Computational modeling for language acquisition: A tutorial with syntactic islands

Lisa S. Pearl and Jon Sprouse
lpearl@uci.edu, jon.sprouse@uconn.edu

Abstract

Purpose: Given the growing prominence of computational modeling in the acquisition research community, we present a tutorial on how to use computational modeling to investigate learning strategies that underlie the acquisition process. This is useful for understanding both typical and atypical linguistic development.

Methods: We provide a general overview of why modeling can be a particularly informative tool and some general considerations when creating a computational acquisition model. We then review a concrete example of a computational acquisition model for complex structural knowledge referred to as syntactic islands. This includes an overview of syntactic islands knowledge, a precise definition of the acquisition task being modeled, the modeling results, and how to meaningfully interpret those results in a way that is relevant for questions about knowledge representation and the learning process.

Conclusions: Computational modeling is a powerful tool that we can use to understand linguistic development. The general approach presented here can be used to investigate any acquisition task and any learning strategy, provided both are precisely defined.

Keywords: language acquisition; computational modeling; syntactic islands; syntax

1 Introduction

This article is intended as a brief tutorial on using computational modeling in an informative way, specifically applied to problems in language acquisition. Because computational modeling is a tool growing in prominence in the acquisition research community, we highlight what it is especially useful for: investigating the strategies that underlie the acquisition process. We then present a general overview of the relevant considerations when modeling, and provide a concrete example of how to apply it to yield results pertinent to researchers investigating acquisition – whether they are interested in how it typically succeeds or the different ways in which it may be delayed or break down completely. Our example study focuses on the acquisition of complex structural knowledge often referred to as *syntactic islands*. We demonstrate how computational modeling forces us to be explicit about all the components in a syntactic island learning strategy, and by doing so, provides a list of individual pieces that may be further investigated. We believe that modeling results like these will be useful both for basic cognitive science (e.g., contributing to debates between linguistic nativists and constructionists) and for clinically-oriented cognitive science (e.g., providing a comparison point for theories of atypical development).

1.1 A tool for language acquisition

Broadly speaking, the goal of the language acquisition field is to determine the learning strategies that children use to acquire their native language. Like all sciences, there are two general steps to this process: *theory specification* and *theory evaluation*. Theory specification involves constructing theories that could potentially solve the learning problem that children face. Theory evaluation involves comparing the predictions of competing theories to data from children to determine which theory is more likely to be correct. Historically, these two steps have been conducted via human theorizing (specification) and experimental research (evaluation), both of which are necessary components for discovering how it is that children acquire language. We believe that computational modeling is additionally beneficial to both theory specification and theory evaluation, in ways that are complementary to classic human theorizing and experimental work. So, this is a core benefit of computational modeling.

On the specification side, computational modeling can allow us to construct theories of language learning strategies that are explicit about all the necessary components. This is not a trivial capability – it often turns out that learning theories that seem explicit to humans do not actually specify all the details necessary to implement the strategies they describe. A computational model highlights where the gaps are, because a computer program can only implement a learning strategy where every relevant detail is specified (Kol, Nir, & Wintner, 2014; Pearl, 2014). So, even if a learning theory has already been developed, a computational model provides a way to flesh out that theory and identify the necessary components.

On the evaluation side, a computational model provides a testing ground for a fleshed out learning theory, since it sets up a particular learning scenario and implements the theory's learning strategy within that learning scenario. If the learning strategy succeeds, this is support for the strategy instantiated that way; if it fails, this is an argument against that instantiation's validity. We emphasize that a computational model of a learning strategy provides a particular kind of

evidence: namely, that this instantiation of the strategy *can* work. It does not mean this is the only instantiation (or only strategy) that can work. Also, if the instantiation of the learning strategy fails, this does *not* mean the strategy in general has been proven to fail, since it is possible that a different instantiation might succeed. So, it is important to interpret computational modeling findings appropriately.

Still, with this information in hand, experimentalists can also conduct their own evaluations of the learning theories by investigating the necessary strategy components as well as any predictions generated by the modeled learning strategy (e.g., the order of acquisition for different pieces of linguistic knowledge or the developmental trajectory of some piece of knowledge). If the strategy components seem to be utilized by children of the appropriate age and the model's predictions match observable behavioral data, then this provides additional empirical support for the learning theory that proposed that strategy.

1.2   What makes a useful computational model

In general, a complete computational model must contain an explicit specification of the components of the acquisition process. Drawing from several recent discussions of the acquisition process (Lidz & Gagliardi, 2015; Omaki & Lidz, 2014; Pearl & Sprouse, 2013b; Pearl, 2014; Pearl & Mis, 2011, in press), we suggest the following components:

1. *initial state:* What knowledge, abilities, and learning biases does the learner already have?

2. *encoding:* How does the modeled child perceive the input?

3. *intake:* What part of the perceived input does the modeled child use for learning?

4. *learning period:* How long does the modeled child have to learn?

5. *inference:* How are updates to the modeled child's internal representation made?

6. *iteration:* How does the internal representation affect subsequent encoding, intake, and observable behavior?

7. *target state:* What does it mean for the modeled child to succeed at learning?

Crucially, we will want each of these components to be cognitively plausible for children. To that end, we can draw on prior theoretical, corpus, and experimental research to determine what is realistic. For example, theoretical research can help define the initial state and the target knowledge state, and proposals for the inference and iteration processes. Corpus research can help define the set of data available to children. Experimental research can help determine what abilities the learner has in the initial state, how the child's developing cognitive abilities affect data encoding and intake, how long the learning period is, what plausible inference and iterations procedures are for the modeled child's age, what behavior children display when they are in the process of acquiring the target knowledge (i.e., the developmental trajectory), and what behavior children display when they have acquired the target knowledge.

We note that making each component "cognitively plausible" is still non-trivial, despite all this information. For example, consider the more general issue of the learner's input, which encompasses both the encoding and intake components. For some acquisition problems (such as the syntactic islands problem we discuss below), data about the distribution of child-directed utterances is sufficient, and a reasonable approximation of this input is available through resources such as the CHILDES database (MacWhinney, 2000). However, for other acquisition problems, richer input data may be required (e.g., visual scenes paired with utterances), and such data may not be available in large enough quantities to easily estimate a learner's input. For this reason, one important contribution to modeling work is the creation of this kind of richer data set, as well as ways to automatically generate it from existing resources (Matusevych, Alishahi, & Vogt, 2013).

The learning period component can also present a plausibility challenge. For some acquisition problems, we have empirical data about exactly what children know when; for others, we may only have the typically developing adult knowledge state or an atypically developing learner's final knowledge state. In the latter case, in may difficult to determine what should count as a plausible learning period. Because modeling requires us to be explicit, modelers may have to simply make an educated guess about how long might be plausible for learning until empirical data do become available.

The inference component also raises practical considerations, in addition to plausibility considerations. For many acquisition problems, the empirical data may be compatible with many different inference algorithms (e.g., Bayesian inference, neural networks, reinforcement learning). For some acquisition problems, the choice of inference algorithm may be more a matter of taste, as different algorithms have different benefits and drawbacks, while for other problems, it may be that certain algorithms lend themselves more naturally to the task (e.g., see Pearl and Goldwater (in press) for discussion of why Bayesian inference may be a natural fit for several language acquisition tasks). As we will see in the model discussed below for syntactic island constraints, a very simple inference algorithm was sufficient for implementing the proposed learning strategy.

The target state can also present practical challenges, as we consider what exactly the learner ought to learn and what behavior that learner ought to produce in order to demonstrate that the learner has indeed learned what it ought to learn. If we have detailed empirical data available about the stages of learning, this can be a reasonable comparison for the learner's output – we can try to capture the appropriate learning trajectory (e.g., Alishahi & Stevenson, 2008). However, if we don't, we may need to rely on other measures of what counts as acquisition success (Pearl, 2014). Perhaps the learner should reach adult-like knowledge and so we can use behavior correlated with adult knowledge as a metric for learner behavior (e.g., Pearl & Sprouse, 2013a, 2013b). Perhaps we have a measure of behavior at one particular age (when the child may not yet have the adult knowledge), and we can use that behavior as a metric of what the learner should have learned by that age (e.g., Pearl & Mis, in press). Perhaps we know that the target knowledge will be used to bootstrap future acquisition processes, and so we should measure how useful the learner's acquired knowledge is, regardless of whether it matches the adult knowledge (e.g., L. Phillips & Pearl, 2014a, 2014b, in press).

