*What do you think that happens?
A quantitative and cognitive modeling analysis
of linguistic evidence across socioeconomic status
for learning syntactic islands

Alandi Bates and Lisa Pearl

1   Introduction

We know that by the time children from low-socioeconomic status (**SES**) backgrounds reach the age of 4, they will have encountered 30 million fewer words than their higher-SES counterparts (Hart and Risley, 1995; Schwab and Lew-Williams, 2016). There are also known differences in the quantity and quality of child-directed speech (**CDS**) at the lexical and foundational syntactic levels (Huttenlocher et al., 2010; Rowe, 2012). However, far less is known about the development of complex syntactic knowledge and the input differences that may exist at this level across SES. We investigate the differences in both the quality and quantity of complex syntactic input in American English CDS between high-SES populations and low-SES populations. More specifically, we look at the distributions of *wh*-dependencies that reflect knowledge of syntactic island constraints (a form of complex syntactic knowledge) in the speech directed at these two populations; we then assess whether the low-SES *wh*-dependency distribution supports the acquisition of syntactic island constraints as well as the high-SES distribution has been shown to (Pearl and Sprouse, 2013).

We first find that the low-SES *wh*-dependency distribution is both qualitatively and quantitatively similar to the high-SES CDS distribution. This suggests low-SES input may support acquisition of syntactic islands as well as high-SES input does. To evaluate this, we cognitively model the acquisition of syntactic island knowledge from the low-SES input, using the same probabilistic learning strategy that was successful with high-SES input (Pearl and Sprouse, 2013). We find that this same learning strategy is successful when learning from the low-SES CDS input, though a crucial syntactic building block involving complementizer *that* comes from a different *wh*-dependency type in low-SES CDS. Taken together, our results suggest that the nature of the input for learning about syntactic islands doesn't fundamentally differ across SES; this notably contrasts with input differences found for more foundational lexical and syntactic knowledge. We discuss implications for linguistic development across SES.

2 SES differences in language development and input

SES significantly impacts certain aspects of linguistic development in children (Huttenlocher et al., 2002; Vasilyeva et al., 2008; Frank et al., 2017), due to known differences across SES in input quantity (e.g., how often mothers talk to their children) and quality (e.g., the length of utterances and the complexity of constructions used). For example, high-SES parents tend to use more word types and tokens, as well as more diversified syntactic constructions than low-SES parents (Huttenlocher et al., 2007; Rowe, 2012). Notably, the input that these populations receive greatly impacts their future linguistic outcomes. For instance, children from high-SES families generally have larger vocabularies and build these vocabularies earlier and faster than their low-SES counterparts (Hart and Risley, 1995; Hoff, 2003). There's also suggestive evidence that effective intervention is possible when needed (Huttenlocher et al., 2002; Rowe et al., 2017), which could potentially mediate the impact of SES-based input differences. For instance, when teachers provided high-quality input to their students, low-SES students showed improvement in measures of comprehension on par with their high-SES peers by the end of the school year (Huttenlocher et al., 2002). Of course, input-based interventions like this are only effective if we know what high-quality input actually consists of. That is, if we identify significant differences in the input across SES, we then know more precisely what's missing and how to fix it. We examine the nature of the input across SES for the complex syntactic knowledge known as syntactic island constraints, which is based on *wh*-dependencies.

2.1 *Wh*-dependencies & syntactic islands for High-SES children

One hallmark of the syntax of human languages is the ability to have long-distance dependencies: relationships between two words in a sentence that are not adjacent to each other. Long-distance dependencies, such as the dependencies between the *wh*-word *what* and *eat* in (1), can be arbitrarily long (Chomsky, 1965; Ross, 1967; Chomsky, 1973).

(1)   a.   What did Falkor eat $\_\_{what}$?
         b.   What did Atreyu see Falkor eat $\_\_{what}$?
         c.   What did the Childlike Empress say Atreyu saw Falkor eat $\_\_{what}$?
         d.   What did Bastian hear the Childlike Empress say Atreyu saw Falkor eat $\_\_{what}$?

However, there are specific syntactic structures that long-distance dependencies can't cross: syntactic islands. Four examples of syntactic islands are in (2) (Chomsky, 1965; Ross, 1967; Chomsky, 1973), with * indicating ungrammaticality and [...] highlighting the proposed island structure that a *wh*-dependency can't cross in English.

