on the learning mechanism (e.g., use statistical learning). This is not a logical necessity, as one could easily imagine a UG bias about the learning mechanism (e.g., use a language-specific learning strategy) as well as a non-UG bias about the hypothesis space (e.g., certain hypotheses are *a priori* less probable in a particular hypothesis space, as is the case in Bayesian inference over a subset-superset hypothesis space).

[2] Notably, however, this does not address the induction problem traditionally associated with structure dependence, which concerns hypothesizing structure-dependent *rules* that utilize these hierarchical representations (Berwick et al. 2011). Just because structured representations are available does not necessarily mean children know to use them when forming rules.

therefore, they proposed a learning preference for learning in cases of local uncertainty, which would rule out the troublesome ambiguous data. Pearl & Mis (2011, 2012) discovered that expanding the range of informative data even further negated the need for the local uncertainty bias; instead, a modeled learner could reproduce the observed behavior of children as long as it recognized the distributional similarities between *one* and other referential pronouns like *it*. Notably, however, this learner did not achieve the adult knowledge state, even though it reproduced child behavior. Pearl & Mis (2011) suggest that an additional strategy is still needed to reach the adult knowledge state. One possibility is the learning strategy investigated by Foraker at el. (2009), in which an ideal Bayesian learner with detailed linguistic knowledge about the link between semantic interpretation and certain syntactic structures (syntactic complements and syntactic modifiers) was able to use the difference in distribution for *one* with these structures to converge on the correct knowledge for *one*. In the Foraker et al. (2009) model, the learning mechanism is domain-general; however, it is still unclear whether the detailed linguistic knowledge that is assumed can be derived through domain-general means or would instead be innate and domain-specific.

These previous studies have made at least two contributions to the language learning debates. First, they have demonstrated a concrete set of methodologies for investigating the types of learning biases that are required by language learning. Specifically, by combining child-directed speech corpora with explicitly defined computational learning models, it is possible to systematically test the necessity of different types of learning biases. Second, they have demonstrated that at least some basic syntactic phenomena (e.g., hierarchical representations and anaphoric *one*) could in principle be learned without innate, domain-specific biases. Notably, however, there are some lingering questions, such as whether the fundamental assumptions of these models could *also* be learned without innate, domain-specific biases, and whether the end-states of the models are identical to the end-states hypothesized for adult speakers. We take these results as the starting point for our investigation of learning biases for syntactic island effects.

1.3 The acquisition of syntactic island effects

Although these findings have substantially advanced our understanding of the acquisition of some aspects of syntax, there are at least three ways that the computational approach to the investigation of language learning (and the nature of learning biases) can be further advanced. First, the phenomena that have been investigated so far are generally not considered central to the syntactic theories of UG proponents. This likely means that the theoretical consequences of the previous studies have been limited due to the (relatively) peripheral nature of the phenomena in current syntactic research. In order to truly test the UG hypothesis, and in order for the resulting acquisition models to have a real impact on existing syntactic theories (Chomsky 1965), we need to choose a set of syntactic phenomena that are central to (UG-based) syntactic theories. Second, while the methodology for testing learning biases is relatively clear, the data required to actually perform those tests are still relatively scarce. Realistic syntactic learning models require child-directed speech corpora annotated with specific syntactic structural information, such as phrase structure trees. Unfortunately, many of the freely available corpora do not yet have this kind of syntactic annotation (though there are other types of syntactic annotation available for some corpora, such as dependency tree annotations in CHILDES (Sagae et al. 2010)). Finally, discussions of all of the assumptions underlying successful computational models can help highlight both the progress that they represent (i.e., moving away from explicitly encoding

syntactic theories) and the challenges that they reveal (i.e., lingering questions about how those assumptions are met). Our goal in this paper is to address these three issues by (i) constructing a corpus of child-directed speech with the syntactic annotations that we need to test syntactic learning models, (ii) investigating the learning biases required to learn a set of phenomena that is undeniably central to (UG-based) syntactic theories – namely, syntactic island constraints, and (iii) explicitly discussing all of the components of the simplest successful model, as well as the consequences of that model for both acquisition and syntactic theories.

With these goals in place, our investigation and the rest of this article are both organized as follows. Section 2 introduces syntactic island effects, and presents the formal acceptability judgment experiments (from Sprouse et al. 2012a) that were used to quantitatively define the target state of learning. Section 3 introduces the syntactic annotation process and the results of the structural search of the child-directed speech corpora that were used as realistic child-directed input for the learning model. This step is particularly important, as it identifies the data from which syntactic islands must be learned, and also serves to formalize the apparent induction problem that has been claimed by linguists, but not universally assumed by all researchers (e.g., Sampson 1989, Sampson 1999, Pullum & Scholz 2002, MacWhinney 2004, and Tomasello 2004 among others). Section 4 describes the simplest statistical learner that successfully learns the pattern of island effects. Section 5 reports the results of that learning strategy when it is trained on realistic input, and discusses its behavior in detail. Section 6 presents a detailed discussion of the biases built into this learner, focusing on (i) the empirical motivation (if any) of the bias, (ii) the classification of the bias according to the schema in section 1.1, and (iii) the questions raised by the bias for the learning debates. Section 7 continues the discussion by highlighting potential empirical problems for this learner raised by current empirical claims in syntactic theory (or conversely, predictions that the learner makes concerning related phenomena in syntactic theory). Section 8 concludes.

2. A brief introduction to syntactic island effects

One of the most interesting aspects of the syntax of human languages is the fact that dependencies can exist between two non-adjacent items in a sentence. For example, in English, Noun Phrases (NPs) typically appear adjacent (or nearly adjacent) to the verbs that select them as semantic arguments (e.g., "Jack likes Lily."). However, in English *wh*-questions, *wh*-words do not appear near the verb that selects them as semantic arguments. Instead, *wh*-words appear at the front of the sentence (1a), resulting in a long-distance dependency between the *wh*-word and the verb that selects it (we will mark the canonical position of the *wh*-word, which is often called the *gap position*, with an underscore). One of the defining characteristics of these long-distance *wh*-dependencies is that they appear to be unconstrained by length (Chomsky 1965, Ross 1967): The distance between the *wh*-word and the verb that selects it can be increased by any number of words and/or clauses (1b-d). Though there is clearly an upper bound on the number of words and/or clauses that an English speaker can keep track of during sentence processing, this restriction appears to be based on the limited nature of human working memory capacity rather than an explicit grammatical restriction on the length of *wh*-dependencies in English. Because of this, syntacticians often describe *wh*-dependencies as *unbounded* or *long-distance* dependencies.

(1) a. What does Jack think __?
b. What does Jack think that Lily said __?

c. What does Jack think that Lily said that Sarah heard __?
d. What does Jack think that Lily said that Sarah heard that David stole __?

Though it is true that *wh*-dependencies are unconstrained by length, they are not entirely unconstrained. Linguists have observed that if the gap position of a *wh*-dependency appears within certain syntactic structures, the resulting sentence will be unacceptable (Chomsky 1965, Ross 1967, Chomsky 1973, Huang 1982, and many others):

(2) a. *What did you make [the claim that Jack bought __]?
b. *What do you think [the joke about __] offended Jack?
c. *What do you wonder [whether Jack bought __]?
d. *What do you worry [if Jack buys __]?
e. *What did you meet [the scientist who invented __]?
f. *What did [that Jack wrote __] offend the editor?
g. *What did Jack buy [a book and __]?
h. *Which did Jack borrow [__ book]?

Drawing on the metaphor that the relevant syntactic structures are *islands* that prevent the *wh*-word from *moving* to the front of the sentence, Ross (1967) called the unacceptability that arises in these constructions *island effects*, and the syntactic constraints that he proposed to capture them *island constraints*. Though island effects are typically exemplified by *wh*-dependencies, it should be noted that island effects arise with several different types of long-distance dependencies in human languages, such as relative-clause formation (3), topicalization (4), and adjective-*though* constructions (5):

(3) a. I like the car that you think [that Jack bought __].
b. *I like the car that you wonder [whether Jack bought __].

(4) a. I don't know who bought most of these cars, but that car, I think [that Jack bought __].
b. *I know who bought most of these cars, but that car, I wonder [whether Jack bought __]?

(5) a. Smart though I think [that Jack is __], I don't trust him to do simple math.
c. *Smart though I wonder [whether Jack is __], I trust him to do simple math.

In the 45 years since island effects were first investigated (Chomsky 1965, Ross 1967), there have been literally hundreds of articles in dozens of languages devoted to the investigation of island effects, resulting in various proposals regarding the nature of island constraints (e.g., Erteschik-Shir 1973, Nishigauchi 1990, Deane 1991, Kluender & Kutas 1993, Szabolcsi & Zwarts 1993, Tsai 1994, Reinhart 1997, Hagstrom 1998, Chomsky 2001, Goldberg 2007, Truswell 2007, Abrusan 2011, and many others), the cross-linguistic variability of island effects (e.g., Engdahl 1980, Huang 1982, Rizzi 1982, Lasnik & Saito 1984, Torrego 1984, Hagstrom 1998), and even the real-time processing of dependencies that contain island effects (e.g., Stowe 1986, Kluender & Kutas 1993, McKinnon & Osterhout 1996, Traxler & Pickering 1996, Phillips 2006, and many others). Though most of this literature is beyond the scope of the present article,

it does serve to underscore the central role that syntactic island effects have played in the development of (generative) syntactic theory. Furthermore, the predominant analysis of syntactic island effects in generative syntactic theory is well known to rely on innate, domain-specific learning biases. For example, in the Government and Binding framework of the 1980s, syntacticians proposed a syntactic constraint called the *Subjacency Condition*, which basically holds that the dependency between a displaced element (e.g., a *wh*-word) and the gap position cannot cross two or more *bounding nodes* (Chomsky 1973, Huang 1982, Lasnik & Saito 1984, and many others). The definition of *bounding nodes* can vary from language to language in order to account for the various patterns of island effects that have been observed cross-linguistically. For example, the bounding nodes in English are argued to be NP (Noun Phrase) and IP (Inflection Phrase) (Chomsky 1973), while the bounding nodes in Italian and Spanish are argued to be NP and CP (Complementizer Phrase) (Rizzi 1980, Torrego 1984). Crucially, this framework assumes that the Subjacency Condition itself is part of UG, as are the possible options for bounding nodes (NP, IP, or CP). The language learner then simply needs to determine which bounding nodes are relevant for her specific language in order to learn syntactic island constraints. Although recent evolutions of syntactic theory have terminologically abandoned Subjacency and bounding nodes, it has been argued that modern incarnations of syntactic constraints (such as *phase impenetrability*) are essentially formal variants of the original Subjacency analysis (Boeckx & Grohmann 2007).

Between the centrality of syntactic island effects as a topic of research in (generative) syntactic theory, and the reliance on a UG-based mechanism for their acquisition, it seems clear that syntactic island effects are an ideal case study in the role of innate, domain-specific learning biases in language acquisition. However, investigating the learning of syntactic island effects requires a formally explicit definition of the target state beyond the diacritics that are typically used to delineate unacceptable sentences in syntactic articles. To that end, we decided to explicitly construct the target state from data from Sprouse et al. (2012a), who collected formal acceptability judgments for four island types using the magnitude estimation task: Complex NP islands (2a), (simple) Subject islands (2b), Whether islands (2c), and (conditional) Adjunct islands (2d). These four islands were selected by Sprouse et al. (2012a) for several reasons. First, they have been argued to be captured by syntactic constraints (e.g., Subjacency or the Condition on Extraction Domains), as opposed to the island types that have historically been captured with semantic constraints (e.g., factive islands, negative islands). Second, dependencies spanning these islands are still somewhat intelligible, and so can provide a more nuanced assessment of unacceptability, rather than being complete "word salad". This is because these islands are the more acceptable incarnations of their particular types: Complex NP islands are more acceptable than Relative Clause islands, simple Subject islands are more acceptable than sentential Subject islands, Whether islands are more acceptable than Wh-islands with full *wh*-words in embedded spec-CP, and conditional Adjunct islands are more acceptable than causal Adjunct islands. Thus, a successful learner must accomplish a harder task than if these islands were the less acceptable varieties: The learner must realize that dependencies spanning these more acceptable islands are still ungrammatical when compared to grammatical dependencies, even though these island-spanning dependencies are still relatively intelligible.

The Sprouse et al. (2012a) results are particularly useful for two reasons. First, the magnitude estimation task employs a continuous scale (the positive number line) for acceptability judgments, which results in gradient responses that are comparable to the probabilistic outputs of statistical learning models. Second, Sprouse et al. used a (2x2) factorial

definition of each island effect (shown in 6-9), which controls for the two salient syntactic properties of island-violating sentences: (i) they contain a long-distance dependency, and (ii) they contain an island structure. By translating each of these properties into separate factors, each with two levels (dependency GAP POSITION: matrix, embedded; STRUCTURE present in question: non-island, island), Sprouse et al. were able to define island effects as a superadditive interaction of the two factors - in other words, an island effect is the additional unacceptability that arises when the two factors are combined, above and beyond the independent contribution of each factor. Specifically, a syntactic island occurs when there is more unacceptability than what the EMBEDDED dependency and the presence of an ISLAND structure in the question contribute by themselves.

(6) Complex NP islands

a.	Who __ claimed that Lily forgot the necklace?	MATRIX | NON-ISLAND
b.	What did the teacher claim that Lily forgot __?	EMBEDDED | NON-ISLAND
c.	Who __ made the claim that Lily forgot the necklace?	MATRIX | ISLAND
d.	*What did the teacher make the claim that Lily forgot __?	EMBEDDED | ISLAND

(7) Subject islands

a.	Who __ thinks the necklace is expensive?	MATRIX | NON-ISLAND
b.	What does Jack think __ is expensive?	EMBEDDED | NON-ISLAND
c.	Who __ thinks the necklace for Lily is expensive?	MATRIX | ISLAND
d.	*Who does Jack think the necklace for __ is expensive?	EMBEDDED | ISLAND

(8) Whether islands

a.	Who __ thinks that Jack stole the necklace?	MATRIX | NON-ISLAND
b.	What does the teacher think that Jack stole __ ?	EMBEDDED | NON-ISLAND
c.	Who __ wonders whether Jack stole the necklace?	MATRIX | ISLAND
d.	*What does the teacher wonder whether Jack stole __ ?	EMBEDDED | ISLAND

(9) Adjunct islands

a.	Who __ thinks that Lily forgot the necklace?	MATRIX | NON-ISLAND
b.	What does the teacher think that Lily forgot __ ?	EMBEDDED | NON-ISLAND
c.	Who __ worries if Lily forgot the necklace?	MATRIX | ISLAND
d.	*What does the teacher worry if Lily forgot __ ?	EMBEDDED | ISLAND

Because the factorial definition treats island effects as a superadditive interaction of two factors, the presence of a syntactic island is also visually salient: If the acceptability of the four question types (as indicated by their z-scores) is plotted in an interaction plot, the presence of a syntactic island appears as two non-parallel lines (the left panel of Figure 1), and results in a significant statistical interaction; the absence of a syntactic island appears as two parallel lines (the right panel of Figure 1), and results in no significant statistical interaction.

Figure 1. Example graphs showing the presence (left panel) and absence (right panel) of a syntactic island using the factorial definition from Sprouse et al. (2012a).