In general, it is important to empirically and psychologically ground the choices for the different components whenever possible. This increases confidence that the computational model we

create is informative about how humans solve the language acquisition problem. When we make choices that are not derived from empirical data, we must be prepared to explain why they are reasonable choices and what impact they have on the acquisition process. For more detailed discussion of when, why, and how to computationally model language acquisition more generally, we recommend acquisition modeling overviews by Pearl (Pearl, 2010; Pearl & Goldwater, in press), Alishahi (Alishahi, 2010; Freudenthal & Alishahi, 2014), and Räsänen (2012), among others.

Still, if all these learning task components are defined explicitly (which may be easier to do for some acquisition tasks than others), and the model succeeds at reaching the defined target state, we can then explore the internal representations constructed by the modeled child to understand exactly *why* it succeeds. Similarly, if a modeled learning strategy fails, we can inspect the internals of the modeled child to understand exactly *why* it failed. This means that computational modeling has the additional benefit of providing information that may not be accessible in studies of actual children, where we can only observe success or failure without seeing inside the child's mind to understand why that success or failure occurred.

In the next section, we demonstrate how to leverage empirical results to define each of these learning task components for a particular acquisition task involving syntactic island constraints. We focus on syntactic island constraints because they are a complex linguistic phenomenon that has been at the center of multiple debates in the linguistics literature, and so are a good demonstration of the benefits of computational modeling. In what follows, we will attempt to illustrate the construction of a modeled learner (i) who uses the data children use (ii) in a way that children might use those data (iii) to learn what children learn from those data (iv) in the same amount of time children have to learn. When we have a learning strategy that successfully accomplishes this, we have a viable option for how the acquisition process could typically occur for that specific acquisition task. It is at that point that we can examine the components of the successful strategy to help us understand why the strategy works as it does, and when it might fail.

2   Modeling syntactic island acquisition

2.1   A brief introduction to syntactic island effects

One of the most interesting aspects of human language structure is that there can be relationships between linguistic elements that are not next to each other (sometimes referred to as *long-distance dependencies*). English *wh*-questions are an excellent example of this. For instance, in (1a), we understand the *wh*-word *what* as the thing that Jack is thinking, even though it doesn't appear after the verb *think*, where the object of *think* would typically appear. We indicate the position where *what* is understood with an underscore, which is often called the *gap* position. So, (1a) shows that there is a long-distance dependency between *what* and the gap where *what* is understood. One defining characteristic of long-distance *wh*-dependencies is that they appear to be unconstrained by length (Chomsky, 1965; Ross, 1967), as shown in (1b)–(1d). The distance between the *wh*-word and where it's understood can be increased by any number of words or clauses.[1]

---

[1]Though there's clearly an upper bound on the number of words and/or clauses that an English speaker can keep track of during language processing, this restriction appears to be based on the limited nature of human working

(1)   a.   What does Jack think ___*what*?
       b.   What does Jack think that Lily said ___*what*?
       c.   What does Jack think that Lily said that Sarah heard ___*what*?
       d.   What does Jack think that Lily said that Sarah heard that the goblins stole ___*what*?

Still, while *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 such as those in square brackets in (2), the resulting utterance will be unacceptable (Chomsky, 1965; Ross, 1967; Chomsky, 1973; Huang, 1982, among many others).[2] These structures are known as *syntactic islands* due to a metaphor where *wh*-questions are formed by moving the *wh*-word from the position where it's understood to the front of the utterance, and these structures prevent that movement because they are islands that cannot be moved out of (Ross, 1967). So, the explanation for the unacceptability of these utterances is that the long-distance dependencies cross a syntactic island, which is not allowed.

(2)   Some examples of island-crossing dependencies, with island structures in brackets
       a.   * What did you make [the claim that Jack bought ___*what*]?
       b.   * What do you think [the joke about ___*what*] was hilarious?
       c.   * What do you wonder [whether Jack bought ___*what*]?
       d.   * What do you worry [if Jack buys ___*what*]?

Island effects are particularly interesting from the perspective of language acquisition for two reasons. First, there is cross-linguistic variability in syntactic island effects (e.g., Engdahl, 1980; Huang, 1982; Rizzi, 1982; Lasnik & Saito, 1984; Torrego, 1984; Hagstrom, 1998) which means that children must learn what the syntactic islands are for their language. This is equivalent to learning which long-distance dependencies are acceptable in their language. We present a model of a learning strategy for doing this, empirically evaluating its performance on English syntactic islands. Second, the differences between sentences containing allowed dependencies and sentences containing island-crossing dependencies are, in many cases, difficult to characterize without making reference to relatively abstract syntactic structure. This has led many syntacticians to postulate relatively complex, abstract constraints to capture island effects in the adult grammar (e.g., Chomsky (1986)), which has in turn led many acquisition researchers to postulate domain-specific, innate knowledge to explain the acquisition of island effects by children. This makes island effects an ideal case study for exploring the power of computational modeling with respect to (i) complex learning problems, and (ii) long-standing debates in cognitive science (i.e., debates about the role of domain-specific, innate knowledge in language acquisition).

2.2   Defining the syntactic islands acquisition task

First, we must precisely define the specific acquisition task, using the components from section 1.2: the initial state, encoding, data intake, learning period, inference, iteration, and target state.

memory capacity rather than an explicit structural restriction on the length of *wh*-dependencies in English.
    [2]Unacceptable utterances are traditionally indicated with a * in front of them.

2.2.1 The initial state

Since the initial state is comprised of the knowledge, abilities, and biases the learner already has, this is where the learning strategy plays a large role. In particular, to implement a particular learning strategy, the learner needs to draw on her existing knowledge, abilities, and biases. So, to know what is required in the learner's initial state, we first need to describe the learning strategy itself in enough detail to know what it depends on for implementation.

The learning strategy developed by Pearl and Sprouse (2013b) involves the learner using certain pieces of phrase structure to define a *wh*-dependency, in particular the phrase structure nodes that contain the dependency (which they call *container nodes*). A phrase structure node contains the dependency if the path from the gap to the *wh*-word must pass through the phrase structure node. To see this demonstrated, consider the utterance *Who did Jack think that the story about penguins amused?* and the phrase structure tree in Figure 1. Starting at the gap, the path must move up through the embedded VP, the embedded IP, the CP, the main VP, and the main IP before it finally reaches a phrase structure node that has the *wh*-word *who* as its child. These are the container nodes for this dependency (also shown in the bracket structure in (3a)).
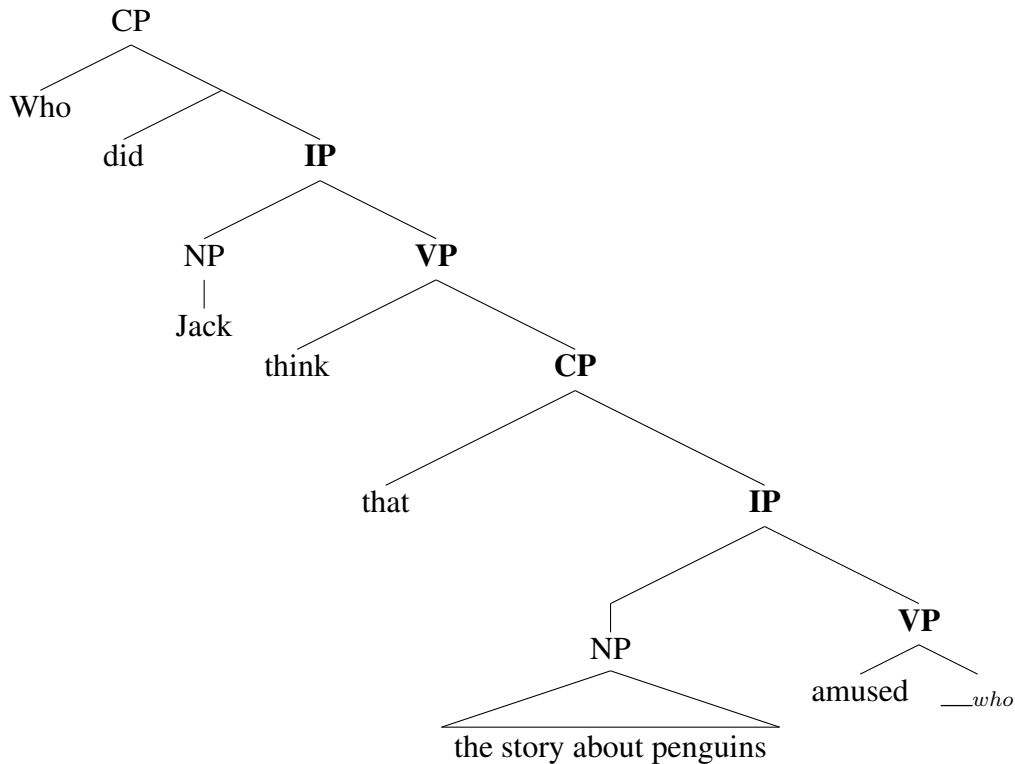