(2) a. **Complex NP island**
        *What did Falkor make [the claim [$_{CP}$ that Atreyu fought $\_{what}$]]?
    b. **Subject island**
        *What did Falkor think [[the joke about $\_{what}$] was hilarious]?
    c. **Whether island**
        *What did Falkor wonder [whether Atreyu fought $\_{what}$]?
    d. **Adjunct island**
        *What did Falkor worry [if Atreyu fights $\_{what}$]?

During language development, children must infer and internalize the constraints on long-distance dependencies (i.e., syntactic island constraints) that allow them to recognize that the questions in (2) are not allowed, while the questions in (1) are fine.

Pearl and Sprouse (2013) constructed a cognitive computational model for learning these syntactic island constraints. This model relies on the idea that children can characterize a long-distance dependency as a syntactic path from the head of the dependency (e.g., *What* in (3)) through a set of structures that contain the tail (e.g., $\_{what}$) of the dependency, as shown in (3a)-(3b). These structures correspond to phrase types such as Verb Phrases (VP), Inflectional Phrases (IP), and Complementizer Phrases (CP), among others. Under this view, children simply need to learn which long-distance dependencies have licit syntactic paths and which don't.

To model this learning process, Pearl and Sprouse (2013) implemented a probabilistic learning algorithm that tracks local pieces of these syntactic paths. It breaks the syntactic path into a collection of syntactic trigrams that can be combined to reproduce the original syntactic path, as shown in (3c). The learning model then tracks the frequencies of these syntactic trigrams in the input. It later uses them to calculate probabilities for all syntactic trigrams comprising a *wh*-dependency[1] and so generate the probability of any *wh*-dependency (as shown in (4)- (5)). The generated probability corresponds to whether that dependency is allowed, with higher probabilities indicating grammatical dependencies and lower probabilities indicating ungrammatical dependencies.

(3)   What did Falkor claim that Atreyu fought $\_{what}$?

    a. Syntactic structures containing the *wh*-dependency:
        What did [$_{IP}$ Falkor [$_{VP}$ claim [$_{CP}$ that [$_{IP}$ Atreyu [$_{VP}$ fought $\_{what}$]]]]]?
    b. Syntactic path of *wh*-dependency:
        start-IP-VP-CP$_{that}$-IP-VP-*end*

---

[1]It smooths these probabilities by adding 0.5 to all trigram counts. This allows the model to accept dependencies composed of trigrams it's never seen before, though it gives them a much lower probability than dependencies composed of trigrams it has in fact seen before. See Pearl and Sprouse (2013, 2015) for further discussion of this point.

    c.   Syntactic trigrams $T \in$ syntactic path:
  = $start$-IP-VP
     IP-VP-CP$_{that}$
        VP-CP$_{that}$-IP
           CP$_{that}$-IP-VP
              IP-VP-end

(4)   p($start$-IP-VP) $\approx \frac{count(start-IP-VP)}{total\ count\ all\ trigrams}$

    ...

    p(IP-VP-end) $\approx \frac{count(IP-VP-end)}{total\ count\ all\ trigrams}$

(5)   p(What did Engywook tell Atreyu $\_\_\_what$?)
    = p($start$-IP-VP-end)   trigrams = $start$-IP-VP, IP-VP-end
    = p($start$-IP-VP)*p(IP-VP-end)

To evaluate high-SES syntactic input quality, Pearl and Sprouse (2013) let the model learn from a realistic sample of high-SES American English CDS equivalent to the quantity of data children typically encounter during the time when they're learning about syntactic island constraints. The high-SES input data were sampled from the structurally-annotated Brown-Adam (Brown, 1973), Brown-Eve (Brown, 1973), Valian (Valian, 1991), and Suppes (Suppes, 1974) corpora from the CHILDES Treebank (Pearl and Sprouse, 2013), comprising 102K utterances with 21K wh-dependencies. With this input, the model estimated syntactic trigram probabilities and could then generate probabilities for any desired wh-dependency.
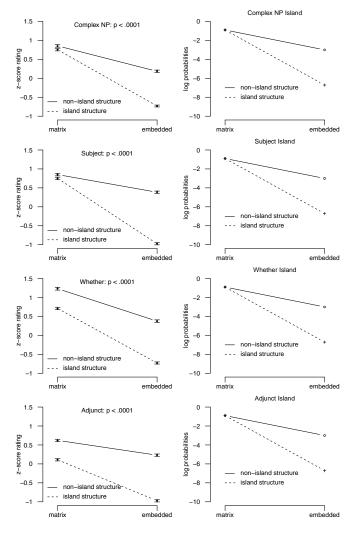
The wh-dependencies that the model needed to generate probabilities were those that American English adults had given acceptability judgments for in Sprouse et al. (2012) – (6) shows a sample set for Complex NP islands, with island structures indicated with [...]. These stimuli were designed using a 2x2 factorial design, involving dependency length (matrix vs. embedded) and presence of an island structure in the utterance (non-island vs. island).