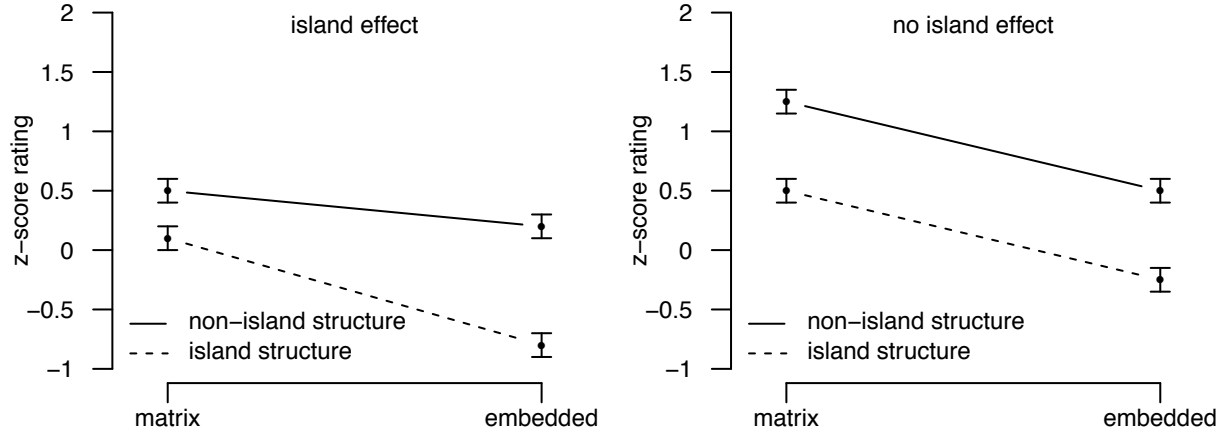
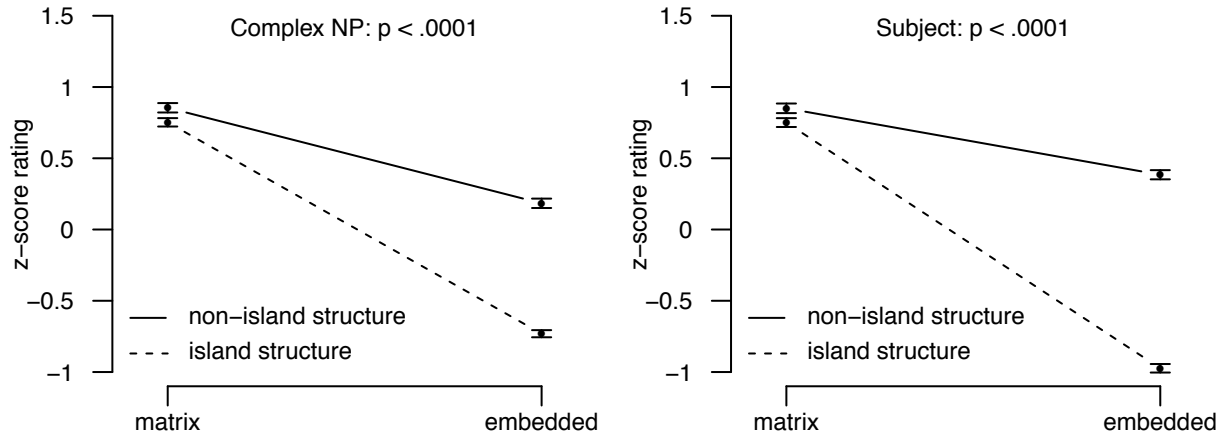
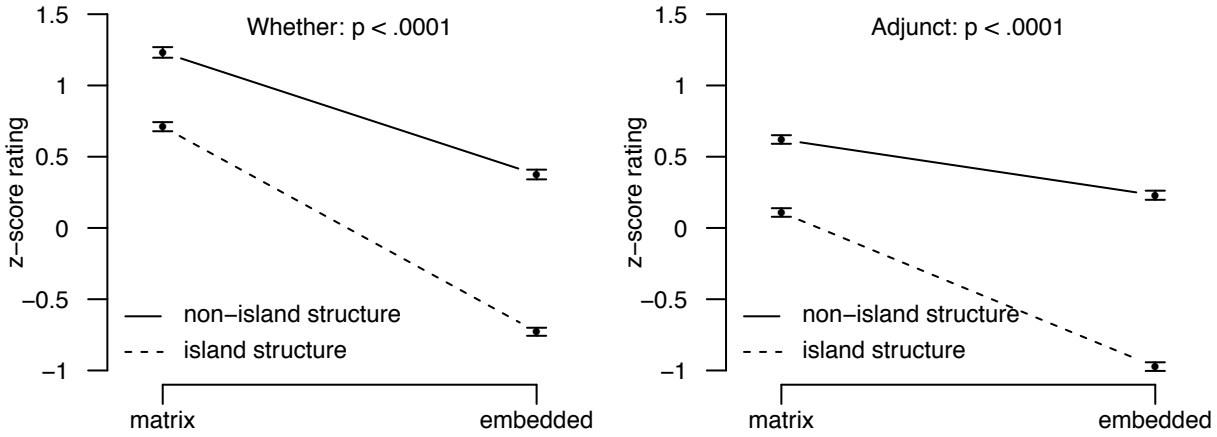


Figure 2 plots the experimentally obtained judgments for the island types investigated in Sprouse et al. (2012a), which shows that adult speakers appear to have implicit knowledge of these four syntactic islands. We can thus use the superadditive interactions for the four island types in Figure 2 as an explicit target state for our statistical learner. [3]

Figure 2. Experimentally derived acceptability judgments for the four island types from Sprouse et al. (2012a) (N=173).



[3] We follow the field of syntax in assuming that well-controlled acceptability judgments can be used to infer grammaticality (see Chomsky 1965, Schütze 1996, Schütze & Sprouse in press, Sprouse & Almeida in press). We also follow the conclusion in Sprouse et al. (2012a, 2012b) that the acceptability judgment pattern observed for syntactic islands is due to grammatical constraints, and likely cannot be explained as an epiphenomenon of sentence processing.



3. Identifying the induction problem using syntactically annotated corpora

To identify an induction problem, we must determine the data available to children, since this is the input they would use to reach the target state knowledge. To assess a child's input for constraints on *wh*-dependencies (and, specifically, the data in the input directly relevant for generating the judgments in Sprouse et al. 2012a), we examined child-directed speech samples to determine the frequency of the structures used as experimental stimuli in Sprouse et al. (2012a). While the CHILDES database has many corpora that are annotated with syntactic dependency information (Sagae et al. 2010), it is difficult to automatically extract the kind of *wh*-dependency information we needed to identify. For this reason, we selected five well-known corpora of child-directed speech from the CHILDES database (MacWhinney 2000) to annotate with phrase structure tree information: the Adam, Eve, and Sarah corpora from the Brown data set (Brown 1973), the Valian dataset (Valian 1991), and the Suppes dataset (Suppes 1974). We first automatically parsed the child-directed speech utterances using a freely available syntactic parser (the Charniak parser[4]), yielding the basic phrase structure trees. However, due to the conversational nature of the data, there were many errors. We subsequently had the parser's output hand-checked by two separate annotators from a group of UC Irvine undergraduates who had syntax training, with the idea that errors that slipped past the first annotator would be caught by the second.[5] We additionally hand-checked the output of our automatic extraction scripts when identifying the frequency of *wh*-dependencies used as experimental stimuli in Sprouse et al. (2012a) in order to provide a third level of error detection.

The data from these five corpora comprise child-directed speech to 25 children between the ages of one and five years old, with 813,036 word tokens total. Of all the utterances, 31,247 contained *wh*-words and verbs, and so were likely to contain syntactic dependencies. Table 1 shows the number of utterances found containing the structures and dependencies examined in Sprouse et al. (2012a).

[4] Available at ftp://ftp.cs.brown.edu/pub/nlparser/.

[5] This work was conducted as part of NSF grant BCS-0843896, and the parsed corpora are available at http://www.socsci.uci.edu/~lpearl/CoLaLab/TestingUG/index.html.

Table 1. The corpus analysis of the child-directed speech samples from CHILDES, given the experimental stimuli used in Sprouse et al. (2012a) for the four island types examined. The syntactic island condition (which is ungrammatical) is italicized.[6]

	MATRIX | NON-ISLAND	EMBEDDED | NON-ISLAND	MATRIX | ISLAND	*EMBEDDED | ISLAND*
Complex NP	7	295	0	0
Subject	7	29	0	0
Whether	7	295	0	0
Adjunct	7	295	15	0

From Table 1, we can see that these utterance types are fairly rare in general, with the most frequent type (EMBEDDED | NON-ISLAND) appearing in only 0.9% of all *wh*-utterances (295 of 31,247). Secondly, we see that being grammatical doesn't necessarily mean an utterance type will occur in the input. Specifically, while both the MATRIX | NON-ISLAND and MATRIX | ISLAND utterance types are grammatical, they rarely occur in the input (7 for MATRIX | NON-ISLAND, either 0 or 15 for MATRIX | ISLAND). This is problematic from a learning standpoint if a learner is keying grammaticality directly to input frequency. Unless the child is very sensitive to small frequency differences (even 15 out of 31,247 is less than 0.05% of the relevant input), the difference between the frequency of grammatical MATRIX | ISLAND or MATRIX | NON-ISLAND utterances and that of ungrammatical EMBEDDED | ISLAND utterances is very small for Adjunct island effects. It's even worse for Complex NP, Subject, and Whether island effects, since the difference between grammatical MATRIX | ISLAND utterances and ungrammatical EMBEDDED | ISLAND structures is nonexistent. Thus, it appears that child-directed speech input presents an induction problem to a learner attempting to acquire an adult grammar for dependencies crossing syntactic islands.

The existence of an induction problem then requires some sort of learning bias in order for children to end up with the correct adult grammar. We note that this induction problem arises when we assume that children are limiting their attention to direct evidence of the language knowledge of interest (something Pearl & Mis (2012) call the *direct evidence assumption*) – in this case, utterances containing *wh*-dependencies and certain linguistic structures. One useful bias may involve children expanding their view of which data are relevant (Foraker et al. 2009, Pearl & Mis 2011, Perfors, Tenenbaum, & Regier 2011), and thus including *indirect positive evidence* (Pearl & Mis 2012) for syntactic islands in their input.[7] We explore this option in the learning strategy we describe in the next section.

---

[6] Note that the number of MATRIX | NON-ISLAND data are identical for all four island types since that control structure was identical for each island type (a *wh*-dependency linked to the subject position in the main clause, with the main clause verb (e.g., *thinks*) taking a tensed subordinate clause (e.g., *Lily forgot the necklace*)). Similarly, the number of EMBEDDED | NON-ISLAND data are identical for Complex NP, Whether, and Adjunct islands since that control structure was identical for those island types (a *wh*-dependency linked to the object position in the embedded clause, with the main clause verb taking a tensed subordinate clause).

[7] Interestingly, the idea of indirect positive evidence is similar in spirit to what linguistic parameters are meant to do in generative linguistic theory - if multiple linguistic phenomena are

4. A statistical learning algorithm for syntactic islands

Though there appears to be an induction problem for syntactic islands, children clearly must utilize some learning procedure in order for them to become adults who have the acceptability judgments observed in Sprouse et al. (2012a). The essence of the acquisition process involves applying learning procedures to the available input in order to produce knowledge about language (Niyogi & Berwick 1996, Yang 2002, among many others). Pearl & Lidz (2009) suggest that the process can be further specified by considering the following components:

(i) children's representations of the hypothesis space
(ii) the set of input children learn from (the data *intake* (Fodor 1998b)), and how that input set is identified and represented
(iii) the updating procedure, and how it uses the intake

In this section, we will use these three components to organize the presentation of our learning algorithm, albeit in a slightly different order: the representation of the input, the representation of the hypothesis space given the input, and the updating procedure given the input. We describe the performance of this learning strategy based on realistic input in section 5. We postpone discussion of the nature of the components of the learning strategy until section 6.

4.1 The representation of the input

Syntactic island effects are constraints on long-distance dependencies; therefore it is clear that the algorithm must operate over sentences that have been parsed into a phrase structure representation, and must also have the ability to track the structural information of the dependency itself (see Fodor 1998a, Fodor 1998b, Sakas & Fodor 2001, and Fodor 2009 for discussions of the utility of parsing during acquisition). Specifically, we propose that the algorithm extracts all of the phrasal nodes that dominate (or "contain") the gap location but not the *wh*-element, resulting in what we call the *container node sequence*. For example, given the sentence (and associated phrase structure representation) in (10a), the container nodes would be the unclosed left brackets that dominate the gap but not the *wh*-element as in (10b), resulting in the container node sequence in (10c). Another example is shown in (11a-c). Here, the gap position associated with the *wh*-element *who* is dominated by several nodes (11b), which can be represented by the container node sequence in (11c).

(10) a. $[_{CP}$ Who did $[_{IP}$ she $[_{VP}$ like __]]]?
b. IP VP
c. IP-VP

(11) a. $[_{CP}$ Who did $[_{IP}$ she $[_{VP}$ think $[_{CP}$ $[_{IP}$ $[_{NP}$ the gift] $[_{VP}$ was $[_{PP}$ from __]]]]]]]]?
b. IP VP CP IP VP PP
c. IP-VP-CP-IP-VP-PP

---

controlled by the same parameter, data for any of these phenomena can be treated as an equivalence class, where learning about some linguistic phenomena yields information about others (Chomsky 1981, Viau & Lidz 2011, Pearl & Lidz in press).

Although container nodes appear to be a relatively complex piece of information to extract from the input, they are not unmotivated, as they play an integral role in all syntactic formulations of island constraints (Ross 1967, Chomsky 1973, among others). Furthermore, the sentence-processing literature has repeatedly established that the search for the gap location is an active process (Crain & Fodor 1985, Stowe 1986, Frazier & Flores d'Arcais 1989) that tracks the container nodes of the gap location (see Phillips 2006 for a review of real-time studies that have demonstrated the parser's sensitivity to island boundaries). In this way, the assumption that the learner could in principle have access to this information from the phrase structure is a well-established fact of the behavior of the human sentence parser (though there is a difference between having access to information and actually using that information, which we will discuss in detail in section 6).

In order to track container node sequences, the learning algorithm must also specify the set of possible container nodes. For the current algorithm, we assume phrase structure nodes that are relatively universal across syntactic theories (e.g., NP, VP, IP, CP). However, the definition of island effects in section 2 and the corpus study in section 3 make it clear that CP nodes must be subcategorized in order to successfully learn syntactic islands. For example, without subcategorizing the CP node, the container node sequence for the grammatical EMBEDDED | NON-ISLAND sentence in the Whether island design would be identical to the ungrammatical EMBEDDED | ISLAND condition: IP-VP-CP-IP-VP. In order to separate these two conditions, the algorithm must track the lexical item that introduces the CP (*that* versus *whether*): IP-VP-CP_{that}-IP-VP versus IP-VP-$CP_{whether}$-IP-VP. This is an empirical necessity; however, we discuss potential empirical motivation for this assumption, as well as the questions it raises, in section 6.

4.2 The representation of the hypothesis space

Given an input representation based on container node sequences, the hypothesis space consists of container node sequences, only some of which are grammatical. This can be formalized through a learning algorithm that assigns some probability to each possible container node sequence, either explicitly or implicitly. However, we already know from the corpus search in section 3 that a learning algorithm that assigns a probability to the full container node sequence based solely on the frequency of that sequence will be unsuccessful, because there are container node sequences that are rated acceptable by adults that nonetheless have a frequency of 0 (or near 0) in child-directed speech. This suggests that the learning algorithm must decompose the container node sequences in some way, prior to assigning probabilities based on the child-directed input.

To solve this problem, the proposed algorithm tracks the frequency of *trigrams* of container nodes (i.e., a continually updated sequence of three container nodes) in the input utterances.[8] For example, the container node sequences from (10c) would be represented as a

[8] Note that this means the learner is learning from data containing dependencies besides the one of interest, treating the other dependencies as indirect positive evidence (Pearl & Mis 2012). For example, a learner deciding about the sequence IP-VP-CP_{that}-IP-VP would learn from IP-VP dependencies that the trigram *start-IP-VP* appears. This is a learning bias that expands the relevant intake set of the learner – all dependencies are informative, not just the ones being judged as grammatical or ungrammatical.

sequence of trigrams as in (12c), and the container node sequences from (11c) would be represented as a sequence of trigrams as in (13c):

(12) a. [$_{CP}$ Who did [$_{IP}$ she [$_{VP}$ like __]]]?
b. IP VP
c. start-IP-VP-end =
start-IP-VP
IP-VP-end

(13) a. [$_{CP}$ Who did [$_{IP}$ she [$_{VP}$ think [$_{CP}$ [$_{IP}$ [$_{NP}$ the gift] [$_{VP}$ was [$_{PP}$ from __]]]]]]]]?
b. IP VP CP_{null} IP VP PP
c. start-IP-VP-CP_{null}-IP-VP-PP-end =
start-IP-VP
IP-VP-CP_{null}
VP-CP_{null}-IP
CP_{null}-IP-VP
IP-VP-PP
VP-PP-end

The ability to track trigrams of container nodes is also an empirical necessity: Neither tracking only unigrams or only bigrams will succeed, as there are grammatical dependencies that contain each of the unigrams and bigrams that exist in the container node sequences in the ungrammatical island violations (see section 6.5). Conversely, 4-grams (and above) will require special treatment for *wh*-dependencies where the gap is in the subject of the matrix clause, as the container node sequence consists of only 3 units (start-IP-end) (again, see section 6.5).