Figure 1: Abbreviated phrase structure tree with phrase structure node labels in **bold** that are relevant for characterizing the *wh*-dependency in *Who did Jack think that the story about penguins amused?*

The *wh*-dependency can be characterized as the sequence of container nodes from the *wh*-word to the gap. We also include the labels *start* and *end* in the sequence to indicate the beginning and

ending of the dependency, as shown in (3b). One important component of the characterization is how specific the container nodes are. They could just be the basic container nodes themselves, such as CP and VP. However, they could also be described by what word is the head of the phrase for the container node, such as the complementizer *that* for the CP or the verb *think* for the main clause VP – that is, they could be subcategorized, and so represented as CP_{that} or VP_{think}. Pearl and Sprouse (2013b) investigated a version of dependency characterization which used basic nodes except for CP, which was subcategorized by the complementizer (3b).

The learner then extracts a certain kind of information from every dependency in the input by breaking the container node sequence characterizing the dependency into a sequence of trigrams (3c). It is the frequency of these smaller pieces that the learner tracks. So, after encountering the utterance *Who did Jack think that the story about penguins amused?*, the learner has observed one instance of each of the five container node trigrams that comprise the *wh*-dependency. After encountering many *wh*-dependencies, the learner has a collection of frequencies for each of the container node trigrams observed. Importantly, that is *all* the learner needs to learn.[3]

(3)   Who did Jack think that the story about penguins amused __$_{who}$?

    a.   Phrase structure nodes containing the *wh*-dependency:
        Who did [$_{IP}$ Jack [$_{VP}$ think [$_{CP}$ that [$_{IP}$ the story about penguins [$_{VP}$ amused __$_{who}$]]]]]?

    b.   Container node characterization of *wh*-dependency with CP subcategorization:
        start-IP-VP-CP$_{that}$-IP-VP-*end*

    c.   Trigrams of container nodes $\in Trigrams_{start-IP-VP-CP_{that}-IP-VP-end}$:

        = *start*-IP-VP
              IP-VP-CP$_{that}$
                 VP-CP$_{that}$-IP
                    CP$_{that}$-IP-VP
                         IP-VP-*end*

How then does a learner come to have judgments about a dependency? Whether the dependency has been encountered before or not, the learner follows similar steps to generate a judgment:

1. Characterize the dependency as a sequence of container nodes.

2. Identify the container node trigrams comprising the dependency.

3. Combine the frequency information from each container node trigram in the dependency to generate a probability for that dependency.

The dependency probability can then be translated into a judgment: More probable dependencies are judged more acceptable while less probable dependencies are judged less acceptable.

---

[3]We note that the empirical support for young children's ability to track frequencies in the input comes from the extensive experimental literature on children's implicit statistical learning abilities, including Saffran, Aslin, and Newport (1996), Gomez and Gerken (1999), and Mintz (2006), among others.

The particular way Pearl and Sprouse (2013b) implemented the third step of combining the frequency information from the trigrams was to use the product of the individual trigram probabilities, as shown in (4). A trigram's probability is calculated by dividing the trigram frequency observed from the learner's input by the total number of trigrams observed in the learner's input.[4]

(4)　　p(Who did Jack think that the story about penguins amused __$_{who}$?)

$$= \prod_{trigram \in Trigrams_{start-IP-VP-CP_{that}-IP-VP-end}} p(trigram)$$

$$= p(\textit{start}\text{-IP-VP}) * p(\text{IP-VP-CP}_{that}) * p(\text{VP-CP}_{that}\text{-IP}) * p(\text{CP}_{that}\text{-IP-VP}) * p(\text{IP-VP-}\textit{end})$$

$$= \frac{\#\ start-IP-VP}{\#\ total\ trigrams} * \frac{\#\ IP-VP-CP_{that}}{\#\ total\ trigrams} * \frac{\#\ VP-CP_{that}-IP}{\#\ total\ trigrams} * \frac{\#\ CP_{that}-IP-VP}{\#\ total\ trigrams} * \frac{\#\ IP-VP-end}{\#\ total\ trigrams}$$

This strategy relies on the learner being able to do a number of things. First, to identify the container nodes that describe a dependency, the learner must be able to recognize the appropriate phrase structure in an input utterance. Second, to characterize a *wh*-dependency as a sequence of container nodes, with the appropriate level of specificity (e.g., CP vs. CP$_{that}$), the learner must recognize the container nodes and be able to use them. Third, the learner must know to break the container node sequence that characterizes a dependency into trigrams, and track the frequency of these container node trigrams in the input. Fourth, the learner must know to combine the container node trigram information to generate a probability for the dependency and be able to do so successfully. These comprise the initial state of the learner, shown in (5).

(5)　　Knowledge, abilities, and biases in the learner's initial state
　　　a.　Phrase structure
　　　　(i)　Knowledge: Know phrase structure.
　　　　(ii)　Ability: Successfully recognize phrase structure of utterances.
　　　b.　Container nodes
　　　　(i)　Bias: Characterize *wh*-dependencies using container nodes.
　　　　(ii)　Ability: Successfully recognize container nodes.
　　　　(iii)　Ability: Successfully characterize dependencies using container nodes.
　　　c.　Container node trigrams
　　　　(i)　Bias: Break container node sequences into trigrams.
　　　　(ii)　Ability: Successfully break container node sequences into trigrams.
　　　　(iii)　Bias: Track frequency of container node trigrams.
　　　　(iv)　Ability: Successfully track frequency of container node trigrams.
　　　d.　Generating *wh*-dependency judgments
　　　　(i)　Bias: Use container node trigram probabilities to generate judgments.
　　　　(ii)　Knowledge: How to combine trigram probabilities the right way.
　　　　(iii)　Ability: Successfully combine trigram probabilities.

---

[4]There is a small additional step in calculating a trigram's probability called *smoothing*, which ensures that no trigram actually has zero probability, even if it has never been observed in the input. In essence, the smoothing step allows trigrams that have never before been observed to have a very small non-zero probability anyway, just in case they're acceptable but simply haven't shown up in the learner's input yet. See Pearl and Sprouse (2013b) for details of the smoothing calculation.

2.2.2   Data encoding and data intake

The data the learner actually uses are some subset of the data available in the child's input, and this subset is defined by both the learner's developing cognitive abilities and the biases in the learner's initial state. First, however, it's useful to identify what children's input actually looks like.

The CHILDES database (MacWhinney, 2000) is an excellent resource for the input children typically encounter, as it provides transcripts and recordings of naturalistic interactions between children and their caretakers (among other types of data). The data that come from American English are most relevant for the English syntactic island effects Pearl and Sprouse (2013b) investigated, and so they examined the transcripts of American English speech directed at children of a variety of ages. This served as a reasonable sample of the linguistic data available to children. In particular, Pearl and Sprouse (2013b) used a sample of 101,838 child-directed utterances aggregated from several commonly used American English corpora in the CHILDES database: the Adam and Eve corpora by Brown (1973), the Valian (1991) corpus, and the Suppes (1974) corpus. Collectively, these utterances were directed at 24 children between the ages of 1;6 and 5;2.