(6)   Sample Complex NP Island stimuli

    a.   matrix+non-island:
        Who $\_\_\_who$ claimed that Atreyu fought the goblin?
    b.   embedded+non-island:
        Who did Falkor claim that Atreyu fought $\_\_\_who$?
    c.   matrix+island:
        Who $\_\_\_who$ made [the claim that Atreyu fought the goblin]?
    d.   embedded+island:
        *Who did Falkor make [the claim that Atreyu fought $\_\_\_who$]?

This design demonstrates the existence of syntactic island knowledge as a superadditive interaction of acceptability judgments, which appears as non-parallel lines in an interaction plot, such as those in Figure 1. In particular, if we consider

the Complex NP plot in the top row , there are four acceptability judgments, one for each of the stimuli in (6).

Figure 1: Left column: High-SES adult judgments demonstrating implicit knowledge of four syntactic islands via a superadditive interaction. Right column: Modeled high-SES child judgments demonstrating the same implicit knowledge via a superadditive interaction.

The matrix+non-island dependency of (6a) has a certain acceptability score – this is the top-lefthand point. There is a (slight) drop in acceptability when the matrix+island dependency of (6a) is judged in comparison to (6c) – this is

5

the lower-lefthand point. We can interpret this as the unacceptability associated with simply having an island structure in the utterance. There's also a drop in acceptability when the embedded+non-island dependency of (6b) is judged in comparison to (6a) – this is the upper-righthand point. We can interpret this as the unacceptability associated with simply having an embedded *wh*-dependency. If the unacceptability of the embedded+island dependency of (6d) were simply the result of those two unacceptabilities (having an island structure in the utterance and having an embedded *wh*-dependency), the drop in unacceptability would be additive and the lower-righthand point would be just below the upper-righthand point (and so look just like the points on the lefthand side). But this isn't what we see – instead, the acceptability of (6d) is much lower than this. This is a superadditive effect for the embedded+island stimuli. So, the additional unacceptability of an island-crossing-dependency like (6d) – i.e., implicit knowledge of syntactic islands – appears as a superadditive interaction in these types of acceptability judgement plots.

The left column of Figure 1 shows the results of collecting acceptability judgments from high-SES adult speakers using that design. The visible superadditive interactions demonstrate implicit knowledge of the four syntactic islands in (2) in English. The right coumn of Figure 1 shows the log probability for the same stimuli for each of the four islands, as predicted by the learning model in Pearl and Sprouse (2013). Log probabilities are reported for each dependency because the probabilities are very small numbers (due to the multiplication of syntactic trigram probabilities). The visible superadditive interactions indicate that the high-SES input was sufficient to scaffold the development of these syntactic island constraints.

2.2 Low-SES input for syntactic islands

Here we assess low-SES input, focusing on the information necessary for the development of the implicit syntactic islands knowledge that was previously assessed by Pearl and Sprouse (2013) for high-SES input. We first want to identify if there are any quantitative differences between the high-SES and low-SES input samples we have in terms of the *wh*-dependencies available, as these dependencies are the foundation of the development of syntactic island constraints. We will answer this question via quantitative analysis of the distribution of *wh*-dependencies available. We then want to identify if there are any qualitative differences between the high-SES and low-SES input in terms of how well the *wh*-dependencies available scaffold the development of syntactic island constraints. That is, whether any quantitative differences exist or not, does low-SES input differ from high-SES input in how it allows complex syntactic development to occur? We'll answer this question by applying the same computational learning model from Pearl and Sprouse (2013) that allows successful acquisition of this knowledge from high-SES input. If successful acquisition of island constraints occurs when learning from low-SES input, this would suggest low-SES is not qualitatively different from high-SES input in this respect. In contrast, if successful acquisition doesn't

occur when learning from low-SES input, this would implicate a qualitative difference for complex syntactic acquisition between low-SES and high-SES input.