4.3 The updating procedure

The learner generates the probability of a given container node trigram based on the observed data. Then, to gauge the grammaticality of any given container node chain (such as one that crosses an island), the learner calculates the probability of observing that sequence of container node trigrams, which is simply the product of the trigram probabilities.[9] For example, in (12), the sequence IP-VP would have a probability equal to the product of the trigram *start-IP-VP* and the

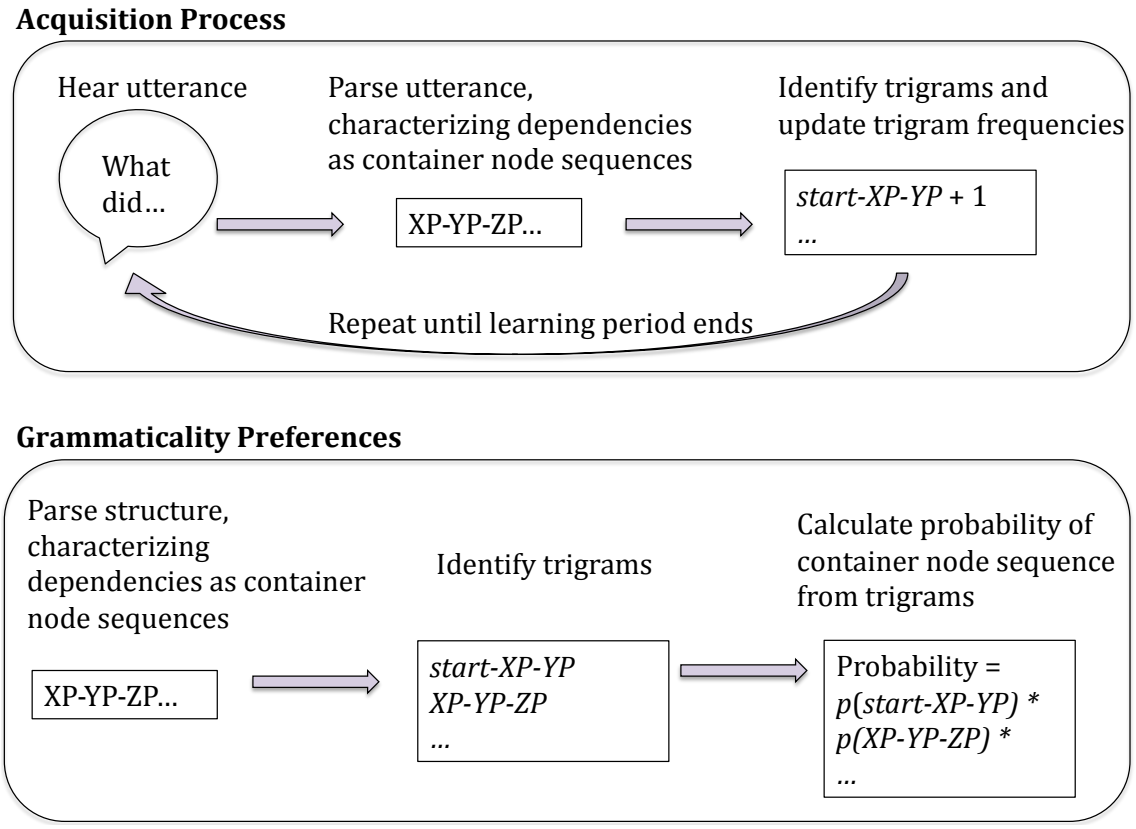
[9] We note that the learner we implement in section 4.4 uses smoothed trigram probabilities (using Lidstone's Law (Manning & Schütze 1999) with smoothing constant $\alpha = 0.5$), so unobserved trigrams have a frequency slightly above 0. Thus, the equation for a trigram *t*'s probability is

$$\frac{\text{total observations of } t \; + \; \alpha}{\text{total observations of all N trigrams} \; + \; N\alpha}$$

Specifically, the learner imagines that unobserved trigrams have been observed α times, rather than 0 times, and all other trigrams have been observed α + their actual observed occurrences. We note also that the overall trend of results we observe later on does not critically depend on the value of α, which effectively serves to distinguish trigrams that rarely occur from trigrams that never occur. The smaller α is, the more these are distinguished.

trigram *IP-VP-end*. The learning algorithm and calculation of grammaticality preferences[10] are schematized in Figure 3, and two examples of grammaticality preferences are shown in (14) and (15). A more formal description of the learning algorithm and generation of grammaticality preferences is provided in Appendix A.

Figure 3. Steps in the acquisition process and calculation of grammaticality preferences.



(14) "Where does the reporter think Jack stole from?"
[$_{CP}$ Where does [$_{IP}$ [$_{NP}$ the reporter] [$_{VP}$ think [$_{CP}$ [$_{IP}$ [$_{NP}$ Jack] [$_{VP}$ stole [$_{PP}$ from __]]]]]]]?"

	IP	VP	CP_{null}	IP	VP	PP
Sequence:	start-IP-VP-CP_{null}-IP-VP-PP-end					
Trigrams:	start-IP-VP					
	IP-VP-CP_{null}					
	VP-CP_{null}-IP					

[10] Here and throughout we will use the term *grammaticality preference* to refer to the result of the learning algorithm (a probability), and *acceptability judgments* to refer to the actual observed behavior of adults in an experimental setting (e.g., Sprouse et al. 2012a). As discussed at the end of section 4, an acceptability judgment is the result of several factors, of which the grammaticality preferences generated by our learner are just one. Other factors affecting acceptability judgments include semantic plausibility, lexical properties, and parsing difficulty.

CP_{null}-IP-VP
IP-VP-PP
VP-PP-end

Probability(IP-VP-CP_{null}-IP-VP-PP) =
p(start-IP-VP)*p(IP-VP-CP_{null})*p(VP-CP_{null}-IP)*p(CP_{null}-IP-VP)*p(IP-VP-PP)*p(VP-PP-end)

(15) *"Who does Jack think the necklace for is expensive?"
[$_{CP}$ Who does [$_{IP}$ [$_{NP}$ Jack] [$_{VP}$ think [$_{CP}$ [$_{IP}$ [$_{NP}$ the necklace [$_{PP}$ for __]] [$_{VP}$ is expensive]]]]]]?
IP VP CP_{null} IP NP PP
Sequence: start-IP-VP-CP_{null}-IP-NP-PP-end
Trigrams: start-IP-VP
IP-VP-CP_{null}
VP-CP_{null}-IP
CP_{null}-IP-NP
IP-NP-PP-
NP-PP-end

Probability(IP-VP-CP_{null}-IP-NP-PP) =
p(start-IP-VP)*p(IP-VP-CP_{null})*p(VP-CP_{null}-IP)*p(CP_{null}-IP-NP)*p(IP-NP-PP)*p(NP-PP-end)

Given this learning algorithm, a child can generate a grammaticality preference for a given dependency at any point during learning, based on the input previously observed, by calculating its probability from the frequency of the trigrams that comprise it (see Figure 3). Similarly, a relative grammaticality preference can be calculated by comparing the probabilities of two dependencies' container node sequences. This will allow us, for example, to compare the inferred grammaticality of dependencies spanning island structures versus dependencies spanning non-island structures.

5. The performance of the algorithm

In this section, we evaluate the performance of the proposed algorithm for both child-directed speech and adult-directed input (both speech and text, which is likely more similar to an adult's linguistic input). We include both types of input in order to assess the performance of the model under slightly different input environments, and to quantify the differences between child- and adult-directed corpora (especially given the scarcity of the former and the relative abundance of the latter). After presenting the results of the algorithm for both input types, we then discuss the detailed behavior of the algorithm to uncover exactly how it is that the set of biases described in section 4 combine to learn the superadditive pattern of island effects.

5.1 Empirically grounding the learner

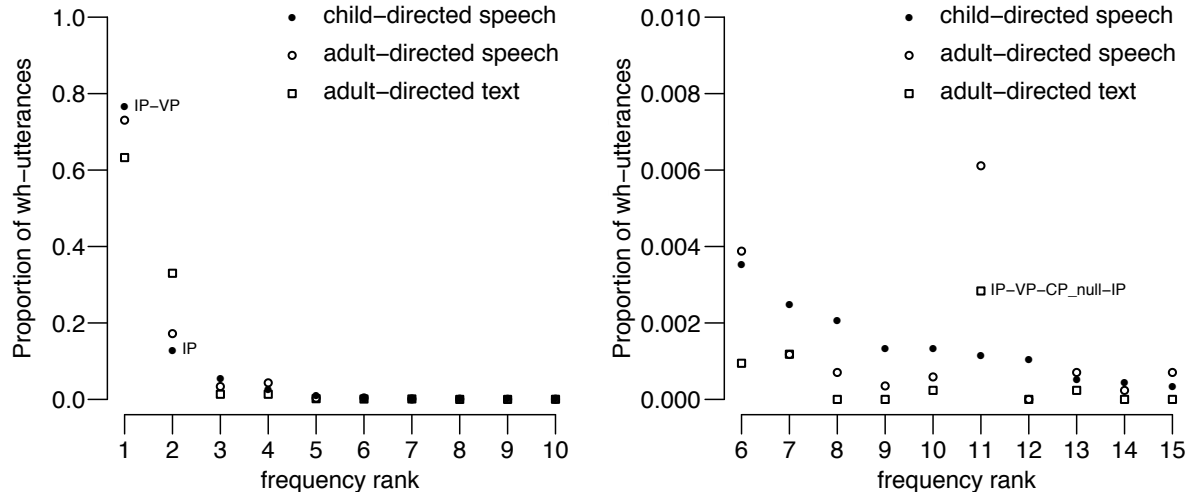
The two datasets used as input were comprised of six corpora across three corpus types: child-directed speech from the Adam and Eve corpora from Brown (1973), the Valian corpus (Valian

1991), and the Suppes corpus (Suppes 1974) of CHILDES (MacWhinney 2000), adult-directed speech from the Switchboard section of the Treebank-3 corpus (Marcus et al. 1999) and adult-directed text from the Brown section of the Treebank-3 corpus (Marcus et al. 1999). Table 3 presents the basic composition of the three corpus types. Figure 4 provides a compact representation of the distribution of the types of *wh*-dependencies in each corpus, while Appendix B provides a detailed description of the composition of each corpus that can be used by readers to construct additional algorithms (or to replicate the performance of the current algorithm).

Table 3: Basic composition of the child-directed and adult-directed input corpora.

	Child-directed: speech	Adult-directed: speech	Adult-directed: text
total utterances	101838	74576	24243
total *wh*-dependencies	20923	8508	4230

Figure 4. The 15 most frequent *wh*-dependency types in the three corpora types. The left panel displays the 10 most frequent *wh*-dependency types for each of the three corpora types, with IP-VP and IP dominating all three corpora types (IP-VP: rank 1, IP: rank 2). The right panel displays the 6th-15th most frequent *wh*-dependency types on a smaller y-axis scale (0-.01) in order to highlight the small amount of variation between corpora types for these dependency types.



These results suggest that two sequences account for a substantial portion of the input of all three corpora: IP-VP, which corresponds to a gap in the matrix object position, and IP, which corresponds to a gap in the matrix subject position. These two dependency types account for between 90 and 95% of the *wh*-dependencies in the input, depending on the corpus type. This analysis also suggests that child-directed speech is similar to adult-directed speech in terms of the proportion of *wh*-dependencies, with IP-VP accounting for a substantially larger proportion of the input than IP (child-directed speech: 76.7% versus 12.8%, adult-directed speech: 73.0% versus 17.2%). This suggests that, at the current level of abstraction, child-directed speech and

adult-directed speech are fairly equivalent, which is not necessarily the case for less abstract representations such as complete phrase structure trees, grammatical category sequences, or vocabulary items. In contrast, adult-directed written text tends to be biased slightly more towards main clause subject dependencies (IP), though main clause object dependencies (IP-VP) are still far more prevalent (IP-VP: 63.3% versus IP: 33.0%). Also, it should be noted that overt complementizers (such as *that*, indicated as CP_{that} in the table in Appendix B) are rare in general. This will be relevant when we examine the learned grammaticality preferences for dependencies involving the complementizer *that*.

In addition to specifying the composition of the input, computational models also require a specification of the amount of input that the algorithm receives in the form of a learning period. We based the current learning period on empirical data from Hart & Risley (1995), who found that children are exposed to approximately 1 million utterances between birth and 3 years of age. Assuming that syntactic islands are acquired within a three year period (perhaps between the ages of 2 and 5 years old; see Goodluck, Foley, & Sedivy 1992, De Villiers & Roeper 1995, De Villiers, Roeper, Bland-Stewart, & Pearson 2008, and Roeper & de Villiers 2011), we can use the composition of the annotated corpora to estimate the number of *wh*-dependencies that would occur in those one million utterances. Given child-directed speech samples from Adam and Eve (Brown 1973), Valian (Valian 1991), and Suppes (Suppes 1974), we estimate the proportion of *wh*-dependencies (20,923) to total utterances (101,823) as approximately 0.2. We thus set the learning period to 200,000 *wh*-dependency data points. This means that the current algorithm will encounter 200,000 data points containing *wh*-dependencies, drawn randomly from a distribution characterized by the corpora in the table in Appendix B.

5.2 Success metrics and learner implementation

We can test the current algorithm by comparing the learned grammaticality preferences to empirical data on adult acceptability judgments from Sprouse et al. (2012a). The container node sequences that arise for the sentence types in (6-9) above are given in (16-19). It should be noted that the current algorithm will compare syntactic island violations to only three types of grammatical container node sequences, despite the number of superficial sentence types involved: IP, IP-VP-CP_{that}-IP-VP, and IP-VP-CP_{null}-IP.[11]

(16) Complex NP islands

a.	IP	MATRIX | NON-ISLAND
b.	IP-VP-CP_{that}-IP-VP	EMBEDDED | NON-ISLAND

[11] This shows that actual process of generating acceptability judgments is likely more nuanced than the basic implementation in the current algorithm. One clear difference is that the current algorithm does not factor in the portion of the utterance beyond the gap position, whereas the actual process in humans likely does. For example, *Who saw it?* is not judged as equivalent to *Who thought that Jack said that Lily saw it*?, even though both are IP dependencies. Similarly, the current algorithm does not factor lexical or semantic properties into the judgments, whereas the actual process in humans likely does. This is why experimental studies have to balance the lexical, structural, and semantic properties of the experimental materials, as Sprouse et al. (2012a) did.

c. IP — MATRIX | ISLAND
d. *IP-VP-NP-CP_{that}-IP-VP — EMBEDDED | ISLAND

(17) Subject islands

a. IP — MATRIX | NON-ISLAND
b. IP-VP-CP_{null}-IP — EMBEDDED | NON-ISLAND
c. IP — MATRIX | ISLAND
d. *IP-VP-CP_{null}-IP-NP-PP — EMBEDDED | ISLAND

(18) Whether islands

a. IP — MATRIX | NON-ISLAND
b. IP-VP-CP_{that}-IP-VP — EMBEDDED | NON-ISLAND
c. IP — MATRIX | ISLAND
d. *IP-VP-$CP_{whether}$-IP-VP — EMBEDDED | ISLAND

(19) Adjunct islands

a. IP — MATRIX | NON-ISLAND
b. IP-VP-CP_{that}-IP-VP — EMBEDDED | NON-ISLAND
c. IP — MATRIX | ISLAND
d. *IP-VP-CP_{if}-IP-VP — EMBEDDED | ISLAND

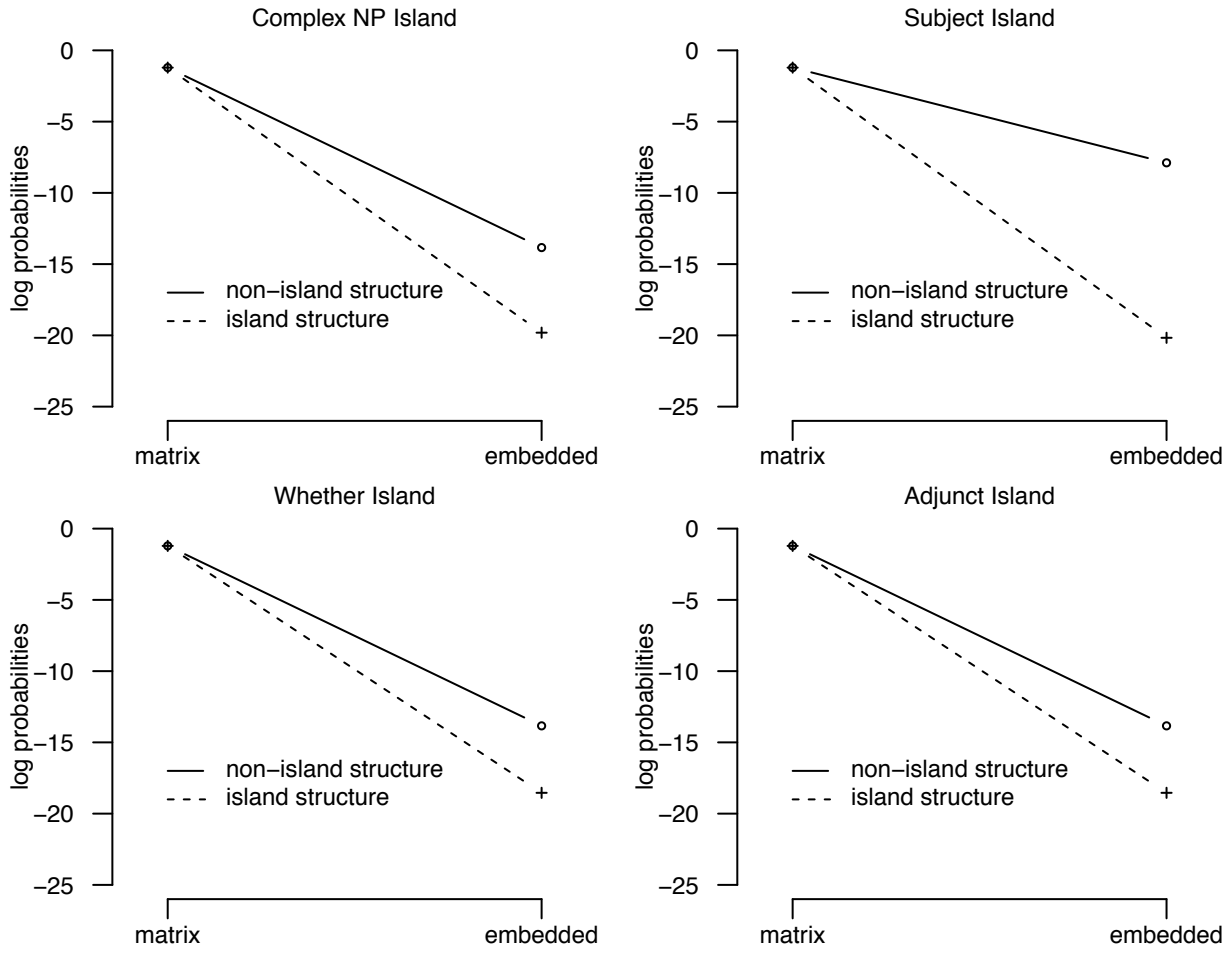
Recall that this factorial definition of island effects makes the presence of island effects visually salient. If the acceptability of the four utterance types is plotted in an interaction plot, the presence of an island effect shows up as two non-parallel lines (e.g., the left panel of Figure 1), while the absence of an island effect shows up as two parallel lines (e.g., the right panel of Figure 1). Sprouse et al. (2012a) found an island effect pattern for all four island types; therefore, a successful algorithm will also reveal an island effect pattern for all four island types.