So, this is likely a representative enough sample of the linguistic input for learning about syntactic islands. Now, how does the modeled learner encode these data? Put simply, just because the information is available doesn't mean the child will be able to encode all that information, or encode what's there correctly, due to immature or atypically developing cognitive abilities (Lidz & Gagliardi, 2015; Omaki & Lidz, 2014). Pearl and Sprouse (2013b) assumed that the modeled learner had sufficient linguistic knowledge and cognitive resources to recognize the syntactic structure of the utterances encountered (i.e., the learner can construct the the phrase structure for an utterance). They also assumed that the modeled learner had sufficient working memory to be able to identify the gap associated with any dependency.

Now, what's the modeled learner's intake? The data intake for learning about syntactic islands, based on the learning strategy implemented in the learner's initial state, is the part of the data deemed relevant for syntactic islands. So, the modeled learner focuses on the structure of the *wh*-dependencies in the child-directed speech utterances, since the container node trigrams are what's being tracked, and those characterize *wh*-dependencies (5b). That is, each *wh*-dependency is perceived by the learner as the sequence of container nodes that characterize it.

Pearl and Sprouse (2013b) discovered that approximately 20% of their linguistic sample contained *wh*-dependencies (20,923 of 101,838). They derived the container node characterization of these dependencies from structurally parsed annotations of the *wh*-dependencies, available as part of the CHILDES Treebank.[5]

Notably, it turned out that most of the *wh*-dependencies (89.5%) were of two types: those with the gap in the main clause object like *What did she see?* (container node sequence: *start*-IP-VP-*end* = 76.7% of *wh*-dependencies) and those with the gap in the main clause subject like *Who saw it?* (container node sequence: *start*-IP-*end* = 12.8% of *wh*-dependencies).[6] Still, the remaining

---

[5] *http://www.socsci.uci.edu/~lpearl/CoLaLab/CHILDESTreebank/childestreebank.html*

[6] Interestingly, when Pearl and Sprouse (2013b) investigated adult-directed conversational speech, they discovered a very similar distribution of *wh*-dependencies, consisting of mostly main-clause object (73.0%) and main-clause subject (17.2%) *wh*-dependencies. This demonstrates a rare case when child-directed speech and adult-directed speech do *not* appear to differ in their distribution of some linguistic attribute.

10.5% of the *wh*-dependencies included 24 different dependency types, so there are a variety of *wh*-dependencies in the learner's data intake.

2.2.3   The learning period

How long the learner has to learn can be translated into how much data the learner encounters before learning is complete. Hart and Risley (1995) determined that American children in their samples were exposed to approximately one million utterances between birth and three years old, so this provides an empirical basis for the modeled learner of Pearl and Sprouse (2013b). In particular, Pearl and Sprouse (2013b) assumed that intuitions about syntactic islands are acquired in a three year period, such as between the ages of two and five, based on available experimental studies (Goodluck, Foley, & Sedivy, 1992; de Villiers & Roeper, 1995; de Villiers, Roeper, Bland-Stewart, & Pearson, 2008; Roeper & de Villiers, 2011).[7] So, they assumed that the modeled learner would encounter one million utterances distributed similarly to the CHILDES-based sample of American English, which meant that approximately 20% (200,000 utterances) would contain *wh*-dependencies. Since only *wh*-dependencies are relevant for the modeled learner, this means the learning period is therefore 200,000 *wh*-dependencies, and these are the data that the modeled learner applies the learning strategy to. Importantly, these 200,000 data points are encountered with the same relative frequency as the CHILDES data sample indicated, so approximately 9 out of 10 are main-clause object or main-clause subject dependencies, while the rest are distributed across the remaining 24 *wh*-dependency types.

2.2.4   Inference and iteration

The inference process for the modeled learner defines how updates to the learner's internal representation are made, based on the data intake. For the syntactic islands learning strategy, the internal representation is simply the frequencies of the container node trigrams, since these are what the learner updates when encountering the input. These are also what the learner draws on to create judgments about new dependencies, and so represent the part of the iteration process that relates to the learner's observable behavior. The other two parts of iteration, which relate to how the updated internal representation affects subsequent encoding and intake, are not impacted, since neither encoding nor intake depend on the container node trigram frequencies.

2.2.5   The target state

Now, what is the learner supposed to learn? While we do not have experimental data indicating the trajectory of development for the four syntactic islands of interest, we do have data demonstrating adult knowledge of these syntactic islands. This provides the ultimate target state of acquisition.

---

[7]This is a simplifying assumption, since no experimental studies currently exist about when children acquire the four specific syntactic islands Pearl and Sprouse (2013b) investigated. Notably, this means Pearl & Sprouse's modeling results can be used to predict what experimental studies *will* find about the trajectory of acquisition for these specific syntactic islands.

Sprouse, Wagers, and Phillips (2012) collected formal acceptability judgments from 173 participants for the four island types in (2), using the magnitude estimation task: Complex NP islands (6d), (simple) Subject islands (7d), Whether islands (8d), and (conditional) Adjunct islands (9d). Notably, the acceptability judgments not only provide a demonstration of target knowledge, but also offer a concrete set of behaviors that a modeled learner should aim to reproduce. This allows us to match the modeled learner's output to adult linguistic output. In particular, Pearl and Sprouse (2013b) compared the *wh*-dependency judgments generated by the modeled learner against the acceptability judgments generated by American English adults.

Sprouse et al. (2012) were very careful to control the acceptability judgment collection, and used a (2x2) factorial definition of each island effect (shown in (6)-(9)) that controlled for two salient properties of utterances that crossed islands, like those in (2). First, these utterances contained an embedded-clause dependency (EMBEDDED), rather than a matrix-clause dependency (MATRIX) (in syntax, *matrix* is a technical term for main clauses). Second, these utterances contained a syntactic island structure (ISLAND), as opposed to not (NON-ISLAND). So, four different stimuli types could be created that separated the individual contributions of these two factors.

(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

The factorial definition of island effects treats the unacceptability of the island-crossing dependencies in (6d)-(9d) as a superadditive effect of the two factors (an EMBEDDED dependency and the presence of an ISLAND structure in the utterance). In particular, an island is apparent when there is additional unacceptability that arises beyond the dependency being an embedded clause