2.3 Low-SES CDS corpus

We assessed low-SES CDS from a subpart of the HSLLD corpus (Dickinson and Tabors, 2001) in CHILDES (MacWhinney, 2000), which came from the Elicited Report, Mealtime, and Toy Play sections of Home Visit 1. This sample contained 31,875 utterances and 3,904 *wh*-dependencies, directed at 78 children between the ages of 3 and 5.

We extracted and syntactically annotated all *wh*-dependencies following the format of the CHILDES Treebank (Pearl and Sprouse, 2013), and then coded the syntactic paths of the dependencies (as in (3b) and shown below with a different example in (7)). Following Pearl and Sprouse (2013), the *CP* nodes were further subcategorized by the lexical item serving as complementizer, such as CP_{that}, $CP_{whether}$, CP_{if}, and CP_{null}. This allows the modeled learner of Pearl and Sprouse (2013) to distinguish grammatical dependencies like (7a) from ungrammatical ones like (7b). With these syntactic paths in hand, we can then assess the distribution of these *wh*-dependencies, characterized this way, in the low-SES input sample.

(7) a. Who do you think ___$_{who}$ read the book?
          syntactic path: *start-IP-VP-CP$_{null}$-IP-end*
     b. *Who do you think that ___$_{who}$ read the book?
          syntactic path: **start-IP-VP-CP$_{that}$-IP-end*

3 *Wh*-dependencies across SES

Descriptive corpus analysis. Our corpus analysis revealed 16 *wh*-dependency types in the low-SES input, 12 of which also appeared in the high-SES corpus analysis of Pearl and Sprouse (2013).[2] Additionally, the low-SES input contained 3 *wh*-dependency types not in the high-SES input:

- *start-IP-VP-CP$_{null}$-IP-VP-NP-PP-end*
  (e.g., *What did he think it was a movie of ___$_{what}$?*)

- *start-IP-VP-IP-VP-IP-VP-PP-IP-VP-end*
  (e.g., *What did you want to try to plan on doing ___$_{what}$?*)

- *start-IP-VP-CP$_{that}$-IP-end*
  (e.g., *What do you think that ___$_{what}$ happens?*)

---

[2]A more detailed description of the *wh*-dependency distribution across SES is available in Appendix A in the supplementary materials at http://sites.uci.edu/alandibates/publications/ and http://socsci.uci.edu/~lpearl/research.html.

Interestingly, this last dependency type is an example of a "*that*-trace" violation and is ungrammatical in the high-SES dialect. Also, the two dependency types that account for the vast majority of the low-SES *wh*-dependency input (85.8%) are the same two that account for the vast majority of the high-SES input (89.5%), and they occur in about the same proportions:

- 75.5% low-SES, 76.7% high-SES: *start-IP-VP-end*
  (e.g., *What did Lily read ___$_{what}$?*)

- 10.3% low-SES, 12.8% high-SES: *start-IP-end*
  (e.g., *What ___$_{what}$ happened?*)

This suggests a high-level qualitative similarity in the *wh*-dependency input across SES.

Quantitative analysis. To more precisely quantify how similar the input distributions are, we use the Kullbeck-Leibler (**KL**) divergence (Kullback and Leibler, 1951) to assess the similarity in the *wh*-dependency distributions in low-SES vs. high-SES CDS. We additionally use KL divergence to assess how similar these distributions are to distributions in the high-SES adult-directed speech (**ADS**) and adult-directed text (**ADT**) from Pearl and Sprouse (2013) as a comparison baseline (the corpora are described in Table 1).

Table 1: Corpora statistics for low-SES CDS (L-CDS), high-SES CDS (H-CDS), high-SES adult-directed speech (H-ADS), and high-SES adult-directed text (H-ADT).

corpora	# utterances	# *wh*-dependencies	# children	ages
L-CDS	31,875	3,904	78	3 - 5
H-CDS	101,838	20,923	25	1 - 5
H-ADS	74,576	8,508	N/A	N/A
H-ADT	24,243	4,230	N/A	N/A

We note that the KL divergence was only calculated over the distribution of the 9 *wh*-dependencies (see Table 2) these four corpora had in common, which accounted for 99.1%-99.6% of the total *wh*-dependencies in these corpora.