To evaluate the success of the current algorithm, we can plot the predicted grammaticality preferences in a similar interaction plot: If the lines are non-parallel, then the learner has acquired the knowledge required to implement island constraints; if the lines are parallel, then the learner did not acquire the knowledge required to implement island constraints. The current algorithm will follow the grammaticality preference calculation process outlined in Figure 3 and Appendix A. In particular, it will receive data incrementally, identify the container node sequence and trigrams contained in that sequence, and update the corresponding trigram frequencies. It will then use these trigram frequencies to infer a probability for a given *wh*-dependency, which can be equated to the judged acceptability of that dependency – more probable dependencies are more acceptable, while less probable dependencies are less acceptable. Though the inferred acceptability can be generated at any point during learning (based on the trigram frequencies at that point), we will show results only from the end of the learning period.

5.3 Modeling results

Because the result of a grammaticality preference calculation is often a very small number (due to multiplying many probabilities together), we will instead report the log probability. This allows for easier comparison with acceptability judgments. All log probabilities are negative (this is because raw probabilities are between 0 and 1, and the logarithm of numbers less than 1 is negative). The more positive numbers (i.e. closer to zero) represent "more acceptable" structures while more negative numbers (i.e., farther from zero) represent "less acceptable" structures.[12] Figures 5 and 6 represent the results of the proposed algorithm given child-directed and adult-directed input, respectively. Table 4 lists the log probabilities depicted in Figures 5 and 6.

Figure 5. Log probabilities derived from a learner using child-directed speech.



[12] This measurement is similar to *surprisal*, which is traditionally defined as the negative log probability of occurrence (Tribus 1961) and has been used recently within the sentence processing literature (Hale 2001, Jaeger & Snider 2008, Levy 2008, Levy 2011). Under this view, less acceptable dependencies are more surprising.

Figure 6: Log probabilities derived from a learner using adult-directed speech and text.

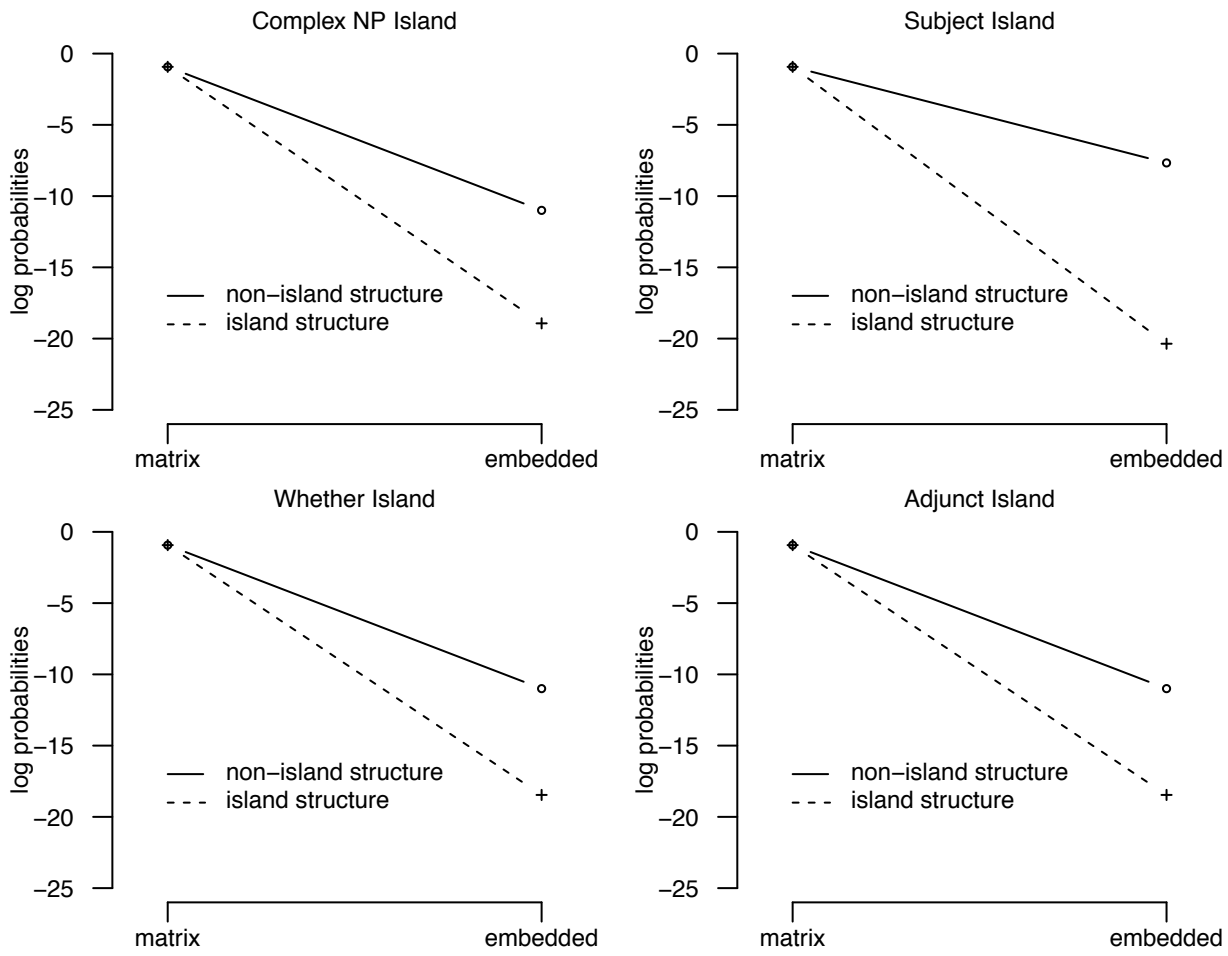


Table 4. Inferred grammaticality of different *wh*-dependencies from Sprouse et al. (2012a), represented with log probability.

		Child-directed speech	Adult-directed speech & text
Grammatical dependencies			
matrix subject	IP	-1.21	-0.93
embedded subject	IP-VP-CP_{null}-IP	-7.89	-7.67
embedded object	IP-VP-CP_{that}-IP-VP	-13.84	-11.00
Island-spanning dependencies			
Complex NP	IP-VP-NP-CP_{that}-IP-VP	-19.81	-18.93
Subject	IP-VP-CP_{null}-IP-NP-PP	-20.17	-20.36
Whether	IP-VP-$CP_{whether}$-IP-VP	-18.54	-18.46
Adjunct	IP-VP-CP_{if}-IP-VP	-18.54	-18.46

Figures 5 and 6 indicate that learners using either child-directed or adult-directed input and the proposed algorithm would arrive at the correct pattern of grammaticality preferences (a superadditive interaction) for all four islands. Furthermore, the log probabilities suggest that the ungrammatical island violations are substantially less acceptable than the grammatical control conditions in the factorial design. This can be seen by subtracting the log probabilities of the two conditions that one wishes to compare: Because subtraction in logarithmic space is equivalent to division in the raw space, the difference between two log probabilities indicates the number of times larger or smaller one probability is than the other. For example, the log probability of Subject island violations (-20.17) is 12.28 less than the log probability of an embedded subject dependency (-7.89). This indicates that the proposed algorithm rates Subject island violations as 12.28 times less acceptable than embedded subject dependencies. This measure is also known as the *log-odds* of the comparison. All of the island violations are at least 4 times less acceptable than the grammatical control conditions, and often more than 10 times less acceptable.

Although these results demonstrate that our modeled learner can acquire the general superadditive interaction pattern observed in the actual acceptability judgment experiments, it should be noted that there are noticeable differences between the observed acceptability judgments and the inferred grammaticality preferences learned by this model. The reason for this is that actual acceptability judgments are based on dozens of factors that are not included in this model. For example, lexical items, semantic probability, and processing difficulty have all been demonstrated to impact acceptability judgments (Schütze 1996, Cowart 1997, Keller 2000, Sprouse 2009). The inferred grammaticality of this particular model would constitute only one (relatively large) factor among many that affect acceptability. In other words, the grammaticality preferences of this model are themselves limited to the dependency alone – they ignore all of the other properties of the sentence.

5.4 Understanding the behavior of the algorithm

The results of the previous section suggest that the proposed algorithm can successfully learn the superadditive pattern of syntactic island effects from realistic child-directed or adult-directed input. The question then is how tracking container node trigrams leads to such success. The answer requires a closer examination of the container node trigram probabilities involved in each island-spanning dependency, as shown in Table 5 below. Crucially, for each of the island-spanning dependencies, there is at least one extremely low probability container node trigram in the container node sequence of the dependency. These trigrams are assigned low probabilities because these trigram sequences are never observed in the input – it is only the smoothing parameter that prevents these probabilities from being 0. Note that some trigrams are low probability due to being rarely encountered in the input (e.g., CP_{that}-IP-VP in child-directed speech) – but, crucially, this is still more than never. Even though CP_{that} rarely appears, it *does* appear, and so it is assigned a probability that is substantially non-zero.

Table 5. Container node trigram probabilities for each of the island-crossing dependencies, as well as two grammatical dependencies, after the learning period has finished. Very low probability container node trigrams, which are never observed in the input, are in **bold**. Log probability for the complete dependency is also shown.

Dependency	Trigram	Child-directed probabilities	Adult-directed probabilities
Complex NP island	start-IP-VP	.42	.41
IP-VP-NP-CP_{that}-IP-VP	IP-VP-NP	.0015	.0011
	VP-NP-CP_{that}	**.0000012**	**.0000013**
	NP-CP_{that}-IP	**.0000012**	**.0000013**
	CP_{that}-IP-VP	.000044	.00004
	IP-VP-end	.4	.38
log(probability)		-19.81	-18.46
Subject island	start-IP-VP	.42	.41
IP-VP-CP_{null}-IP-NP-PP-end	IP-VP-CP_{null}	.0073	.0045
	VP-CP_{null}-IP	.0073	.0045
	CP_{null}-IP-NP	**.0000012**	**.0000013**
	IP-NP-PP	**.0000012**	**.0000013**
	NP-PP-end	.00021	.0003
log(probability)		-20.17	-20.36
Whether island	start-IP-VP	.42	.41
IP-VP-$CP_{whether}$-IP-VP	**IP-VP-$CP_{whether}$**	**.0000012**	**.0000013**
	VP-$CP_{whether}$-IP	**.0000012**	**.0000013**
	$CP_{whether}$-IP-VP	**.0000012**	**.0000013**
	IP-VP-end	.4	.38
log(probability)		-18.54	-18.46
Adjunct island	start-IP-VP	.42	.41
IP-VP-CP_{if}-IP-VP	**IP-VP-CP_{if}**	**.0000012**	**.0000013**
	VP-CP_{if}-IP	**.0000012**	**.0000013**
	CP_{if}-IP-VP	**.0000012**	**.0000013**
	IP-VP-end	.4	.38
log(probability)		-18.54	-18.46
Triple object, no CPs	start-IP-VP	.42	.41
IP-VP-IP-VP-IP-VP	IP-VP-IP x 2	.031	.017
	VP-IP-VP x 2	.031	.017
	IP-VP-end	.4	.38
log(probability)		-6.81	-7.89
Triple object, with CPs	start-IP-VP	.42	.41
IP-VP-CP_{null}-IP-VP-CP_{null}-IP-VP	IP-VP-CP_{null} x 2	.0073	.0045
	VP-CP_{null}-IP x 2	.0073	.0045
	CP_{null}-IP-VP x 2	.0067	.002
	IP-VP-end	.4	.38
log(probability)		-13.67	-15.59

In addition to highlighting the role of low probability trigrams in determining the acceptability of syntactic island violations, Table 5 also highlights the tension between the length of the dependency and its acceptability. Given that the proposed algorithm calculates the probability of the dependency as the product of the probability of the trigrams that compose the sequence, longer dependencies will tend to be less probable than shorter dependencies because longer dependencies by definition involve the multiplication of more probabilities, and those probabilities are always less than 1. Despite this general tendency to prefer shorter dependencies to longer dependencies, the specific frequencies of the individual trigrams comprising those dependencies still have a large effect. As a concrete example, Table 5 lists two grammatical dependencies that are relatively long: Both are triply embedded object dependencies, one with no CP container nodes (IP-VP-IP-VP-IP-VP: e.g., *What does Lily want to pretend to steal?*), and one with CP container nodes (IP-VP-CP_{null}-IP-VP-CP_{null}-IP-VP : e.g., *What does Lily think Jack heard she stole?*). In both cases, these grammatical dependencies are categorized by the algorithm as more probable than the island violations (as shown by their log probabilities) despite being substantially longer than the island violations. This is because the container node trigrams that comprise the grammatical dependencies occur with some frequency in the input.

One concern with this approach is that it might be seen to equate difficulty with ungrammaticality (Phillips 2012b).[13] In particular, one might worry that very long dependencies would start to resemble ungrammatical dependencies with respect to acceptability under the proposed algorithm, even though native speakers report a qualitative perceptual difference between them. This may in fact be a general problem for acceptability judgments as a measure of grammaticality. For example, the formal acceptability judgment experiments of Alexopoulou and Keller (2007) concretely demonstrate that very long dependencies (i.e., dependencies that cross two or more clause boundaries) are often rated identically to island violations in acceptability judgment experiments, suggesting that acceptability alone is not enough to capture the qualitative difference between sentences whose ungrammaticality leads to low acceptability and sentences whose length leads to low acceptability. Phillips is correct that this general problem is maintained in the current algorithm, because the current algorithm is designed to capture acceptability effects. However, it should be noted that the current algorithm can in principle distinguish sentences that are low probability because of length (i.e., very many trigrams) and sentences that are low probability because one or more trigrams never occur. This is because the algorithm itself collects precisely that information (the number of trigrams and the relative probability of each trigram). In other words, the qualitative distinction between the two types of sentences can be recovered by looking at how the low probability was calculated. It is an open question how it is that a learner using this algorithm would access this calculation; however, as one anonymous reviewer observes, one simple formal way to capture this is with the geometric mean, which is calculated by multiplying a sequence of numbers together and taking the n^{th}-root of the product where n is equal to the number of items that were originally multiplied (a formula is given in 20). The geometric mean of an ungrammatical dependency will be substantially lower than the geometric mean of a longer dependency, due to the presence of the trigrams that never occur.

(20) Geometric mean of probabilities $p_1 \ldots p_n = \sqrt[n]{\prod_{i=1 \text{ to } n} p_i}$

[13] We are especially grateful to Colin Phillips for his thoughts and suggestions concerning this.