dependency and beyond the island structure being present in the utterance. The presence of a syntactic island effect (which can be thought of as implicit knowledge of syntactic islands) then becomes visually salient: If the acceptability of the four stimuli types for each island (as indicated by their z-scores) is plotted on an interaction plot, the presence of a syntactic island appears as two non-parallel lines (see Figure 2), due to the additional unacceptability that arises for the island-crossing dependency. This interaction is also statistically significant. In contrast, the lack of a syntactic island appears as two parallel lines and results in no significant statistical interaction, as the unacceptability of the utterance is completely explainable by the summed effects of it being an embedded clause dependency and having an island structure in it. Notably, Sprouse et al. (2012) found superadditivity (i.e., non-parallel lines and a statistically significant interaction) for all four islands investigated, as shown in Figure 2. This demonstrates the implicit knowledge about these four syntactic islands that adults possess.
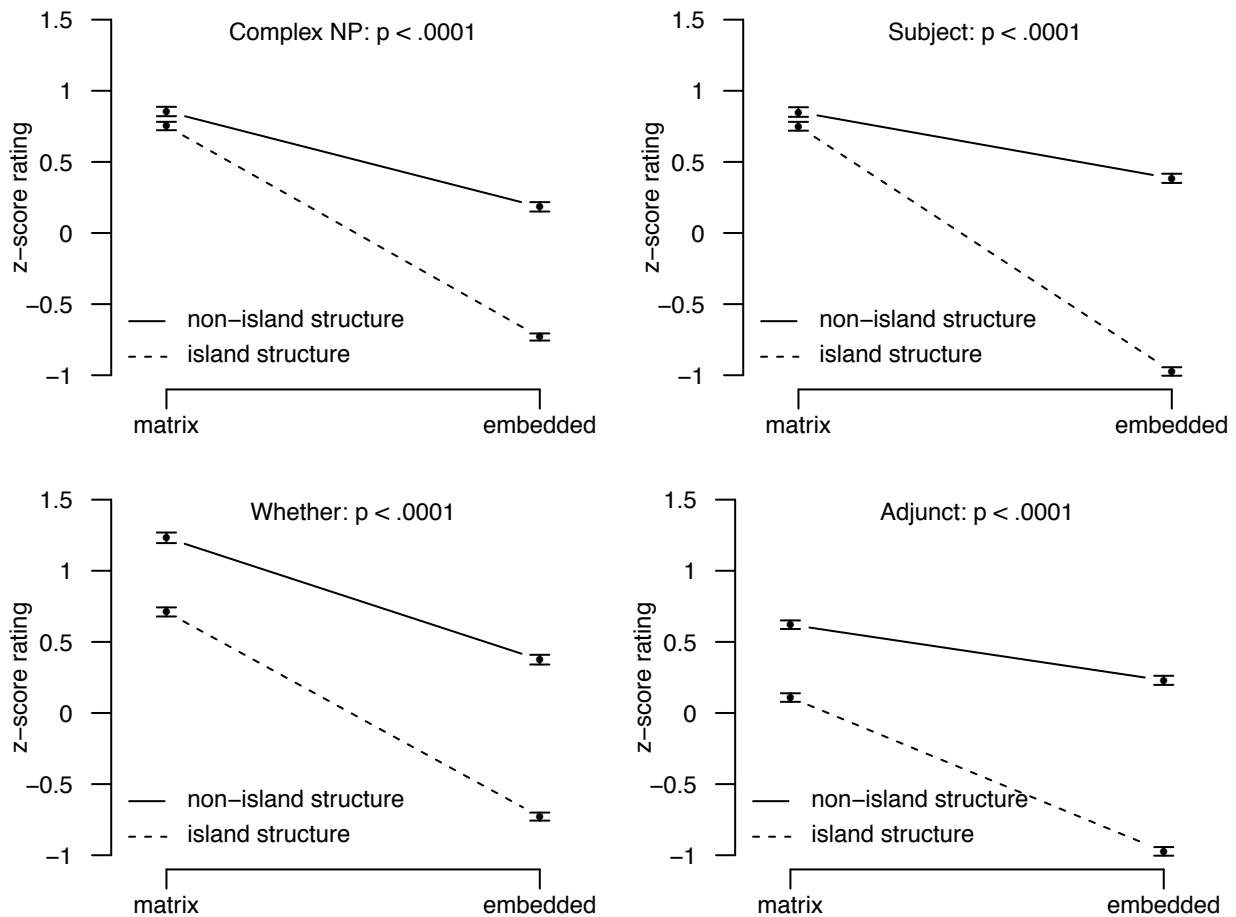

Figure 2: Experimentally derived acceptability judgments for the four island types (N=173) from Sprouse et al. (2012).

So, this is the target behavior we would like the modeled learner to exhibit as well, in order

to demonstrate that it too has implicit knowledge of these four syntactic islands. The way it can demonstrate this target behavior is by assigning probabilities to each of the stimuli for the four island types after it completes the learning period, based on the characterizations of the utterances' *wh*-dependencies. For example, for the Complex NP stimuli, the modeled learner will generate a probability for the utterances in (6a)-(6d). These probabilities can then be plotted on an interaction plot, just as the z-scores of human judgments were plotted in Figure 2. If we see superadditivity in the modeled learner's generated judgments, the modeled learner has also demonstrated implicit knowledge of these four syntactic islands.

2.3   Results

Figure 3 shows the modeled learner's generated judgments for each of the four island types after learning from child-directed speech data, with the log probability on the y-axis.

Pearl and Sprouse (2013b) used the logarithm[8] of the raw probabilities instead of the raw probabilities themselves for several reasons. First, the raw probability calculated for a *wh*-dependency is often a very small number due to multiplying many probabilities together, such as 0.0000001, and it can be visually taxing to compare very small numbers with a large number of decimal places. In contrast, the logarithm of a small number is a negative integer (e.g., log(0.000001) = -6).

Second, the integer of the log probability represents the number of decimal places in the probability. For example, $\log(0.000001) = \log(10^{-6}) = -6$, whereas $\log(0.001) = \log(10^{-3}) = -3$. This means that logarithms allow us to quickly compare the order of magnitude of the probability (e.g., a probability whose log is -6 is one thousand times (10^3) smaller than a probability whose log is -3).

Finally, addition of log probabilities is equivalent to multiplication of the raw probabilities. This is handy when considering a collection of trigram probabilities that we need to multiply together, as the modeled learner needs to do to generate a judgment about a *wh*-dependency (recall the calculation in (4)). We can simply add the log probabilities together, which is often simpler than multiplying raw probabilities with multiple decimal places.

As mentioned above, all log probabilities are negative because raw probabilities are between 0 and 1, so the log of a probability is between negative infinity and 0 (e.g., log(0.000001) = -6 while log(1) = 0). This means the numbers closer to zero are more positive and appear higher on the y axis – these represent structures judged by the modeled learner as "more acceptable". Numbers further from zero are more negative and appear lower on the y axis – these represent structures judged "less acceptable".

As we can see from the interaction plots, the modeled learner displays the qualitative target behavior indicating implicit knowledge of these four syntactic islands: for each one, we observe superadditivity for the island-spanning dependency. This indicates that the modeled learner, like the adults from Sprouse et al. (2012), has learned that there is additional unacceptability that arises in these utterances beyond simply having an embedded dependency and having an island structure

---

[8]The base ten logarithm was used by Pearl and Sprouse (2013b), but the natural logarithm would also accomplish the same goal of scaling the probabilities appropriately. The important aspect is the transformational scaling property of the logarithm, rather than the particular base of the logarithm.
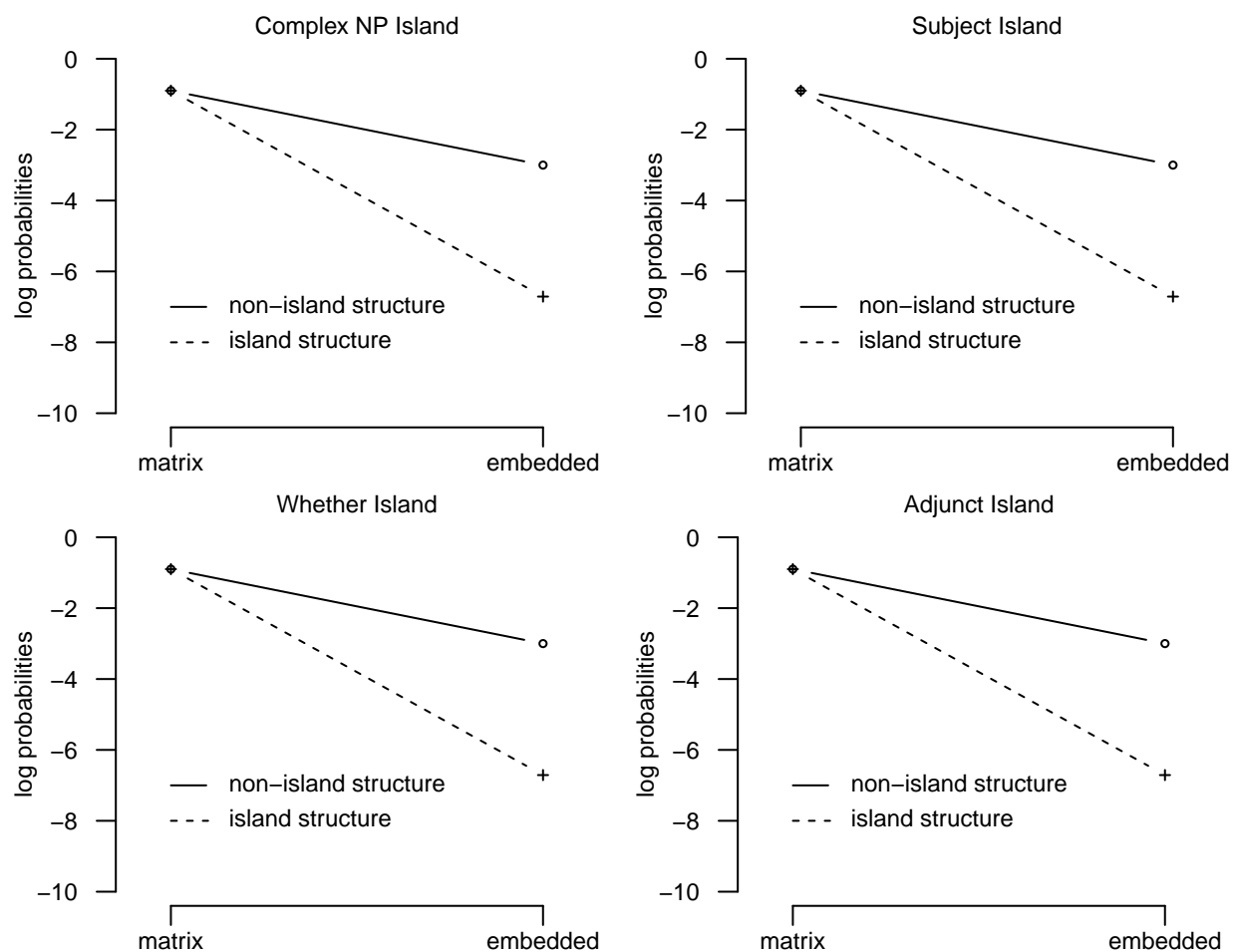
Figure 3: Modeled learner results, after learning from child-directed speech data.