Figure 2 shows the results of this analysis. Higher KL values indicate greater divergence in the distributions, while values closer to zero indicate distributions that are more similar. We see that low-SES CDS and high-SES CDS are the most similar in *wh*-dependency distribution (KL: 0.01324), and appear to be twice as similar as the next closest comparison, which is high-SES CDS vs. high-SES ADS (KL: 0.02658). This affirms a quantitative similarity across SES in child *wh*-dependency input. Moreover, these results highlight that CDS *across* SES is more similar than CDS vs. ADS *within* SES. That is, whether the speech is

Table 2: The nine *wh*-dependencies shared across all four corpora that are used in the KL divergence analysis.

Shared dependencies	Example utterance	Corpora percentage
start-IP-*end*	*Who saw it?*	10.3% - 33.0%
start-IP-VP-*end*	*Who did she see?*	63.3% - 76.7%
start-IP-VP-CP$_{null}$-IP-*end*	*Who did he think stole it?*	0.1% - 0.6%
start-IP-VP-CP$_{null}$-IP-VP-*end*	*What did he think she stole?*	0.2% - 1.1%
start-IP-VP-CP$_{null}$-IP-VP-PP-*end*	*What did he think she wanted it for?*	<0.1% - 0.1%
start-IP-VP-IP-VP-*end*	*What did he want her to steal?*	1.3% - 7.5%
start-IP-VP-IP-VP-IP-VP-*end*	*What did he want her to pretend to steal?*	<0.1%
start-IP-VP-IP-VP-PP-*end*	*What did she want to get out from under?*	<0.1% - 0.8%
start-IP-VP-PP-*end*	*Who did she steal from?*	1.3% - 4.3%

directed at children or adults matters more than whether speech is coming from a high-SES or low-SES population.
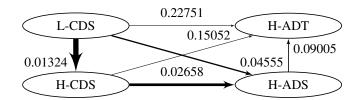

Figure 2: KL divergence for low-SES CDS (L-CDS), high-SES CDS (H-CDS), high-SES adult-directed speech (H-ADS), and high-SES adult-directed text (H-ADT). Line thickness corresponds to similarity, with thicker lines indicating more similar distributions.

We also note that these KL divergences accord with intuitions that speech of any kind is more similar to other speech than it is to text: high-SES ADS diverges more from high-SES ADT (KL: 0.09005) than it does from either high-SES CDS (KL: 0.02658) or low-SES CDS (KL: 0.04555).

4 Learning about syntactic islands

We can also assess qualitative similarity of input in terms of how that input affects learning outcomes. We use the same cognitive learning model developed by Pearl and Sprouse (2013); the modeled learner learns from the *wh*-dependency distribution in low-SES CDS input and generates probabilities for the four sets of experimental stimuli of Sprouse et al. (2012), which correspond to Complex NP, Subject, Whether, and Adjunct islands. These experimental stimuli can be characterized by the syntactic paths shown in Table 3. Note that many of the grammatical dependencies are characterized by the same syntactic path (e.g., *start*-IP-*end*); this is why Table 4, which shows the model's generated log probabilities of the relevant *wh*-dependencies, has only three grammatical dependency syntactic paths

listed. Figure 3 shows the low-SES CDS log probabilities plotted on interaction plots for each of the four island types. Table 4 also shows the log probabilities generated by learners learning from the high-SES CDS and high-SES ADS and ADT reported in Pearl and Sprouse (2013).

Table 3: Syntactic paths for experimental stimuli that acceptability judgments are generated for, in a 2x2 factorial design varying dependency length (*mat*rix vs. *emb*edded) and presence of an island structure (*non*-island vs. *island*). Ungrammatical island-spanning dependencies are indicated with *.