Phillips (2012b) suggests two additional ways around this issue. One is to make the smoothing factor α much smaller (e.g., .00005 instead of 0.5). This effectively further penalizes trigrams that have never been observed – their probability, though non-zero, is significantly smaller and thus lowers the probability of the dependency they are part of. Another way is to back off from the notion of a combined probability for the entire dependency. Instead, a learner could simply note the presence of a very low probability trigram in any given dependency - this might arise naturally if that part of the dependency is difficult to process, because that container node trigram hasn't been encountered before.

5.5 The success of the algorithm

These results suggest that syntactic island effects can be learned from realistic child-directed input using an algorithm that does not directly encode syntactic island constraints. The proposed algorithm does require relatively sophisticated biases, such as (i) the parsing of sentences into phrase structure trees, (ii) the extraction of sequences of container nodes for the dependencies, (iii) the tracking of the frequency of trigrams of container nodes, and (iv) the calculation of the probability of the complete container node sequence for the dependency, based on its trigrams. The results also suggest that two desirable properties of acceptability judgments fall out of this algorithm: (i) a general preference for shorter dependencies, and (ii) a qualitative distinction between long dependencies and ungrammatical dependencies (at least in principle). The construction of this algorithm represents substantial progress in understanding the space of possible learning theories for complex syntactic phenomena like syntactic island effects, but it also raises difficult questions about how the component biases of such an algorithm actually arise in the learner. We turn to these questions in the next section.

6. A discussion of the biases of the algorithm

The previous section presented the successes of the proposed algorithm. In this section, we discuss each bias that comprises the algorithm individually, with a particular focus on (i) the empirical motivation for each bias, (ii) the potential classification of each bias according to the framework laid out in section 1, and (iii) the unanswered questions (for future research) raised by the empirical necessity of each bias.

6.1 Syntactic category and phrase structure information

One of the most basic components of the proposed learning algorithm is that it operates over input that has been parsed into phrase structure trees. It therefore assumes that both syntactic category information and phrase structure information have already been acquired (or are in the process of being acquired). We do not have too much to say about this assumption because basic syntactic phenomena like syntactic categories and phrase structure parsing are required by nearly every syntactic phenomenon. It may be the case that the acquisition of syntactic categories or phrase structure requires at least one innate, domain-specific bias, in which case every syntactic phenomenon, including syntactic islands, would (strictly speaking) require such a bias. Nonetheless, this would not be a fact that is specific to syntactic islands, but rather a general fact of every syntactic phenomenon. We are specifically interested in the consequences of syntactic

islands for learning theories, not the consequences of every syntactic phenomenon. That being said, for recent work investigating the acquisition of syntactic categories from child-directed input, see Mintz (2003) and (2006), and for recent work investigating the acquisition of hierarchical structure given syntactic categories as input, see Klein & Manning (2002).

6.2 Tracking frequencies and calculating probabilities

Another basic component of the proposed algorithm is that the learner has the ability to track the frequency of units in the input, and then calculate the probabilities of those units. This is a relatively uncontroversial assumption, as many learning theories, both in language and other cognitive domains, assume that the learner can track frequencies and calculate probabilities. The ability to track frequencies and calculate probabilities is likely an innate, domain-general ability. Still, the interesting question about the ability to track frequencies and calculate probabilities is not so much the existence of the ability itself, but rather the units that are tracked – a question that we turn to in the next four subsections.

6.3 Restricting the input to *wh*-dependencies

The proposed algorithm assumes that only *wh*-dependencies are used as input by the learner, at least for the acquisition of syntactic island effects with *wh*-dependencies. This assumption is not as neutral as it first appears. First, many syntactic theories recognize similarities between *wh*-dependencies and other types of dependencies, such as relative-clause-dependencies (*rc*-dependencies), by postulating syntactic mechanisms that are common to both (e.g., A′-movement). This might then suggest that, for the purposes of acquisition, the two types of dependencies should be treated as equal. However, recent formal acceptability judgment experiments suggest that, at least in English, *wh*-dependencies and *rc*-dependencies display different sets of island effects, although both do demonstrate some island effects (Sprouse, Caponigro, Greco, and Ceccheto *submitted*). This suggests that these two dependencies must be tracked separately for the purposes of acquisition, and so both should be input (separately) into the syntactic islands learning algorithm.

Second, other dependencies, such as the binding dependencies that hold between nouns and pronouns, do not display syntactic island effects at all. The fact that binding dependencies lack island effects is ambiguous: It could either mean that binding dependencies are never subjected to the syntactic islands learning algorithm, or it could mean that the input for binding dependencies contains the trigrams that are never observed in the input for *wh*-dependencies, and so island effects are not observed for binding dependencies. Although teasing apart these two possibilities is beyond the scope of this article, we can at least say that some of the low probability trigrams that lead to island effects, such as VP-$CP_{whether}$-IP are certainly possible in binding dependencies, as in *John wonders whether you like him*. This suggests that even if the syntactic islands learning algorithm were applied to binding dependencies, the algorithm would likely not lead to syntactic island effects for binding dependencies. In a similar vein, both of the logical possibilities suggest that binding dependencies must be tracked separately from *wh*-dependencies.

The fact that it is empirically necessary to separate *wh*-dependencies from other dependency types does not explain how it is that the acquisition system knows to separate the input. While it is logically possible to achieve this type of separation without necessarily

invoking innate, domain-specific biases, we simply do not have enough information about the learnability of these other dependency types to evaluate the possibilities. What we can say is that the learning strategy proposed here highlights the fact that the any theory of the acquisition of syntactic islands must be able to track *wh*-dependencies separately from *rc*-dependencies and binding dependencies.

6.4 Tracking sequences of container nodes

Much like the assumption that the input must be restricted to *wh*-dependencies, the bias in the proposed algorithm to track sequences of container nodes appears relatively neutral at first glance; after all, syntactic island effects are constraints on dependencies, and therefore the algorithm should track information about the dependencies. However, this assumption is far from neutral, as it is in essence informing the system that long-distance dependencies may have constraints on them and so information about them should be tracked. Of course this is an empirical necessity: There are such things as syntactic island effects, and they do appear to vary both across languages (Rizzi 1982) and across constructions (Sprouse et al. *submitted*). Therefore, the acquisition system must learn (something about) them. But nothing about this algorithm explains why the system *attempts* to learn constraints on long-distance dependencies. For attempts to explain the existence of syntactic islands based on considerations of computational (parsing) efficiency, see Fodor (1978, 1983), Berwick & Weinberg (1984), and Hawkins (1999).

Beyond encoding the very existence of constraints on long-distance dependencies, the bias to track sequences of container nodes also raises the question of how it is that the algorithm knows to track container nodes rather than some other piece of information about a dependency. In other words, why couldn't the constraints be stated over the number of nouns in the dependency, or the number of prepositions, or even stated over certain semantic categories such as temporal modifiers? It is true, as mentioned in section 4, that the fact that the parsing of long-distance dependencies is an active process means that the sequence of container nodes is information that is likely available to the language system, but *availability* is distinct from *attention*. The current algorithm is biased to attend to container nodes instead of all of the other logically possible types of information about dependencies that are potentially available. This bias is likely domain-specific, as long-distance dependencies (and their constraints) have not been clearly demonstrated in any other domain of cognition. It is, however, an open question whether this bias is also innate, or whether it can be derived from other biases. Nonetheless, it seems to be the case that any theory of syntactic islands that postulates a structurally-defined constraint will likely track container nodes, and therefore will be confronted with this difficult question.

6.5 Tracking trigrams instead of other n-grams

The proposed algorithm decomposes the sequence into trigrams (a moving window of three container nodes). Once again, this is an empirical necessity. The corpus analysis in section 3 suggests that the learning algorithm must decompose the container node sequences into smaller units, otherwise three of the (grammatical) MATRIX | ISLAND conditions would be erroneously characterized as ungrammatical. A unigram model will successfully learn Whether and Adjunct islands, as there are container nodes in these dependencies that never appear in grammatical

dependencies ($CP_{whether}$ and CP_{if}), but will fail to learn Complex NP and Subject islands, as all of the container nodes in these islands are shared with grammatical dependencies. This is problematic under the assumption that all four island types should be learned by the same algorithm, although it is logically possible that different island types arise due to different algorithms (perhaps according to some of the theoretical distinctions that have been postulated in the syntax literature). In effect, there is tension between the size of the *n*-grams that the algorithm tracks, and the number of learning algorithms that are necessary: Decreasing one requires an increase in the other to capture the empirical facts of syntactic islands.

A similar problem arises for a bigram model: At least for Subject islands, there is no bigram that occurs in a Subject island violation but not in any grammatical dependencies. The most likely candidate for such a bigram is IP-NP, as this is precisely the configuration that suggests extraction from the subject position (and thus distinguishes Subject islands from grammatical extraction from objects, which would be VP-NP). However, sentences such as *What, again, about Jack impresses you?* or *What did you say about the movie scared you?* suggest that a gap can arise inside of NPs, as long as the extraction is of the head noun (*what*), not of the noun complement of the preposition. One way to circumvent this problem is to assign a different structural analysis to these sentences such that the container node sequence no longer contains IP-NP. In this case, there is a tension between the simplicity of the structural analysis of certain sentences, and the size of the *n*-grams that the algorithm tracks. Another option is to postulate a distinct learning algorithm for Subject islands, as previously discussed for unigrams.

Although trigrams are the smallest *n*-gram that captures all four island effects without postulating a second learning algorithm, one could ask if increasing the size of the *n*-grams would result in better empirical coverage. The problem with an approach that assumes an *n* greater than 3 is that there is no straightforward way to accommodate extraction from the matrix subject position, which only results in a single container node (IP). It is possible to accommodate these sequences in a trigram model by assuming symbols for *start* and *end*, resulting in *start*-IP-*end*. *Start* and *end* symbols may not be part of phrase structure grammars, but they are at least psychologically principled in that the algorithm needs to track the beginning and end of dependencies at some level. However there is no obviously principled way to incorporate an additional symbol in a 4-gram model to capture matrix subject dependencies. Alternatively, as an anonymous reviewer observes, one might argue that a learner could simply not divide these dependencies into 4-grams; rather, the learner would divide dependencies into 4-grams if they can be divided, and would leave them alone otherwise. This runs into a similar problem, however: Namely, something special must be done for dependencies that cannot be divided into 4-grams (such as matrix subject dependencies) so that they can be both learned from and learned about. This suggests that a trigram model is simpler because the 4-gram model will require an exception for these dependencies. This problem holds for every *n*-gram above 3.

Like the previous biases, the bias to track trigrams appears to be an empirical necessity (unless the learner uses additional or more complex algorithms). Also like the previous biases, it is an open question how this bias arises. Learning models based on sequences of three units have been proposed and are consistent with children's observable behavior for other linguistic knowledge (e.g., the comparison of three sequential transitional probabilities for word segmentation: Saffran et al. 1996, Aslin et al. 1998, Graf Estes et al. 2007, Pelucchi et al. 2009a, Pelucchi et al. 2009b; frequent frames consisting of three sequential units for grammatical categorization: Mintz 2006, Wang & Mintz 2008); additionally, these learning models are consistent with human behavior for non-linguistic phenomena (Saffran et al. 1996) and also with

learning behavior in non-human primates (Saffran et al. 2008). Given this, such a bias is likely domain-general; however, the fact that trigrams are an available option does not explain how it is that the learning algorithm knows to leverage trigrams (as opposed to other n-grams) for syntactic islands. The existence of certain syntactic islands in English appears to be predicated upon this choice (Complex NP and Subject islands do not arise under other choices), so this bias is inextricably linked to the question that arises throughout this discussion: Why is it that syntactic island effects exist in language at all? The explicit algorithm proposed here makes it clear that this is a problem that any theory of the learning of syntactic island constraints must address at some point.

6.6 Subcategorization of CP

In addition to a bias to track trigrams of container nodes, the proposed algorithm has a bias to track subcategories of CP based on the lexical item that introduces the CP (*that*, *whether*, *if*, and the null complementizer). Much like the other biases, this is empirically necessary: An algorithm that treats all CPs identically will fail to learn Whether islands and Adjunct islands, because the only difference between Whether and Adjunct violations and their non-island control conditions is in the type of CP (*that* versus *whether*, and *that* versus *if*). Again, like the other biases, this raises the question of how the algorithm knows what the proper set of container nodes to track is. It is logically possible to subcategorize any number of maximal projections, or none at all, or even to count intermediate projections (e.g., N′) as a container node.

The fact that CPs *can* be subcategorized is relatively straightforward. Different CPs introduce different types of clauses, with substantial semantic differences: *that* introduces declarative clauses (which are semantically propositions), *whether* introduces questions (which are semantically sets of propositions), and *if* introduces condition clauses. It may also be possible to quantify the degree of semantic difference captured by the subcategorization of different types of maximal projections, such that one could argue that the differences between CPs is greater than the differences between NPs, or the differences between VPs. However, the fact that this type of information is available to the language system does not explain how it is that the learner knows to pursue this particular strategy (or knows where to draw the line between types of container nodes). It may be possible to capture part of this behavior with innate, domain-general preferences for certain types of hypotheses (either more specific hypotheses, such as subcategorize all container nodes, or more general hypotheses, such as subcategorize no container nodes) coupled with a domain-specific proposal about the types of information that could be used to correct mistaken hypotheses. But this simply pushes back the question to one about how the system knows which evidence to look for to correct mistaken hypotheses (i.e., is it innate or derived?). In short, much like the previous biases, the empirical necessity of subcategorizing CPs raises difficult questions for any theory of the acquisition of syntactic islands.

6.7 The problems raised by the acquisition of syntactic island effects

In this section we have attempted to illustrate (i) the empirical necessity of each of the biases of the proposed learning algorithm, and (ii) the difficult questions raised by the empirical necessity of these biases. Some of the basic components of the algorithm will be part of the learning theory for any syntactic phenomenon (e.g., assigning phrase structure and tracking frequencies), but

others appear to be specific to syntactic island effects, such as restricting the input to *wh*-dependencies, tracking sequences of container nodes, segmenting container node sequences into trigrams, and subcategorizing CP container nodes by the lexical item that introduces them. These biases are interesting because on the one hand, they are significantly less specific than previous approaches to the acquisition of island effects (which tended to directly encode syntactic constraints in the learning algorithm); on the other hand, they are still specific enough to raise difficult questions about how they could arise in the learner. The explicit modeling procedure here (based on realistic input) suggests that any theory that seeks to learn syntactic islands as a type of grammatical constraint will be forced grapple with the empirical necessity of these specific biases. The other option, of course, is to deny that syntactic island effects are the result of grammatical knowledge, as has been proposed by several researchers in the past (e.g., Givón 1979, Deane 1991, Pritchett 1991, Kluender & Kutas 1993, Kluender 1998, Kluender 2004, Hofmeister & Sag 2010). The problem with the non-grammatical solution to this problem is that the currently available empirical evidence, from sentence processing studies to cross-linguistic syntactic studies, suggests that a grammatical approach to syntactic islands is much more likely (see Sprouse et al. 2012b for a brief review).

7. Consequences for syntactic theory

Historically, there have been relatively close ties between syntactic theories and acquisition theories. These ties are bi-directional: One of the goals of syntactic theory is to systematically explain the properties of language acquisition through the form of syntactic theories (beginning with at least Chomsky 1965), and the goal of acquisition theories is to explain the learning trajectory from birth to the adult target state, as defined by syntactic theories. This means that any proposal regarding the acquisition of syntactic phenomena will likely interact with both the form and empirical coverage of syntactic theories. In this section, we investigate the interaction between the proposed algorithm and existing syntactic theories, to highlight both potential problems for the proposed algorithm, and potential areas for future research.

7.1 Parasitic gaps and Across-the-Board constructions

Though this statistical learning model demonstrates that syntactic islands can in principle be learned from child-directed input, this particular model cannot capture certain known exceptions to syntactic island constraints, such as *parasitic gap* constructions (Ross 1967, Engdahl 1983). Parasitic gaps are constructions where a displaced element is associated with two gap positions, one which is licit and one which is inside a syntactic island (21). While a single gap inside an island structure results in unacceptability (21a), an additional gap outside the island seems to eliminate the unacceptability (21b) (see Phillips 2006 for experimentally collected judgments). The two gaps in these constructions are often described as the *true gap*, which occurs outside of the island, and the *parasitic gap*, which occurs inside of the island. The name is a metaphorical reference to the fact that the parasitic gap could not exist without the true gap, much like a parasite cannot exist without a host.

(21) a. *Which book did you laugh [before reading __]?
Ungrammatical gap dependency: IP-VP-CP$_{before}$-IP-VP
b. Which book did you judge __$_{true}$ [before reading __$_{parasitic}$]?