present in the utterance.

Still, there are noticeable differences between the observed acceptability judgments from adults and the grammaticality preferences inferred by the model learner, despite the qualitative similarity. For example, adults assign some unacceptability to having an island structure present in the utterance (i.e., the lefthand data point for the island structure line in each of the plots in Figure 2 is lower than the lefthand data point for the non-island structure line). In contrast, the modeled learner does not assign some unacceptability to the presence of an island structure (all the lefthand data points for the island structure lines overlap the lefthand data points for the non-island structure lines in Figure 3). This is because actual acceptability judgments are based on dozens of factors that are not included in this model, such as lexical item choice, semantic probability, and processing difficulty (Schütze, 1996; Cowart, 1997; Keller, 2000; Sprouse, 2009). The grammaticality inferred by this modeled learner would constitute only one factor among many that affect acceptability, though it is likely a very large factor since the qualitative pattern emerges so distinctly. So, this is why we only look for qualitative similarity between the modeled results and human judgment data, rather than

looking for an exact quantitative fit. Since the qualitative behavioral pattern is in fact there, we interpret this to mean that the target knowledge adults possess about these four syntactic islands is learnable from realistic child-directed speech data, and the modeled learning strategy demonstrates one way to accomplish this.

3   Interpreting the model results

As mentioned in the introduction, one of the most useful aspects of computational modeling is its explanatory power: when we find a learning strategy that works, we can look to see precisely *why* it works. Here, Pearl and Sprouse (2013b) discovered a learning strategy for syntactic islands that involves characterizing *wh*-dependencies using sequences of container nodes. Now, why does it work?

Crucially, for each of the island-spanning dependencies, there is at least one extremely low probability container node trigram in the container node sequence characterizing the dependency (10). These trigrams are assigned low probabilities because these trigrams are never observed in the input.[9] Note that some trigrams are low probability due to being rarely encountered in the child-directed input (e.g., CP_{that}-IP-VP) – but, importantly, this is still more than never. For example, even though CP_{that}-IP-VP rarely appears, it *does* appear, and so it is assigned a probability that is substantially non-zero. This contrasts with each of the trigrams listed in (10).

(10)   Container node trigram sequences for island-spanning *wh*-dependencies

    a.   Complex NP

       (i)   * What did [$_{IP}$ the teacher [$_{VP}$ make [$_{NP}$ the claim $_{CP_{that}}$ that [$_{IP}$ Lily $_{VP}$ forgot __]]]]]?

       (ii)   *start*-IP-VP-NP-CP_{that}-IP-VP-*end*

       (iii)   Low probability:

             VP-NP-CP_{that}

             NP-CP_{that}-IP

    b.   Subject

       (i)   * Who does [$_{IP}$ Jack [$_{VP}$ think [$_{CP_{null}}$ [$_{IP}$ [$_{NP}$ the necklace [$_{PP}$ for __]] is expensive]]]]?

       (ii)   *start*-IP-VP-CP_{null}-IP-NP-PP-*end*

       (iii)   Low probability:

             CP_{null}-IP-NP

             IP-NP-PP

    c.   Whether

       (i)   * What does [$_{IP}$ the teacher [$_{VP}$ wonder [$_{CP_{whether}}$ whether [$_{IP}$ Jack [$_{VP}$ stole __]]]]]?

       (ii)   *start*-IP-VP-$CP_{whether}$-IP-VP-*end*

       (iii)   Low probability:

             IP-VP-$CP_{whether}$

             VP-$CP_{whether}$-IP

             $CP_{whether}$-IP-VP

---

[9]It is only the smoothing step in the calculation process that prevents these probabilities from being 0.

d. Adjunct
   (i)   * What does [$_{IP}$ the teacher [$_{VP}$ worry [$_{CP_{if}}$ if [$_{IP}$ Lily [$_{VP}$ forgot __]]]]]?
   (ii) *start*-IP-VP-CP$_{if}$-IP-VP-*end*
   (iii) Low probability:
            IP-VP-CP$_{if}$
            VP-CP$_{if}$-IP
            CP$_{if}$-IP-VP

So, implicit knowledge of these syntactic islands can be translated to implicit recognition that the *wh*-dependencies in question contain several container node trigrams that never occur in the input, and this is what learners leverage to make their judgments. On the representational side, this helps us understand how knowledge of constraints on *wh*-dependencies may be represented in the mind. In particular, rather than needing to know about specific island structural constraints, humans could instead be sensitive to local pieces of structure captured by container node trigrams.[10]

On the learning side, this helps us understand exactly what is required for humans to leverage these local pieces of structure and end up with the grammaticality preferences we see demonstrated behaviorally by adults for syntactic islands. If this is indeed the way the children learn about syntactic islands, then we know what this learning strategy requires to succeed: the knowledge, biases, and abilities listed in (5). These fall into four major areas: (i) parsing utterances into phrase structure trees, (ii) characterizing dependencies as container node sequences, (iii) tracking the frequency of container node trigrams in the input, and (iv) calculating the probability of a container node sequence for a given *wh*-dependency based on its container node trigrams.

3.1 Consequences for the debate between linguistic nativists and constructionists

Once a successful learning strategy has been identified, one question that we can ask is how the components of that learning strategy may come to exist in the child. Pearl and Sprouse (2013) suggest that there are (at least) two dimensions along which components of a learning strategy might vary: components can either be (i) innately specified (*innate*) or derived from other components (*derived*), and (ii) generally available across cognitive domains (*domain-general*) or available to one specific cognitive domain alone (*domain-specific*). This means that there are four types of learning components (11).

(11) a. innate, domain-general
      b. derived, domain-general
      c. derived, domain-specific
      d. innate, domain-specific

---

[10]In fact, this is intuitively what the *Subjacency Condition* captured (Chomsky, 1973; Huang, 1982; Lasnik & Saito, 1984; Chomsky, 1986, among many others), though its exact instantiation of local structure differed from the container node trigrams investigated by Pearl and Sprouse (2013b).