		Complex NP islands	*Subject islands*
mat	non	*start*-IP-*end*	*start*-IP-*end*
emb	non	*start*-IP-VP-CP$_{that}$-IP-VP-*end*	*start*-IP-VP-CP$_{null}$-IP-*end*
mat	island	*start*-IP-*end*	*start*-IP-*end*
emb	island	**start*-IP-VP-NP-CP$_{that}$-IP-VP-*end*	**start*-IP-VP-CP$_{null}$-IP-NP-PP-*end*
		Whether islands	*Adjunct islands*
mat	non	*start*-IP-*end*	*start*-IP-*end*
emb	non	*start*-IP-VP-CP$_{that}$-IP-VP-*end*	*start*-IP-VP-CP$_{that}$-IP-VP-*end*
mat	island	*start*-IP-*end*	*start*-IP-*end*
emb	island	**start*-IP-VP-CP$_{whether}$-IP-VP-*end*	**start*-IP-VP-CP$_{if}$-IP-VP-*end*

Table 4: Log probabilities of different *wh*-dependencies, representing acceptability judgments, for modeled learners learning from low-SES child-directed speech (L-CDS), as well as prior results from Pearl & Sprouse (2013) of modeled learners learning from high-SES child-directed speech (H-CDS) and high-SES adult-directed speech and text (H-ADS+H-ADT).

	L-CDS	H-CDS	H-ADS + H-ADT
Grammatical dependencies			
start-IP-*end*	-0.48	-1.21	-0.93
start-IP-VP-CP$_{null}$-IP-*end*	-8.11	-7.89	-7.67
start-IP-VP-CP$_{that}$-IP-VP-*end*	-15.88	-13.84	-11.00
Island-spanning dependencies			
start-IP-VP-NP-CP$_{that}$-IP-VP-*end*	-22.13	-19.81	-18.93
start-IP-VP-CP$_{null}$-IP-NP-PP-*end*	-20.12	-20.17	-20.36
start-IP-VP-CP$_{whether}$-IP-VP-*end*	-19.25	-18.54	-18.46
start-IP-VP-CP$_{if}$-IP-VP-*end*	-19.25	-18.54	-18.46

We can see that a core pattern emerges when learning from low-SES CDS: all grammatical dependencies have higher probabilities (equivalent to less negative log probabilities) than the island-spanning dependencies. In particular, grammatical dependencies have log probabilities ranging from -0.48 to -15.88, while island-spanning dependencies range from -19.25 to -22.13. This is the same pat-
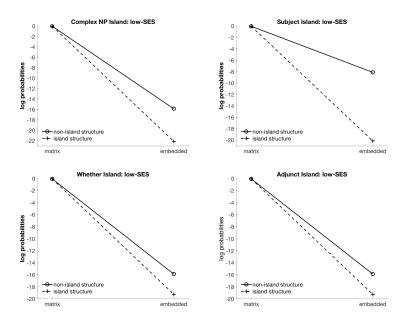
Figure 3: Judgments derived from a modeled learner using low-SES CDS, demonstrating implicit knowledge of syntactic islands as indicated by super-additivity (which appears as non-parallel lines in these interaction plots).

tern which was found when learning from either high-SES child-directed or adult-directed input (high-SES grammatical: -0.93 to -13.84; high-SES island-spanning: -18.46 to -20.36). Importantly, in Figure 3, we see the superadditivity that indicates implicit knowledge of syntactic island constraints. That is, just as with the log probabilities generated from the high-SES data and the acceptability judgments from high-SES adults, island-spanning dependencies are more unacceptable than would be predicted, given that they're embedded dependencies and they have an island structure in the utterance. This affirms what the low KL divergence between the low-SES and high-SES CDS *wh*-dependencies suggested: the input quality is the same across SES, with respect to the development of the complex syntactic knowledge of syntactic island constraints.

5   Discussion

Our results suggest that the *wh*-dependency input that low-SES children receive is quantitatively and qualitatively similar to the input of high-SES children. This similarity allows a modeled child to learn implicit knowledge of syntactic islands from low-SES input just as easily as from high-SES input, as demonstrated by the modeled judgment behavior.

Interestingly, there's a striking difference in the exact *wh*-dependency distribution across SES that turns out to be crucial for acquisition success for two of the syntactic island types. This difference involves a particular structural building block, which comes from dependencies that are characterized with CP_{that}.

As noted in (7), the only distinction between certain grammatical dependencies and certain ungrammatical dependencies is the complementizer. Example (7) showed this for a grammatical dependency with the *null* complementizer and an ungrammatical dependency with complementizer *that*. Another key example is the difference between grammatical dependencies with complementizer *that* (8a) and ungrammatical dependencies with complementizers like *whether* (whether islands) or *if* (adjunct islands) (8b).