Parasitic gap dependency: IP-VP-CP$_{before}$-IP-VP

The proposed algorithm fails to capture the acceptability of parasitic gaps because the probability of the dependency involving the parasitic gap would be identical to the probability of the dependency in the structurally identical syntactic island violation (as shown by the container node sequences in (21)). So, both (21b) and (21a) would be judged as ungrammatical, as they both contain the same low probability dependency. This, of course, is not the adult target state for acquisition, and therefore suggests that the current strategy is not the precise strategy used by human language learners.

Across-the-Board (ATB) constructions (Williams 1978) also involve a displaced element that is associated with two gap positions; however, in ATB constructions, *both* of the gap positions would ordinarily be illicit, as they appear in the two conjuncts of a coordinate structure. In this way, each gap on its own would violate the Coordinate Structure Constraint (Ross 1967), but together appear to be a grammatical option, as in (22b-c).

(22) a. What did you read __ and then review __?
ATB gap dependencies (both): IP-VP-VP
b. *What did you read __ and then review the book?
Ungrammatical gap dependency: IP-VP-VP
c. *What did you read the book and then review __?
Ungrammatical gap dependency: IP-VP-VP

As with parasitic gaps, the current learning strategy would not fare well on ATB extractions: The ungrammatical gaps in (22b-c) and the grammatical gaps in (22a) are all characterized by the same container node sequence, and thus would be assigned the same status (either all grammatical, or all ungrammatical), contrary to the human adult target state.

Crucially, both parasitic gaps and ATB extractions involve combination of the information coming from each gap (which the current strategy would represent as combining the container node sequences characterizing each gap). Success in both cases involves recognizing that a gap that would be ungrammatical on its own becomes grammatical when it is combined with another gap of a specific kind. The question then becomes how children know combination *can* occur, and *what* the precise combination operation is.[14]

A preliminary investigation of 9 child-directed speech corpora containing approximately 675,000 words shows that ATB extractions do occur in the input (albeit rarely – we found only 78 examples), while parasitic gaps do not occur at all (0 examples). The fact that ATB constructions do exist in child-directed input suggests that the properties of ATB construction can in principle be learned from the input (with some combination of learning biases); however, the lack of parasitic gap examples suggests that these may either be acquired later (assuming that parasitic gaps do appear in adult-directed input), or may be learned in conjunction with ATB constructions. This latter possibility receives some support from the fact that parasitic gap and

[14] We note that if a combination operation is *always* part of the learner's treatment of utterances containing gaps, this should not affect our current results on dependencies associated with a single gap. This is because single gap dependencies would presumably be a special case for the combination operation where no combination of dependency information would need to occur.

ATB constructions share a number of complex syntactic properties, some of which are given in (23-25) (see Munn 2001 for a review and references).

(23) The dependencies must be A'-dependencies, not A-dependencies:
  a. *The book was borrowed __ after Jack read __.
  b. *The book was borrowed __ and Jack read __.

(24) The true/first conjunct gap must not c-command the parasitic/second conjunct gap
  a. *Who __ read the paper after John talked to __?
  b. *Who __ read the paper but John didn't talk to __?

(25) The parasitic/second conjunct gap cannot be inside an additional island
  a. *Which report did you file __ after wondering whether you read __?
  b. *Which report did you file __ and wonder whether you read __?

To account for these properties, some syntactic analyses (e.g., Chomsky 1986) have postulated a null operator that (i) appears at the left periphery of the adjunct phrase in parasitic gap constructions (26a) and at the left periphery of the second conjunct phrase in ATB constructions (26b), and (ii) binds the parasitic/second conjunct gap instead of the displaced *wh*-phrase:

(26) a. Which report$_i$ did you read $\_\__i$ [OP$_j$ before filing $\_\__j$]?
  b. Which report$_i$ did you read $\_\__i$ [OP$_j$ and then file $\_\__j$]?

The dependency (or chain) formed by the null operator must then be combined with the dependency between the *wh*-phrase and the true/first conjunct gap in order for the parasitic/second conjunct gap to receive the correct interpretation. The properties in (23-25) can then be accounted for by the existence of the null operator and by postulating constraints on the dependency combination operation.

Beyond providing a mechanism by which parasitic gap constructions could potentially be learned from exposure to ATB constructions, the null operator analysis provides a potential avenue for solving the problem posed by these constructions for the current learning strategy. In particular, the null operator analysis postulates two distinct dependencies in these constructions: one that holds between the *wh*-phrase and the true/first conjunct gap, and one that holds between the null operator and the parasitic/second conjunct gap. These two dependencies inside a single sentence will each have distinct container node trigram sequences. In the case of parasitic gap constructions, the syntactic island barrier (the adjunct node) will not be part of either of these sequences, thus eliminating the low probability container node sequence that leads the strategy to incorrectly predict parasitic gap constructions to be ungrammatical. The only way to introduce a syntactic island container node into these two dependencies is to insert a syntactic island in the main clause or in the adjunct clause, which, as predicted, will lead to ungrammaticality (25). Of course, the situation is more complicated for ATB constructions: Whereas the null operator dependency in an ATB construction will not contain any low probability container nodes, the dependency between the *wh*-phrase and the first conjunct gap would be illicit by itself (a Coordinate Structure Constraint (CSC) violation), so it presumably contains at least one low probability container node trigram (although we have not yet tested the CSC using the current algorithm). This means that the current learning strategy must still be expanded to include a

mechanism whereby the first conjunct gap is licit only if there is a gap in the second conjunct. That this type of complex grammatical knowledge exists in the adult state has recently been experimentally confirmed by Wagers and Phillips (2009), who demonstrated both that ATB constructions are rated as acceptable by native speakers, and that the human sentence parser actively searches for a second conjunct gap after encountering a first conjunct gap.

Beyond the problem raised by the first conjunct gap in ATB constructions, the null operator analysis also raises questions about the acquisition of the dependency combination operation: To what extent can the existence of the dependency combination operation be constructed from non-UG biases? To what extent can the constraints on the dependency combination operation that are required to fully capture properties (23-25) above be learned from non-UG biases? These are difficult questions that can only be addressed after we have expanded our child-directed speech corpora to have a better estimate of the relative frequencies of ATB and parasitic gap constructions in child-directed speech. They will require a systematic evaluation of the performance of our models with both classic examples of ATB and parasitic gaps and the ungrammatical sentences in (23-25) that have been used to identify the properties of ATB and parasitic gap constructions.

7.2 Italian *wh*-islands: High probability trigrams that are ungrammatical

Just as the current learning strategy would be forced to treat parasitic gaps as ungrammatical because any dependency that contains a very low frequency trigram is ungrammatical, the algorithm would similarly treat all dependencies that contain only higher frequency trigrams as grammatical. This is not problematic in English, as all such dependencies are in fact grammatical. However, Rizzi (1982) reports an interesting paradigm in Italian in which it looks as though simply doubling a grammatical sequence of trigrams leads to ungrammaticality (Phillips 2012b). Rizzi (1982) reports that Italian does not have *wh*-island effects the way that English does, as an extraction of an NP from a *wh*-island structure is grammatical ((27) = Rizzi's (6a)):

(27) Tuo fratello, a cui mi domando che storie abbiano raccontato, era molto preoccupato.
your brother, to whom$_1$ I wonder which stories$_2$ they have told __$_2$ __$_1$, was very worried.

…to whom$_1$ [$_{IP}$ I [$_{VP}$ wonder [$_{CP}$ which stories$_2$ [$_{IP}$ they [$_{VP}$ have told __$_2$ __$_1$]]]]]

Dependency for *to whom*: IP-VP-CP$_{wh}$-IP-VP

Rizzi analyzes this fact as evidence that the (Subjacency-based) bounding nodes in Italian are NP and CP, which correctly captures the fact that extraction from a CP is possible even when the specifier of CP is filled with a wh-phrase. This analysis makes an interesting prediction: If CP is a bounding node, extraction should not be able to cross two CPs with filled specifier positions. Rizzi reports that this prediction appears is borne out ((28)=Rizzi's (15b)):

(28) *Questo argomento, di cui mi sto domandando a chi potrei chiedere quando dovrò parlare, mi sembra sempre più complicato.
*this topic, of which$_1$ I am wondering to whom$_2$ I may ask __$_2$ when$_3$ I'll have to speak __$_1$ __$_3$, to me seems ever more complicated

…of which$_1$ [$_{IP}$ I [$_{VP}$ am wondering [$_{CP}$ to whom$_2$ [$_{IP}$ I [$_{VP}$ may ask __$_2$ [$_{CP}$ when$_3$ [$_{IP}$ I [$_{VP}$ 'll have [$_{IP}$ to [$_{VP}$ speak __$_1$ __$_3$]]]]]]]

Dependency for *of which*: IP-VP-CP$_{wh}$-IP-VP-CP$_{wh}$-IP-VP-IP-VP

The problem for the current algorithm is that the container node sequence of the ungrammatical sentence in (28) (CP$_{wh}$-IP-VP-CP$_{wh}$-IP-VP) consists of the very same trigrams that are in the grammatical sentence in (27) (CP$_{wh}$-IP-VP, IP-VP-CP$_{wh}$, and VP-CP$_{wh}$-IP). Therefore the current algorithm will treat it as grammatical. Whether sentences such as (28) are unacceptable or not is an empirical question. Nonetheless, the example serves to illustrate one of the primary limitations of the current algorithm: The grammaticality of each sentence is predicated solely upon the frequency of the individual "parts", where the parts are trigrams of container nodes. If any one trigram is low frequency, as in parasitic gaps, the model will treat the sentence as ungrammatical; if all of the trigrams are higher frequency, as in example (28), the model will treat the sentence as grammatical.

7.3 Cross-linguistic variation

The current learning strategy primarily learns the pattern of island effects for a given language from the input that it is presented. There are no additional constraints on the possible patterns of island effects imposed by the learning mechanism itself. What this means in practice is that this model predicts no constraints on the variation of island effects cross-linguistically: Any potential pattern of results (for the four island types investigated) can be derived given the correct input. The problem posed by constrained variation in island effects for the current strategy is straightforward: If there is indeed constrained variation in island effects cross-linguistically, then the current strategy would force us to conclude that the apparent constraint is simply a coincidence. The inputs of the languages in question just happened to not include the information that would be necessary to lead to the unobserved patterns of island effects.

It has been claimed in the syntactic literature that the cross-linguistic pattern of island effects is constrained. A classic example of this is again provided by Italian. Rizzi (1982) observes that whereas English exhibits WH, Complex NP, and Subject islands, Italian only appears to exhibit Complex NP islands:

(29) WH ISLAND
Tuo fratello, a cui mi domando che storie abbiano raccontato, era molto preoccupato.
your brother, to whom$_1$ I wonder which stories$_2$ have.3PL told __$_2$ __$_1$, was very worried.

(30) COMPLEX NP ISLAND
*Questo incarico, che non sapevo la novità che avrebbero affidato a te,…
this task, that not knew.1SG the news that have.3PL assigned __ to you

(31) SUBJECT ISLAND
Questo autore, di cui so che il primo libro è stato pubblicato recentemente, …
This author, by whom know.1SG that the first book __ has been published recently, …

In other words, it appears as though WH islands and Subject islands tend to covary (if a language has one, it will have the other; if it lacks one, it will lack the other). This pattern was also corroborated by Torrego (1984) for Spanish, suggesting that it may be a prevalent pattern for Romance languages. Whether this pattern holds for all languages that have displacement phenomena and island effects is an open empirical question, but it is clear that if it did, it would be a problem for the proposed strategy (unless, again, we are willing to subscribe it to coincidence). The formal experimental results of Sprouse et al. *submitted* suggest that English relative clause dependencies exhibit Subject island effects but not WH island effects, which casts some doubt on the claim that WH and Subject islands always covary. However, only future studies in several additional languages can settle this empirical question.

7.4 The complementizer *that*

Another potential issue for the current algorithm concerns complementizer *that* – specifically, because of the rarity of complementizer *that* in the input data, a learner using this model will generally disprefer dependencies using complementizer *that* (Phillips 2012b). In some cases, this may be desirable. For example, so-called *that*-trace effects are unacceptability that occurs when a gap immediately follows the complementizer *that* (32a), but does not arise when *that* is omitted (32b) (see Cowart 1997 for experimentally collected acceptability judgments). The current learning strategy can capture the distinction between these, shown in (32), using either child-directed or adult-directed input (the log-odds of (32b) versus (32a) is 7.12 for child-directed input and 5.40 for adult-directed input).

(32)

a. *Who do [$_{IP}$ you [$_{VP}$ think [$_{CP}$ that [$_{IP}$ __ [$_{VP}$ read the book]]]]]?
b. Who do [$_{IP}$ you [$_{VP}$ think [$_{CP}$ [$_{IP}$ __ [$_{VP}$ read the book]]]]]?

However, the current learning strategy will also generate a preference for object gaps without *that* (33b) compared to object gaps with *that* (33a) (the log-odds of (33b) versus (33a) is 6.61 for child-directed input and 2.81 for adult-directed input).

(33)

a. What do [$_{IP}$ you [$_{VP}$ think [$_{CP}$ that [$_{IP}$ Jack [$_{VP}$ read __]]]]]?
b. What do [$_{IP}$ you [$_{VP}$ think [$_{CP}$ [$_{IP}$ Jack [$_{VP}$ read __]]]]]?

Interestingly, Cowart (1997) reports that there is a small preference in adult acceptability judgments for (33b) over (33a), but it is significantly smaller than the preference for (32b) over (32a). In other words, there is an object *that*-trace effect, but it is much smaller than the subject *that*-trace effect. The current strategy generates relatively equal dispreference for (32a) and (33a) when using the child-directed corpora (7.12 versus 6.61), which contain relatively few instances of *that*. However, the model generates an asymmetrical dispreference that is more in line with Cowart's (1997) data when using the adult-directed corpora (5.40 versus 2.81), which contain more instances of *that*. This could be taken to be a developmental prediction of the current algorithm: Children may disprefer object gaps in embedded *that*-CP clauses more than adults,

and this dispreference will weaken as they are exposed to additional tokens of *that* in utterances containing dependencies.

7.5 The proposed algorithm and syntactic theory

In many ways, the algorithm proposed here looks very similar to existing theories of syntactic islands: Island effects arise due to constraints on sequences of abstract units derivable from phrase structure trees. This similarity is to be expected given that the syntactic analysis of long-studied phenomena such as syntactic islands have substantial empirical support (e.g., Ross 1967, Chomsky 1973, Rizzi 1980, Huang 1982, Lasnik & Saito 1984, Torrego 1984, Chomsky 1986, among many others). However, it should be noted the proposed algorithm is not a direct instantiation of any existing syntactic analysis that we are aware of. For example, while Head-Driven Phrase Structure Grammar (HPSG) does make use of container node sequences directly in the form of *extraction paths*, HPSG analyses generally do not postulate syntactic constraints on extraction paths to explain island effects (e.g., Pollard & Sag 1994). In contrast, while Government and Binding (GB) Theory does postulate syntactic constraints to explain island effects, GB analyses generally define those constraints over sequences of *bounding nodes* or *barriers*, not sequences of container nodes (e.g., Chomsky 1973, 1986). Whether the current approach of defining constraints over (trigrams of) container node sequences is more appropriate than these other approaches is an empirical question. Nonetheless, the close ties between syntactic theories and acquisition theories allow for a productive investigation of both the potential predictions and potential problems inherent in the proposed algorithm.

8. Conclusion

Given the rate of progress in cognitive science, the most lasting contribution of the current study is likely the construction of structurally annotated child-directed speech corpora that can be used by researchers interested in the acquisition of complex syntactic phenomena (freely available at http://www.socsci.uci.edu/~lpearl/CoLaLab/TestingUG/index.html as well as the derived corpora section of CHILDES: http://childes.psy.cmu.edu/derived/). We have also seen that at the level of abstraction necessary for syntactic islands, the composition of adult-directed input is not substantially different from the composition of child-directed speech. This is an important methodological point for researchers interested in syntactic acquisition, as it is often the case that large samples of syntactically annotated adult-directed input are more easily accessible and readily available than syntactically annotated child-directed speech. This suggests that it may be the case that other complex syntactic phenomena can also be studied using adult-directed input; however, given that this is an empirical question, we recommend using structurally annotated child-directed speech whenever possible.

At a theoretical level, we have also seen that syntactic islands can be learned from realistic child-directed speech without directly encoding syntactic constraints into the learning strategy. The learning strategy proposed here has some desirable properties, such as resembling the target state postulated by syntactic theories, capturing the well-known dispreference for longer dependencies, and maintaining a qualitative distinction (in principle) between dispreferred longer dependencies and truly ungrammatical dependencies. The proposed strategy also makes some interesting empirical predictions when compared to syntactic theories, some of which are beginning to find empirical support in recent formal acceptability judgment experiments.