All learning theories require innate, domain-general (type a) components in order for any learning at all to exist. Most learning theories also allow for the possibility of derived components regardless of whether they are domain-general or domain-specific (types b and c), as the derivation of more complex knowledge is the very definition of learning. However, the possibility of innate, domain-specific components (type d) in language acquisition has led to a longstanding debate in the field of cognitive science between linguistic nativists, who allow for the possibility of type d components, and constructionists, who only allow for the possibility of types a through c. Crucially, computational models like the one presented here provide an explicit list of components that can be evaluated relative to this typology in order to fruitfully contribute to this debate.

Pearl and Sprouse (2013) discuss each of the components of the syntactic islands learning strategy in detail, so we will only provide a brief discussion of each of the four major components here. The first component is the learner's ability to parse utterances into phrase structure trees. This ability is not specific to the learning of syntactic island effects, as it underlies all of syntactic knowledge. So, the question of whether it can be accomplished without innate, domain-specific components is an area of active research in the field of acquisition.

The second component is the ability to characterize dependencies as container node sequences. At first glance, this component appears to be a straightforward consequence of the ability to parse the input into phrase structure trees: once the tree is available, the sequence of container nodes is also (in principle) available. This is because the search for the gap position when processing an utterance is an active process (Crain & Fodor, 1985; Stowe, 1986; Frazier & Flores D'Arcais, 1989) that tracks the container nodes of the gap location (see C. Phillips (2006) for a review). However, this does not explain why the learner chooses to attend to the sequence of container nodes exclusively as opposed to all of the other information that is available in a phrase structure tree, such as the number of nouns that intervene between the *wh*-word and the gap, the number of verbs, or the number of functional categories, and so on. In addition, the modeled strategy requires the learner to characterize *wh*-dependencies at a particular level of specificity – namely what might be considered "basic" phrasal nodes (e.g., IP, VP, NP, and so on) with the exception of CP, which is subcategorized by the lexical item in complementizer position (e.g., CP_{null}, CP_{that}, etc.). This subcategorization is very important, since the learning strategy fails to work for two of the four syntactic islands otherwise (see Pearl and Sprouse (2013b) for details). The fact that the learner must attend to this one piece of information over all others, and this specific instantiation of it besides, is a potential mystery that requires additional research. It may be that this can be explained without innate, domain-specific learning biases, but it may also turn out that the bias to attend to these container node sequences is innate and domain-specific.

The third component is the ability to track the frequency of container node trigrams. Once again, this component can be divided into two parts: the ability to track frequencies, which is likely innate and domain-general (it is assumed by nearly all learning theories), and the bias to attend specifically to trigrams (rather than any other n-grams), which is currently a mystery. Additional research is necessary to determine how it is that the learner settles on trigrams as the appropriate unit size for learning syntactic islands, though there is evidence that young children rely on sequences of three units for other parts of language acquisition, such as word segmentation and grammatical categorization (see discussion below in section 3.2).

The final component is the ability to calculate the probability of a given container node sequence given the trigram frequencies. The ability to calculate probabilities is likely innate and domain-general, as it is part of nearly all learning theories.

3.2 Consequences for studies of atypical development

Another question we can ask is how the successful learning strategy might break down during atypical development. The components of the learning theory again provide a roadmap for potential investigation.

For phrase structure, can the child recognize the appropriate phrase structure for a given utterance, whether it has a dependency in it or not? This is one of the most basic components of the modeled learning strategy, as phrase structure is an underlying representation required for syntactic knowledge. Being able to successfully parse an utterance requires the child to already recognize appropriate syntactic category information and construct the appropriate phrase structures based on those syntactic categories.

For container nodes, does the child recognize the appropriate container nodes and use them to characterize *wh*-dependencies? We know that container nodes are actively tracked during typical processing of dependencies. So, when we observe atypical development, we can investigate whether the child is able to actively track container nodes for dependencies. We also know that the modeled strategy requires a specific instantiation of container nodes in order to succeed. So, another important component is whether the child is characterizing the dependencies using the right kind of container nodes. If she characterizes them using basic nodes only, we expect to see failure for two specific island types (see Pearl and Sprouse (2013b) for discussion); if she characterizes them using all subcategorized container nodes (e.g., VP_{think}, PP_{with}, and so on), we may find no learning of syntactic islands at all due to data sparseness.[11]

If phrase structure and container node characterization are working as they ought to, is the child tracking container node trigrams appropriately? First, we may want to assess if she's tracking them for *wh*-dependencies only, and excluding other kinds of dependencies. This is quite important, as it turns out that syntactic islands do not exist in the same way for dependencies between pronouns and their antecedents. For example, the island-spanning dependency in the whether island * *Who did Lily wonder whether Jack likes __?* is perfectly acceptable when it is a dependency between a pronoun and its antecedent: *Lily wonders whether Jack likes her$_{Lily}$.* In particular, container node trigrams that indicate a whether island (e.g., VP-CP$_{whether}$-IP) because they are unobserved in the learner's *wh*-dependency input are actually observed in the pronoun dependency input. So, it is imperative that the child track container node trigrams for these dependency types separately.

Relatedly, is the child able to break sequences into trigrams appropriately and track these in the input? There is significant experimental evidence that this ability to attend to sequences of three units is consistent with children's observable behavior for a variety of acquisition tasks,

---

[11]For example, instances of IP$_{present}$-VP$_{see}$-PP$_{with}$ and IP$_{present}$-VP$_{see}$-PP$_{on}$ would not be recognized as the same type of local structure (IP-VP-PP), and so would be counted individually in the input. Due to the large number of content words in the language (e.g., verbs and prepositions), this leads to many isolated observations of subcategorized trigrams that can't be aggregated into the more abstract representation (IP-VP-PP) used in the modeled learning strategy. This then disrupts successful learning.

e.g., the comparison of three sequential transitional probabilities for word segmentation (Saffran et al., 1996; Aslin, Saffran, & Newport, 1998; Estes, Evans, Alibali, & Saffran, 2007; Pelucchi, Hay, & Saffran, 2009, among others) and frequent frames consisting of three sequential units for grammatical categorization (Mintz, 2006; Wang & Mintz, 2008). So, we might assess an atypical learner's ability to track trigrams for a variety of linguistic stimuli.

If it turns out that trigrams are being tracked correctly, can the child then combine container node trigrams appropriately to generate a grammaticality judgment? The modeled learning strategy requires the child to do this for *wh*-dependency judgements, but we might also look for the general ability to do this for any task that requires combining probabilities, whether linguistic or non-linguistic. So, if this ability is impaired, we should see repercussions in many developmental areas, and not just in language.

In this way, the precise instantiation of the modeled strategy can provide guidance on what to look for in the learning process when we see atypical development occurring. Additionally, a model can provide specific predictions about how different types of input affect a learner using this strategy. For example, Pearl and Sprouse (2013b) investigated what happens when a learner using this strategy is given the amount of data we think typically developing monolingual American English children encounter with the distribution of data estimated from the CHILDES database samples, which were primarily directed at children from high socio-economic status (SES) backgrounds. But we could just as easily investigate what kind of learning would occur if fewer data were encountered or if the relevant data had a different distribution, as may happen for children from different backgrounds.

For children of different SES backgrounds, it may well be that they encounter fewer relevant *wh*-dependency data than children of higher SES backgrounds, since they encounter fewer utterances overall (Hart & Risley, 1995), and it is certainly possible that the distribution of *wh*-dependencies they encounter is different. Similarly, for children from a bilingual background, the amount of language data encountered in one of the two languages is by necessity less than the amount of that language's data encountered by a monolingual child, since the bilingual child's input is divided between two languages while a monolingual child encounters all the input in one language. Also, a bilingual child's input in one language may have a different distribution of *wh*-dependencies. If we have empirical estimates of the input data encountered by children of different SES and linguistic backgrounds, it is straightforward to apply the model to those data and see how they impact learning of syntactic islands. More generally, the impact of different input types is easy to see when we have a precisely implemented learning model that specifies what the child's encoded intake is and how the child uses it.