(8)   a.   What do you think that Jack read __$_{what}$?
              syntactic path: *start-IP-VP-CP$_{that}$-IP-VP-end*
    b.   *What do you wonder whether/if Jack read __$_{what}$?
              syntactic path: **start-IP-VP-CP$_{whether/if}$-IP-VP-end*

So, it's important that the child encounter *wh*-dependencies in her input that involve complementizer *that* (and not ones that involve complementizers *whether* or *if*). When this happens, the probabilistic learning strategy can leverage the CP_{that} building block to predict that (8a) should be judged as better than (8b). However, dependencies involving CP_{that} are actually fairly rare in naturalistic usage. Pearl and Sprouse (2013) only found 2 of 20,923 (0.0096%) in high-SES CDS, 7 of 8,508 (0.082%) in high-SES ADS, and 2 of 4,230 (0.048%) in high-SES ADT.

In the high-SES CDS sample, both dependencies involving CP_{that} are of the same type: *start-IP-VP-CP$_{that}$-IP-VP-end* instances like (8a). However, in our low-SES CDS sample, there are 2 of 3,094 (0.051%) dependencies involving CP_{that}, and they are both of a different type, which happens to be ungrammatical in the high-SES dialect: *start-IP-VP-CP$_{that}$-IP-end* instances like (9).

(9)   What do you think that __$_{what}$ happens?
         What do [$_{IP}$ you [$_{VP}$ think [$_{CP_{that}}$ that [$_{IP}$ __$_{what}$ [$_{VP}$ happens]]]]]?
         syntactic path: *start-IP-VP-CP$_{that}$-IP*

So, the presence of this *wh*-dependency type, which is ungrammatical in the high-SES dialect, provides the crucial CP_{that} building block necessary for the acquisition of whether and adjunct islands. That is, the key linguistic experience that would allow a child learning from low-SES CDS to acquire the same syntactic knowledge as a high-SES child actually comes from data that's ungrammatical for a high-SES child. This underscores the power of learning strategies that generate linguistic knowledge of larger structures from smaller building blocks; a child relying on smaller building blocks may be able to find evidence for those building blocks in unexpected places.

More generally, our results indicate that the input for the development of complex syntactic knowledge may not differ in impactful ways across SES, the

way it does for lexical or more foundational syntactic knowledge. That is, there may not be a "complex syntax gap" across SES. In the specific case of learning about syntactic islands, we would expect that once low-SES children are able to leverage the *wh*-dependency information in their input, they should learn about these syntactic islands as well as high-SES children do.

We note that the ability to leverage the *wh*-dependency information isn't trivial – there are known delays in language processing in low-SES children compared to their high-SES counterparts (Fernald et al., 2013; Weisleder and Fernald, 2013). However, our results here suggest that once the developmental milestones are met which allow successful processing of the available *wh*-dependency information in low-SES children's input, no other gap remains in low-SES children's input.

More concretely, the syntactic islands learning strategy applied here to the low-SES CDS data requires several foundational knowledge components and processing abilities to be "good enough" – that is, what the child must both know and be able to do in real time. First, the child must know about syntactic phrase structure; she must be able to use that phrase structure knowledge to extract the syntactic path of a *wh*-dependency in real time. Second, the child must know to break syntactic paths into smaller trigram building blocks that can be used to generate a probability for any *wh*-dependency; she must be able to identify these syntactic trigrams in real time. Third, the child must know to track the relative frequency of the syntactic trigrams; she must be able to track these frequencies in real time. Fourth, the child must know to combine these syntactic trigrams to generate the probability for a new *wh*-dependency; she must be able to do so in real time. Any or all of these components could be affected by processing deficits that arise from input quantity and quality differences in low-SES CDS, and it remains an open question which ones are in fact adversely affected by low-SES children's prior linguistic experience. Still, our current work has demonstrated that once low-SES children can use the *wh*-dependency information available to them, their input wouldn't cause them to lag behind their high-SES counterparts when it comes to learning about complex syntactic knowledge like syntactic islands.

References

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

Noam Chomsky. 1965. *Aspects of the Theory of Syntax*. The MIT Press, Cambridge.

Noam Chomsky. 1973. Conditions on transformations. In S. Anderson and P. Kiparsky, editors, *A Festschrift for Morris Halle*, Holt, Rinehart, and Winston, New York, pages 237–286.