It is also interesting to note that we were able to successfully model the acquisition of a complex syntactic phenomenon without sophisticated probabilistic inference mechanisms, such as Bayesian inference (e.g., Regier & Gahl 2004, Feldman et al. 2009, Foraker et al. 2009, Frank et al. 2009, Goldwater et al. 2009, Pearl & Lidz 2009, Pearl et al. 2011, Perfors et al. 2011).[15] Instead, a fairly simple probabilistic learning component (tracking frequencies of particular linguistic representations) was sufficient to learn the pattern from child-directed input. Given the relative complexity of syntactic islands with respect to other phenomena in syntactic theory, this suggests that there may be other (complex) syntactic phenomena that can be modeled with similarly simple probabilistic mechanisms. This may eliminate some of the concerns that have been raised about the psychological plausibility of Bayesian inference as a realistic learning mechanism for humans (e.g., see McClelland et al. 2010 for a recent review).

The process of explicitly modeling the proposed learning strategy, and testing it on both child-directed and adult-directed input, also highlighted several interesting properties of the "problem of syntactic island acquisition". We have seen that the biases in the proposed algorithm appear to be empirically necessary, suggesting that biases such as these (or at least biases that solve the same problems as these) will be present in any theory of the acquisition of syntactic island effects. It seems that any theory of islands will have to answer the following questions:

(i) Why does the system attempt to learn constraints on dependencies at all?

(ii) Why does the system treat *wh*-dependencies as separate from other dependencies like *rc*-dependencies and binding dependencies?

(iii) Why does the system track the container nodes of the dependency as opposed to other types of information about the dependencies?

(iv) Why does the system segment container node sequences into trigrams as opposed to other possible subsets?

(v) Why does the system define container nodes as maximal projections as opposed to intermediate or smaller projects?

(vi) Why does the system subcategorize CP container nodes?

Although all of these questions can be encoded with explicit biases (as in the proposed algorithm), and many of them can be characterized using the framework in section 1 such that they are not obviously innate and domain-specific (i.e., UG-based) biases, it is not the case that we can confidently rule out the role of innate, domain-specific assumptions in giving rise to these biases. Future research is necessary to determine whether each of these problems raised by the acquisition of syntactic islands can be resolved without any innate, domain-specific biases. Still, this is much more tractable now that we have access to (i) structurally annotated child-directed

[15] Of course, our model assumes that the phrase structure has already been inferred, and learning phrase structure may require sophisticated probabilistic inference methods. However, once the phrase structure is available, no sophisticated inference is required to learn syntactic island constraints, which is the learning process explicitly modeled here.

input, and (ii) explicit computational models that reveal the importance of these questions for a complete theory of syntactic island acquisition.

Acknowledgements

We would like to thank Colin Phillips, Jeff Lidz, Norbert Hornstein, Julien Musolino, Bob Berwick, Bob Frank, Virginia Valian, Alexander Clark, Misha Becker, Anne Hsu, Kamil Ud Deen, Charles Yang, Julian Pine, Terry Regier, William Sakas, Amy Perfors, Tom Roeper, two anonymous reviewers, the attendees of the Input & Syntactic Acquisition workshop held at the LSA in 2012 and at UC Irvine in 2009, and the audience at the Ecole Normale Supérieure in 2011 for numerous comments and suggestions on previous versions of this work. All errors remain our own. In addition, we are very grateful to Jessica Lee, Uma Patel, Kristen Byer, Christine Thrasher, and other members of the Computation of Language Laboratory who aided in the syntactic annotation of the child-directed speech. This work was supported in part by NSF grant BCS-0843896.

References

Abrusán, Márta. 2011. Presuppositional and Negative Islands: A Semantic Account. *Natural Language Semantics* 19(3). 257-321.

Alexopoulou, Theodora & Frank Keller. 2007. Locality, cyclicity, and resumption: At the interface between the grammar and the human sentence processor. *Language* 83. 110-160.

Aslin, Richard, Jenny Saffran & Elissa Newport. 1998. Computation of Conditional Probability Statistics by 8-Month-Old Infants. *Psychological Science* 9(4). 321-324.

Baker, Carl. 1978. *Introduction to generative transformational syntax.* Englewood Cliffs, NJ: Prentice Hall.

Baker, Carl. 1981. *The Logical Problem of Language Acquisition.* Cambridge: MIT Press.

Bates, Elizabeth, Inge Bretherton & Lynn Snyder. 1988. *From first words to grammar: Individual differences and dissociable mechanisms*. Cambridge, MA: Cambridge University Press.

Berwick, Robert, Paul Pietroski, Beraca Yankama & Noam Chomsky. 2011. Poverty of the Stimulus Revisited. *Cognitive Science* 35. 1207-1242.

Berwick, Robert & Amy Weinberg. 1984. *The grammatical basis of linguistic performance*. Cambridge, MA: The MIT Press.

Boeckx, Cedrix & Kleanthes Grohmann. 2007. Remark: Putting Phases in Perspective. *Syntax* 10. 204-222.

Braunwald, Susan. 1978. Context, word and meaning: Toward a communicational analysis of lexical acquisition. In Andrew Lock (ed.), *Action, gesture and symbol: The emergence of language*, 485-527. London: Academic Press.

Brown, Roger. 1973. *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge: The MIT Press.

Chomsky, Noam. 1973. Conditions on transformations. In Stephen Anderson & Paul Kiparsky (eds.), *A Festschrift for Morris Halle*, 237-286. New York: Holt, Rinehart and Winston.

Chomsky, Noam. 1980. *Rules and representations*. Oxford: Basil Blackwell.

Chomsky, Noam. 1981. *Lectures on Government and Binding.* Dordrecht: Foris.

Chomsky, Noam. 1986. *Barriers*. Cambridge: The MIT Press.

Chomsky, Noam. 1988. *Language and problems of knowledge: The Managua lectures.* Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: A life in language*, 1-52. Cambridge, MA: MIT Press.

Cowart, Wayne. 1997. *Experimental syntax: Applying objective methods to sentence judgments*. Thousand Oaks, CA: Sage.

Crain, Stephen. 1991. Language acquisition in the absence of experience. *Behavioral and Brain Sciences* 14. 597-612.

Crain, Stephen & Janet Fodor. 1985. How can grammars help parsers? In David Dowty, Lauri Karttunen, & Arnold Zwicky (eds.), *Natural language parsing: Psycholinguistic, computational, and theoretical approaches,* 94–128. Cambridge: Cambridge University Press.

Crain, Stephen & Paul Pietroski. 2002. Why language acquisition is a snap. *The Linguistic Review* 19. 163-183.

de Villiers, Jill G. & Tom Roeper. 1995. Relative clauses are barriers to Wh-movement for young children. *Journal of Child Language* 22. 389-404.

de Villiers, Jill G., Tom Roeper, Linda Bland-Stewart & Barbara Pearson. 2008. Answering hard questions: Wh-movement across dialects and disorder. *Applied Psycholinguistics* 29. 67-103.

Deane, Paul. 1991. Limits to attention: a cognitive theory of island phenomena. *Cognitive Linguistics* 2 1-63.

Denison, Stephanie, Christie Reed & Fei Xu. 2011. The emergence of probabilistic reasoning in very young infants. In *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society,* Boston, MA.

Dewar, Kathryn & Fei Xu. 2010. Induction, Overhypothesis, and the Origin of Abstract Knowledge: Evidence from 9-Month-Old Infants, *Psychological Science* 21(12). 1871-1877.

Dresher, Elan. 1999. Charting the learning path: Cues to parameter setting. *Linguistic Inquiry* 30. 27-67.

Dresher, Elan. 2003. Meno's paradox and the acquisition of grammar. In Stefan Ploch (ed.), *Living on the edge: 28 papers in honour of Jonathan Kaye (Studies in Generative Grammar 62),* 7–27. Berlin: Mouton de Gruyter.

Engdahl, Elisabet. 1980. Wh-constructions in Swedish and the relevance of subjacency. In J. T. Jensen (ed.), *Cahiers Linguisticques D'Ottawa: Proceedings of the Tenth Meeting of the North East Linguistic Society*, 89-108. Ottawa, ONT: University of Ottawa Department of Linguistics.

Engdahl, Elisabet. 1983. Parasitic Gaps. *Linguistic Inquiry* 6(1). 5-34.

Erteschik-Shir, Nomi. 1973. *On the nature of island constraints.* Cambridge, MA: MIT dissertation.

Feldman, Naomi, Thomas Griffiths & James Morgan. 2009. Learning phonetic categories by learning a lexicon. In Niels Taatgen & Hedderik van Rijn (eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society,* 2208-2213. Amsterdam, Netherlands.

Fodor, Janet. 1978. Parsing strategies and constraints on transformations. *Linguistic Inquiry* 9. 427-473.

Fodor, Janet. 1983. Phrase structure parsing and the island constraints. *Linguistics and Philosophy* 6. 163-223.

Fodor, Janet. 1998a. Unambiguous Triggers. *Linguistic Inquiry* 29. 1-36.

Fodor, Janet. 1998b. Parsing to learn. *Journal of Psycholinguistic Research* 27(3). 339-374.

Fodor, Janet. 2009. Syntax Acquisition: An Evaluation Measure After All? In Massimo Piatelli Palmarini, Juan Uriagereka, & Pello Salaburu (eds.), *Of Minds and Language: The Basque Country Encounter with Noam Chomsky*, Oxford University Press.

Foraker, Stephani, Terry Regier, Naveen Khetarpal, Amy Perfors, & Joshua Tenenbaum. 2009. Indirect evidence and the poverty of the stimulus: The case of anaphoric *one*. *Cognitive Science* 33. 287–300.

Frank, Michael, Noah Goodman, & Joshua Tenenbaum. 2009. Using Speakers' Referential Intentions to Model Early Cross-Situational Word Learning. *Psychological Science* 20(5). 578-585.

Frazier, Lyn. & Giovanni Flores d'Arcais. 1989. Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language,* 28. 331–344.

Gerken, LouAnn. 2006. Decision, decisions: infant language learning when multiple generalizations are possible. *Cognition*, 98. B67-B74.

Gibson, Edward & Kenneth Wexler. 1994. Triggers, *Linguistic Inquiry,* 25. 355-407.

Givón, Talmy. 1979. *On Understanding Grammar.* New York: Academic Press.

Goldberg, Adele. 2007. *Constructions at work.* Oxford: Oxford University Press.

Goldwater, Sharon, Thomas Griffiths, & Mark Johnson. 2009. A Bayesian Framework for Word Segmentation: Exploring the Effects of Context. *Cognition,* 112(1). 21-54.

Goodluck, Helen, Michele Foley, & Julie Sedivy. 1992. Adjunct islands and acquisition. In Helen Goodluck (ed.), *Island constraints,* 181-194. Dordrecht: Kluwer.

Graf Estes, Katharine, Julia Evans, Martha Alibali, & Jenny Saffran. 2007. Can Infants Map Meaning to Newly Segmented Words? *Psychological Science,* 18(3). 254-260.

Griffiths, Thomas & Joshua Tenenbaum. 2005. Structure and strength in causal induction. *Cognitive Psychology,* 51. 334-384.

Hagstrom, Paul. 1998. *Decomposing Questions*. Cambridge, MA: MIT dissertation.

Hale, John. 2001. A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics*, 159–166. Association for Computational Linguistics: Stroudsburg, PA.

Hart, Betty & Todd Risley. 1995. *Meaningful differences in the everyday experience of young American children.* Baltimore, MD: P.H. Brookes.

Hawkins, John. A. 1999. Processing complexity and filler-gap dependencies across grammars. *Language* 75. 244–285.

Hofmeister, Philip & Ivan Sag. 2010. Cognitive constraints and island effects. *Language* 86. 366–415.

Hornstein, Norbert & David Lightfoot. 1981. Introduction. In Norbert Hornstein (ed.), *Explanation in linguistics: The logical problem of language acquisitions,* 9–31. London: Longman.

Huang, C.-T. James. 1982. Logical relations in Chinese and the theory of grammar. Doctoral dissertation. MIT, Cambridge, MA.

Klein, Dan & Christopher Manning. 2002. A Generative Constituent-Context Model for Improved Grammar Induction. In *Proceedings of the 40th Annual Meeting for the Association for Computational Linguistics,* 128-135. Association for Computational Linguistics: Stroudsburg, PA.

Kluender, Robert. 1998. On the distinction between strong and weak islands: A processing perspective. In Peter Culicover & Louise McNally (eds.), *Syntax and Semantics 29: The Limits of Syntax*, 241–279. New York: Academic Press.

Kluender, Robert. 2004. Are Subject islands subject to a processing account? In Vineeta Chand, Ann Kelleher, Angelo Rodriguez, and Benjamin Schmeiser (eds.), *Proceedings of the West Coast Conference on Formal Linguistics 23*, 101–125. Somerville, MA: Cascadilla Press.

Kluender, Robert, & Marta Kutas. 1993. Subjacency as a processing phenomenon. *Language and Cognitive Processes* 8. 573-633.

Lasnik, Howard. & Mamuro Saito. 1984. On the nature of proper government. *Linguistic Inquiry* 15. 235-289.

Legate, Julie & Charles Yang. 2007. Morphosyntactic learning and the development of tense. *Language Acquisition* 14(3). 315-344.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106. 1126–1177.

Levy, Roger. 2011. Integrating surprisal and uncertain-input models in online sentence comprehension: formal techniques and empirical results. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics: Stroudsburg, PA.

Lidz, Jeffrey, Sandra Waxman, & Jennifer Freedman. 2003. What infants know about syntax but couldn't have learned: experimental evidence for syntactic structure at 18 months. *Cognition* 89. B65–B73.

Lightfoot, David. 1989. The child's trigger experience: degree-0 learnability. *Behavioral and Brain Sciences* 12. 321–334.

Lightfoot, David. 1991. *How to Set Parameters: Arguments from language change*. Cambridge, MA: MIT Press.

Lightfoot, David. 2010. Language acquisition and language change. *Wiley Interdisciplinary Reviews: Cognitive Science* 1. 677-684. doi: 10.1002/wcs.39.

MacWhinney, Brian. 2000. *The CHILDES project: Tools for analyzing talk,* 3rd edition. Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, Brian. 2004. A multiple process solution to the logical problem of language acquisition. *Journal of Child Language* 31. 883–914.

Manning, Christopher & Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

Marcus, Mitchell, Beatrice Santorini, Mary Ann Marcinkiewicz, & Ann Taylor. 1999. *Treebank-3.* Linguistic Data Consortium, Philadelphia.

McClelland, James, Matthew Botvinick, David Noelle, David Plaut, Timothy Rogers, Mark Seidenberg, & Linda Smith. 2010. Letting structure emerge: Connectionist and dynamical systems approaches to understanding cognition. *Trends in Cognitive Sciences* 14. 348-356.

McKinnon, Richard & Lee Osterhout. 1996. Event-related potentials and sentence processing: Evidence for the status of constraints on movement phenomena. *Language and Cognitive Processes* 11(5). 495-523.

McMurray, Bob & George Hollich. 2009. Core computational principles of language acquisition: can statistical learning do the job? Introduction to Special Section. *Developmental Science* 12(3). 365-368.

Mintz, Toben. 2003. Frequent frames as a cue for grammatical categories in child directed speech. *Cognition* 90. 91-117.

Mintz, Toben. 2006. Finding the verbs: distributional cues to categories available to young learners. In Kathy Hirsh-Pasek & Roberta Golinkoff (eds.), *Action Meets Word: How Children Learn Verbs,* 31-63. New York: Oxford University Press.

Mitchener, William & Misha Becker. 2011. Computational Models of Learning the Raising-Control Distinction. *Research on Language and Computation* 8(2). 169-207.

Nishigauchi, Taisuke. 1990. *Quantification in the Theory of Grammar.* Dordrecht: Kluwer.

Niyogi, Partha & Robert Berwick. 1996. A language learning model for finite parameter spaces. *Cognition* 61. 161-193.

Pearl, Lisa. 2008. Putting the Emphasis on Unambiguous: The Feasibility of Data Filtering for Learning English Metrical Phonology. In Harvey Chan, Heather Jacob, & Enkeleida Kapia (eds.) *BUCLD 32: Proceedings of the 32nd Annual Boston University Conference on Child Language Development,* 390-401. Somerville: MA: Cascadilla Press.

Pearl, Lisa. 2011. When unbiased probabilistic learning is not enough: Acquiring a parametric system of metrical phonology. *Language Acquisition* 18(2). 87-120.