4   Conclusion

Computational modeling is a powerful tool that we can use to understand linguistic development. When used correctly, computational modeling is beneficial to both steps of the scientific process. Theory specification benefits because computational modeling forces theorists to be explicit about both the components of the learning model and their interaction. Theory evaluation benefits because computational modeling provides an explicit demonstration of how a learning strategy solves

a particular learning task given realistic language input. Explicit computational models allow us to first see that the learning strategy works, and then see how it works step by step, what makes it work, and all the individual pieces that go into making it work. The results of computational models can be relevant both for debates in basic cognitive science, such as the longstanding debate between linguistic nativists and constructionists, and for practical applications, such as studies and potential interventions in atypical development. We have provided a concrete example of the computational modeling process applied to the task of learning about syntactic islands, and hope this has highlighted both the considerations necessary for modeling and the empirical data needed to adequately implement an informative model. This general computational modeling approach can be used for any acquisition task, provided we have the theoretical, corpus, and experimental data to precisely define both the acquisition task and the learning strategy.

References

Alishahi, A. (2010). Computational modeling of human language acquisition. *Synthesis Lectures on Human Language Technologies*, *3*(1), 1–107.

Alishahi, A., & Stevenson, S. (2008). A computational model of early argument structure acquisition. *Cognitive science*, *32*(5), 789–834.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321–324.

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

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge: The MIT Press.

Chomsky, N. (1973). Conditions on transformations. In S. Anderson & P. Kiparsky (Eds.), *A Festschrift for Morris Halle* (pp. 237–286). New York: Holt, Rinehart, and Winston.

Chomsky, N. (1986). *Barriers* (Vol. 13). MIT press.

Cowart, W. (1997). *Experimental syntax: Applying objective methods to sentence judgments*. Sage Publications.

Crain, S., & Fodor, J. D. (1985). How can grammars help parsers? In D. Dowty, L. Kartunnen, & A. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives* (pp. 94–128). Cambridge, UK: Cambridge University Press.

de Villiers, J., & Roeper, T. (1995). Relative clauses are barriers to wh-movement for young children. *Journal of Child Language*, *22*(2), 389–404.

de Villiers, J., Roeper, T., Bland-Stewart, L., & Pearson, B. (2008). Answering hard questions: Wh-movement across dialects and disorder. *Applied Psycholinguistics*, *29*(1), 67–103.

Engdahl, E. (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* (pp. 89–108). Ottawa, ONT: University of Ottawa Department of Linguistics.

Estes, K. G., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, *18*(3), 254–260.

Frazier, L., & Flores D'Arcais, G. B. (1989). Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language*, *28*(3), 331–344.

Freudenthal, D., & Alishahi, A. (2014). Computational models of language development. In P. J. Brooks & V. Kempe (Eds.), *Encyclopedia of language development.* SAGE Publications.

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*(2), 109–135.

Goodluck, H., Foley, M., & Sedivy, J. (1992). Adjunct islands and acquisition. In H. Goodluck (Ed.), *Island constraints* (pp. 181–194). Dordrecht: Kluwer.

Hagstrom, P. A. (1998). *Decomposing questions.* Unpublished doctoral dissertation, MIT, Cambridge, MA.

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experience of young american children.* Baltimore, MD: P.H. Brookes.

Huang, C.-T. J. (1982). *Logical relations in Chinese and the theory of grammar.* Unpublished doctoral dissertation, MIT, Cambridge, MA.

Keller, F. (2000). *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality.* Unpublished doctoral dissertation, University of Edinburgh, Edinburgh, UK.

Kol, S., Nir, B., & Wintner, S. (2014). Computational evaluation of the Traceback Method. *Journal of Child Language*, *41*(1), 176–199.

Lasnik, H., & Saito, M. (1984). On the nature of proper government. *Linguistic Inquiry*, *15*, 235–289.

Lidz, J., & Gagliardi, A. (2015). How Nature Meets Nurture: Universal Grammar and Statistical Learning. *Annual Review of Linguistics*, *1*(1), 333–352.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk.* Mahwah, NJ: Lawrence Erlbaum Associates.

Matusevych, Y., Alishahi, A., & Vogt, P. (2013). Automatic generation of naturalistic child–adult interaction data. In *Proceedings of the 35th annual meeting of the cognitive science society. austin, tx: Cognitive science society* (pp. 2996–3001).

Mintz, T. (2006). Finding the verbs: Distributional cues to categories available to young learners. In K. Hirsh-Pasek & R. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 31–63). Oxford: Oxford University Press.

Omaki, A., & Lidz, J. (2014). Linking Parser Development to Acquisition of Syntactic Knowledge. *Language Acquisition*, *xx*, xxx–xxx. doi: 10.1080/10489223.2014.943903

Pearl, L. (2010). Using computational modeling in language acquisition research. In E. Blon & S. Unsworth (Eds.), *Experimental Methods in Language Acquisition Research* (pp. 163–184). Amsterdam, The Netherlands: John Benjamins.

Pearl, L. (2014). Evaluating learning strategy components: Being fair. *Language*, *90*(3), e107–e114.

Pearl, L., & Goldwater, S. (in press). Statistical Learning, Inductive Bias, and Bayesian Inference in Language Acquisition. In J. Lidz, W. Snyder, & J. Pater (Eds.), *The Oxford Handbook of Developmental Linguistics.* (pp. xxx–xxx). Oxford, UK: Oxford University Press.

Pearl, L., & Mis, B. (2011). How Far Can Indirect Evidence Take Us? Anaphoric One Revisited. In L. Carlson, C. Höschler, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 879–884). Austin, TX: Cognitive Science Society.

Pearl, L., & Mis, B. (in press). The role of indirect positive evidence in syntactic acquisition: A

look at anaphoric one. *Language*, *xx*, xxx-xxx.

Pearl, L., & Sprouse, J. (2013a). Computational Models of Acquisition for Islands. In J. Sprouse & N. Hornstein (Eds.), *Experimental Syntax and Islands Effects* (pp. 109–131). Cambridge: Cambridge University Press.

Pearl, L., & Sprouse, J. (2013b). Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem. *Language Acquisition*, *20*, 19–64.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, *113*(2), 244–247.

Phillips, C. (2006). The real-time status of island phenomena. *Language*, *82*, 795–823.

Phillips, L., & Pearl, L. (2014a). Bayesian inference as a cross-linguistic word segmentation strategy: Always learning useful things. In *Proceedings of the computational and cognitive models of language acquisition and language processing workshop.* Gothenberg, Sweden: EACL.

Phillips, L., & Pearl, L. (2014b). Bayesian inference as a viable cross-linguistic word segmentation strategy: Its all about whats useful. In *Proceedings of the 36th annual conference of the cognitive science society* (p. 2775-2780). Quebec City, CA: Cognitive Science Society.

Phillips, L., & Pearl, L. (in press). The utility of cognitive plausibility in language acquisition modeling: Evidence from word segmentation. *Cognitive Science*.

Räsänen, O. (2012). Computational modeling of phonetic and lexical learning in early language acquisition: existing models and future directions. *Speech Communication*, *54*(9), 975–997.

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

Roeper, T., & de Villiers, J. (2011). The Acquisition Path for Wh-questions. In J. de Villiers & T. Roeper (Eds.), *Handbook of Generative Approaches to Language Acquisition* (pp. 189–246). New York: Springer.

Ross, J. (1967). *Constraints on variables in syntax.* Unpublished doctoral dissertation, MIT, Cambridge, MA.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

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

Sprouse, J. (2009). Revisiting satiation: Evidence for an equalization response strategy. *Linguistic Inquiry*, *40*(2), 329–341.

Sprouse, J., Wagers, M., & Phillips, C. (2012). A test of the relation between working memory capacity and syntactic island effects. *Language*, *88*(1), 82–124.

Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, *1*(3), 227–245.

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

Torrego, E. (1984). On inversion in Spanish and some of its effects. *Linguistic Inquiry*, *15*, 103–129.

Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition*, *40*(1), 21–81.

Wang, H., & Mintz, T. (2008). A dynamic learning model for categorizing words using frames. In H. Chan, H. Jacob, & E. Kapia (Eds.), *Proceedings of the 32nd annual Boston University Conference on Language Development [BUCLD 32]* (pp. 525–536). Somerville, MA: Cascadilla Press.