Pearl, Lisa, Sharon Goldwater, & Mark Steyvers. 2011. Online Learning Mechanisms for Bayesian Models of Word Segmentation, *Research on Language and Computation* 8(2). 107-132.

Pearl, Lisa & Jeffrey Lidz. 2009. When domain-general learning fails and when it succeeds: Identifying the contribution of domain-specificity. *Language Learning and Development* 5(4). 235–265.

Pearl, Lisa & Jeffrey Lidz. In press. Parameters in Language Acquisition. In Kleanthes Grohmann & Cedric Boeckx (eds.), *The Cambridge Handbook of Biolinguistics.* Cambridge: Cambridge University Press.

Pearl, Lisa & Benjamin Mis. 2011. How Far Can Indirect Evidence Take Us? Anaphoric One Revisited. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society.* Boston, MA: Cognitive Science Society.

Pearl, Lisa & Benjamin Mis. 2012. Induction problems, indirect positive evidence, and Universal Grammar: Anaphoric One revisited. Ms., University of California, Irvine.

Pearl, Lisa & Jon Sprouse. In press. Computational Models of Acquisition for Islands, In Jon Sprouse & Norbert Hornstein (eds), *Experimental Syntax and Islands Effects*. Cambridge University Press.

Pearl, Lisa & Amy Weinberg. 2007. Input Filtering in Syntactic Acquisition: Answers from Language Change Modeling, *Language Learning and Development* 3(1). 43-72.

Perfors, Amy, Joshua Tenenbaum, & Terry Regier. 2011. The learnability of abstract syntactic principles. *Cognition* 118. 306–338.

Pelucchi, Bruna, Jessica Hay, & Jenny Saffran. 2009a. Statistical Learning in Natural Language by 8- Month-Old Infants. *Child Development* 80(3). 674-685.

Pelucchi, Bruna, Jessica Hay, & Jenny Saffran, J. 2009b. Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition* 113(2). 244-247.

Phillips, Colin. 2006. The real-time status of island constraints. *Language* 82. 795-823.

Phillips, Colin. 2012a. On the Nature of Island Constraints I: Language Processing and Reductionist Accounts. In Jon Sprouse and Norbert Hornstein (eds.), *Experimental syntax and island effects.* Cambridge: Cambridge University Press.

Phillips, Colin. 2012b. On the Nature of Island Constraints II: Language Learning and Innateness. In Jon Sprouse and Norbert Hornstein (eds.), *Experimental syntax and island effects.* Cambridge: Cambridge University Press.

Pollard, Carl & Ivan Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Pritchett, Bradley. 1991. Subjacency in a Principle-Based Parser. In Robert Berwick, Steven Abney, & Carol Tenney (eds.), *Principle-Based Parsing: Computation and Psycholinguistics*, 301–345. Dordrecht: Kluwer.

Pullum, Geoffrey & Barbara Scholz. 2002. Empirical assessment of stimulus poverty arguments. *The Linguistic Review* 19. 9–50.

Regier, Terry & Susanne Gahl. 2004. Learning the unlearnable: The role of missing evidence. *Cognition* 93. 147–155.

Reinhart, Tanya. 1997. Quantifier Scope: How Labor is Divided Between QR and Choice Functions. *Linguistics and Philosophy* 20. 335-397.

Rizzi, Luigi. 1982. Violations of the wh-island constraint and the subjacency condition. In Luigi Rizzi (ed.), *Issues in Italian Syntax.* Dordrecht, NL: Foris.

Rizzi, Luigi. 1991. *Relativized minimality.* Cambridge, MA: MIT Press.

Roeper, Tom & Jill de Villiers. 2011. The Acquisition Path for Wh-Questions. In Jill de Villiers & Tom Roeper (eds.), *Handbook of Generative Approaches to Language Acquisition*, *Studies in Theoretical Psycholinguistics 41*, 189-246. Springer: New York.

Ross, John. 1967. Constraints on variables in syntax. Cambridge, MA: MIT dissertation.

Saffran, Jenny, Richard Aslin, & Elissa Newport. 1996. Statistical Learning by 8-Month-Old Infants. *Science* 274. 1926-1928.

Saffran, Jenny, Marc Hauser, Rebecca Seibel, Joshua Kapfhamer, Fritz Tsao, & Fiery Cushman. 2008. Grammatical pattern learning by infants and cotton-top tamarin monkeys. *Cognition* 107. 479-500.

Saffran, Jenny, Elizabeth Johnson, Richard Aslin, & Elissa Newport. 1999. Statistical learning of tone sequences by human infants and adults. *Cognition* 70(1). 27-52.

Sagae, Kenji, Eric Davis, Alon Lavie, Brian MacWhinney, & Shuly Wintner. 2010. Morphosyntactic annotation of CHILDES transcript. *Journal of Child Language* 37(3). 705-729.

Sakas, William & Janet Fodor. 2001. The structural triggers learner. In Stefano Bertolo (ed.), *Language Acquisition and Learnability*, 172-233. Cambridge, UK: Cambridge University Press.

Sampson, Geoffrey. 1989. Language acquisition: growth or learning? *Philosophical Papers* 18. 203- 240.

Sampson, Geoffrey. 1999. Collapse of the language nativists. *The Independent*, *April 9, 1999*, 7.

Scholz, Barbara & Geoffrey Pullum. 2002. Searching for arguments to support linguistic nativism. *The Linguistic Review,* 19. 185–223.

Schütze, Carson. 1996. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: The University of Chicago Press.

Schütze, Carson & Jon Sprouse. In press. Judgment Data. In Devyani Sharma & Robert Podevsa (eds.), *Research Methods in Linguistics*.

Soderstrom, Melanie, Erin Conwell, Naomi Feldman, & James Morgan. 2009. The learner as statistician: three principles of computational success in language acquisition. *Developmental Science* 12(3). 409-411.

Sprouse, Jon. In press. Deriving competing predictions from grammatical approaches and reductionist approaches to island effects. In Jon Sprouse & Norbert Hornstein (eds.), *Experimental Syntax and Island Effects*. Cambridge University Press.

Sprouse, Jon & Diogo Almeida. In press. The role of experimental syntax in an integrated cognitive science of language. In Kleanthes Grohmann & Cedric Boeckx (eds.), *The Cambridge Handbook of Biolinguistics*. Cambridge: Cambridge University Press.

Sprouse, Jon, Ivano Caponigro, Ciro Greco, & Carlo Ceccheto. Submitted. Experimental syntax and the cross-linguistic variation of island effects in English and Italian. Ms. University of California, Irvine.

Sprouse, Jon, Matt Wagers, & Colin Phillips. 2012a. A test of the relation between working memory capacity and syntactic island effects. *Language* 88(1). *82-124.*

Sprouse, Jon, Matt Wagers, & Colin Phillips. 2012b. Working memory capacity and island effects: A reminder of the issues and the facts. *Language* 88(2). *401-407.*

Stowe, Laurie. 1986. Parsing wh-constructions: Evidence for on-line gap location. *Language and Cognitive Processes* 1. 227–245.

Suppes, Patrick. 1974. The semantics of children's language. *American Psychologist* 29. 103–114.

Szabolcsi, Ana & Frans Zwarts. 1993. Weak islands and an algebraic semantics of scope taking. *Natural Language Semantics* 1. 235-284.

Tenenbaum, Joshua & Thomas Griffiths. 2001. Generalization, similarity and Bayesian inference. *Behavioral and Brain Sciences* 24. 629-640.

Tomasello, Michael. 2004. What kind of evidence could refute the UG hypothesis? *Studies in Language* 28(3). 642-645.

Torrego, Esther. 1984. On Inversion in Spanish and Some of Its Effects, *Linguistic Inquiry* 15. 103-129.

Traxler, Matthew & Martin Pickering. 1996. Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language, 35,* 454-475.

Tribus, Myron. 1961. *Thermodynamics and Thermostatics: An Introduction to Energy, Information and States of Matter, with Engineering Applications*. New York, NY: D. Van Nostrand Company Inc.

Truswell, Robert. 2007. Extraction from adjuncts and the structure of events. *Lingua* 117. 1355–1377.

Tsai, Wei-Tien. 1994. On nominal islands and LF extraction in Chinese. *Natural Language and Linguistic Theory* 12. 121-75.

Valian, Virginia. 1991. Syntactic subjects in the early speech of American and Italian children. *Cognition* 40. 21–81.

Viau, Joshua & Jeffrey Lidz. 2011. Selective learning in the acquisition of Kannada ditransitives. *Language* 87. 679-714.

Wagers, Matt & Colin Phillips. 2009. Multiple dependencies and the role of the grammar in real-time comprehension. *Journal of Linguistics* 45. 395-433.

Wang, Hao & Toben Mintz. 2008. A Dynamic Learning Model for Categorizing Words Using Frames. In Harvey Chan, Heather Jacob, & Enkeleida (eds.), *BUCLD 32 Proceedings*, 525-536. Somerville, MA: Cascadilla Press.

Xu, Fei & Joshua Tenenbaum. 2007. Word learning as Bayesian inference. *Psychological Review* 114. 245-272.

Yang, Charles. 2002. *Knowledge and Learning in Natural Language.* Oxford: Oxford University Press.

Yang, Charles. 2004. Universal Grammar, statistics, or both? *Trends in Cognitive Sciences* 8(10). 451-456.

Appendix A. Formal descriptions of procedures.

The descriptions of the procedures used for learning and generating grammaticality preferences are given below in pseudocode format. The description for the grammatical preference generation highlights how the complete set of trigrams is generated (based off the container nodes encountered during learning) and how the trigram smoothing is implemented.

(A1) Pseudocode description of the learning algorithm

```
for each input utterance u
  for each wh-dependency w in u
    characterize w as a sequence of container nodes cns
      divide cns into a sequence of container node trigrams cnts
      for each container node trigram cn1-cn2-cn3 in cnts
        cn1-cn2-cn3_count = cn1-cn2-cn3_count + 1
```

(A2) Pseudocode description of the generation of grammaticality preferences

```
# get full set of possible trigrams
cn_set = set of container nodes encountered during learning period

  # trigrams beginning with start
  for each cn1 in cn_set
    for each cn2 in cn_set
      if start-cn1-cn2 exists as a trigram
          start-cn1-cn2_count = start-cn1-cn2_count + alpha
      else
          start-cn1-cn2_count = alpha
      total_trigram_count = total_trigram_count + start-cn1-cn2_count

  # trigrams with container nodes in all slots
  for each cn1 in cn_set
    for each cn2 in cn_set
      for each cn3 in cn_set
        if cn1-cn2-cn3 exists as a trigram
          cn1-cn2-cn3-count = cn1-cn2-cn3_count + alpha
        else
          cn1-cn2-cn3_count = alpha
        total_trigram_count = total_trigram_count + cn1-cn2-cn3_count

  # trigrams ending with end
  for each cn2 in cn_set
    for each cn3 in cn_set
      if cn2-cn3-end exists as a trigram
          cn2-cn3-end_count = cn2-cn3-end_count + alpha
      else
          cn2-cn3-end_count = alpha
      total_trigram_count = total_trigram_count + cn2-cn3-end_count
```

```
# calculate trigram probabilities for all trigrams
# assume each trigram has the form cnx-cny-cnz
for each trigram cnx-cny-cnz in complete set of trigrams
   cnx-cny-cnz_probability = (cnx-cny-cnz_count)/(total_trigram_count)

# generate grammaticality preferences for wh-dependency w
wh_probability = 1
characterize w as a sequence of container nodes cns
  divide cns into a sequence of container node trigrams cnts
  for each container node trigram cnx-cny-cnz in cnts
    wh_probability = wh_probability * cnx-cny-cnz_probability
```

Appendix B. Distribution of *wh*-dependencies in the input.

Table B1. Description of child-directed and adult-directed input corpora. Percentages are shown for container node sequences, based on the total *wh*-dependencies in each corpus, with the quantity observed in the corpus on the line below. An example of each container node sequence is given below the sequence.

Container node sequence and example utterance	Child-directed: speech	Adult-directed: speech	Adult-directed: text
IP Who saw it?	12.8% 2680	17.2% 1464	33.0% 1396
IP-VP What did she see?	76.7% 16039	73.0% 6215	63.3% 2677
IP-VP-AdjP-IP-VP What are you willing to see?	0.0% 0	<0.1% 1	0.1% 5
IP-VP-AdjP-IP-VP-PP What are you willing to go to?	0.0% 0	<0.1% 1	0.0% 0
IP-VP-AdjP-PP What are they good for?	0.0% 0	<0.1% 1	<0.1% 1
IP-VP-CP_{for}-IP-VP-PP What did she put on for you to dance to?	<0.1% 1	0.0% 0	0.0% 0
IP-VP-CP_{null}-IP Who did he think stole it?	0.1% 24	0.6% 52	0.3% 12
IP-VP-CP_{null}-IP-VP What did he think she stole?	1.1% 236	0.4% 30	0.2% 8
IP-VP-CP_{null}-IP-VP-IP-VP What did he think she wanted to steal?	0.1% 28	<0.1% 3	0.0% 0
IP-VP-CP_{null}-IP-VP-IP-VP-IP-VP What did he think she wanted to pretend to steal?	<0.1% 2	0.0% 0	0.0% 0

IP-VP-CP_{null}-IP-VP-IP-VP-IP-VP-PP Who did he think she wanted to pretend to steal from?	0.0% 0	<0.1% 1	0.0% 0
IP-VP-CP_{null}-IP-VP-IP-VP-PP Who did he think she wanted to steal from?	<0.1% 1	0.0% 0	0.0% 0
IP-VP-CP_{null}-IP-VP-NP What did he think she said about it?	<0.1% 1	<0.1% 5	<0.1% 1
IP-VP-CP_{null}-IP-VP-PP What did he think she wanted it for?	0.1% 28	<0.1% 5	<0.1% 1
IP-VP-CP_{null}-IP-VP-PP-PP What did he think she wanted out of?	<0.1% 1	0.0% 0	0.0% 0
IP-VP-CP_{that}-IP-VP What did he think that she stole?	<0.1% 2	<0.1% 5	<0.1% 2
IP-VP-CP_{that}-IP-VP-IP-VP What did he think that she wanted to steal?	0.0% 0	<0.1% 1	0.0% 0
IP-VP-CP_{that}-IP-VP-PP Who did he think that she wanted to steal from?	0.0% 0	<0.1% 1	0.0% 0
IP-VP-IP Who did he want to steal the necklace?	<0.1% 9	<0.1% 2	0.0% 0
IP-VP-IP-VP What did he want her to steal?	5.6% 1167	3.4% 287	1.3% 57
IP-VP-IP-VP-IP-VP What did he want her to pretend to steal?	<0.1% 11	<0.1% 6	<0.1% 1
IP-VP-IP-VP-IP-VP-PP Who did he want her to pretend to steal from?	0.2% 43	<0.1% 6	0.0% 0
IP-VP-IP-VP-NP What did he want to say about it?	<0.1% 6	0.0% 0	0.0% 0
IP-VP-IP-VP-NP-IP-VP What did he have to give her the opportunity to steal?	0.0% 0	0.0% 0	<0.1% 1
IP-VP-IP-VP-NP-PP What did she want to steal more of?	<0.1% 1	<0.1% 1	0.0% 0
IP-VP-IP-VP-PP	0.4%	0.4%	<0.1%

What did she want to steal from?	74	33	4
IP-VP-IP-VP-PP-PP What did she want to get out from under?	0.0% 0	0.0% 0	<0.1% 1
IP-VP-NP What did she say about the necklace?	0.2% 52	0.1% 10	0.1% 5
IP-VP-NP-IP-VP What did he give her the opportunity to steal?	0.0% 0	<0.1% 1	<0.1% 2
IP-VP-NP-PP What was she a member of?	<0.1% 7	<0.1% 6	0.0% 0
IP-VP-PP Who did she steal from?	2.5% 524	4.3% 369	1.3% 57
IP-VP-PP-CP_{null}-IP What did she feel like was a very good place?	0.0% 0	<0.1% 1	0.0% 0
IP-VP-PP-CP_{null}-IP-VP What did she feel like he saw?	<0.1% 1	0.0% 0	0.0% 0
IP-VP-PP-IP-VP What did she think about buying?	0.0% 0	<0.1% 3	0.0% 0
IP-VP-PP-NP Where was she at in the building?	0.0% 0	<0.1% 2	0.0% 0
IP-VP-PP-NP-PP What do you put it on top of?	<0.1% 2	0.0% 0	0.0% 0
IP-VP-PP-NP-PP-IP-VP What is she in the habit of doing?	0.0% 0	<0.1% 1	0.0% 0
IP-VP-PP-PP What does he eat out of?	0.1% 22	0.0% 0	0.0% 0
IP-VP-PP-IP-VP What did he think about stealing?	<0.1% 1	0.0% 0	0.0% 0