David K Dickinson and Patton O Tabors. 2001. *Beginning literacy with language: Young children learning at home and school.*. Paul H Brookes Publishing.

Anne Fernald, Virginia A Marchman, and Adriana Weisleder. 2013. SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science* 16(2):234–248.

Michael C Frank, Mika Braginsky, Daniel Yurovsky, and Virginia A Marchman. 2017. Wordbank: An open repository for developmental vocabulary data. *Journal of child language* 44(3):677–694.

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

Erika Hoff. 2003. The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development* 74(5):1368–1378.

Janellen Huttenlocher, Marina Vasilyeva, Elina Cymerman, and Susan Levine. 2002. Language input and child syntax. *Cognitive Psychology* 45(3):337–374.

Janellen Huttenlocher, Marina Vasilyeva, Heidi R Waterfall, Jack L Vevea, and Larry V Hedges. 2007. The varieties of speech to young children. *Developmental Psychology* 43(5):1062.

Janellen Huttenlocher, Heidi Waterfall, Marina Vasilyeva, Jack Vevea, and Larry V Hedges. 2010. Sources of variability in children's language growth. *Cognitive Psychology* 61(4):343–365.

Solomon Kullback and Richard A Leibler. 1951. On information and sufficiency. *The annals of mathematical statistics* 22(1):79–86.

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

Lisa Pearl and Jon Sprouse. 2013. Computational Models of Acquisition for Islands. In Jon Sprouse and Norbert Hornstein, editors, *Experimental Syntax and Islands Effects*, Cambridge University Press, Cambridge, pages 109–131.

Lisa Pearl and Jon Sprouse. 2015. Computational modeling for language acquisition: A tutorial with syntactic islands. *Journal of Speech, Language, and Hearing Research* 58:740–753.

John Ross. 1967. *Constraints on variables in syntax*. Ph.D. thesis, MIT, Cambridge, MA.

Meredith L Rowe. 2012. A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development* 83(5):1762–1774.

Meredith L Rowe, Kathryn A Leech, and Natasha Cabrera. 2017. Going beyond input quantity: Wh-questions matter for toddlers' language and cognitive development. *Cognitive Science* 41:162–179.

Jessica F Schwab and Casey Lew-Williams. 2016. Language learning, socioeconomic status, and child-directed speech. *Wiley Interdisciplinary Reviews: Cognitive Science* 7:264–275.

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

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

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

Marina Vasilyeva, Heidi Waterfall, and Janellen Huttenlocher. 2008. Emergence of syntax: Commonalities and differences across children. *Developmental Science* 11(1):84–97.

Adriana Weisleder and Anne Fernald. 2013. Talking to children matters early language experience strengthens processing and builds vocabulary. *Psychological Science* 24(11):2143–2152.

Supplementary Materials

A *Wh*-dependency distribution across SES

Table 5 shows the distribution of *wh*-dependencies across the different corpora, including the low-SES and high-SES child-directed speech, as well as high-SES adult-directed speech and adult-directed text.

Table 5: Distribution of *wh*-dependencies in child-directed Low-SES (L-CDS) and High-SES (H-CDS), as well as High-SES adult-directed speech (H-ADS) and text (H-ADT). Percentages are shown for syntactic paths, 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. Dependencies used in the KL divergence analysis are in teal. The dependency in the Low SES dialect that's ungrammatical in the High SES dialect is in pink.

Distribution of *wh*-dependencies in the input			
Syntactic path and example utterance	L-CDS	H-CDS	H-ADS	H-ADT
IP	10.3%	12.8%	17.2%	33.0%
Who saw it?	402	2680	1464	1396
IP-VP	75.5%	76.7%	73.0%	63.3%
What did she see?	2949	16039	6215	2677
IP-VP-AdjP-IP-VP	0.0%	0.0%	<0.1%	0.1%
What are you willing to see?	0	0	1	5
IP-VP-AdjP-IP-VP-PP	0.0%	0.0%	<0.1%	0.0%
What are you willing to go to?	0	0	1	0
IP-VP-AdjP-PP	0.0%	0.0%	<0.1%	<0.1%
What are they good for?	0	0	1	1
IP-VP-CP$_{for}$-IP-VP-PP	0.0%	<0.1%	0.0%	0.0%
What did she put on for you to dance to?	0	1	0	0
IP-VP-CP$_{null}$-IP	0.1%	0.1%	0.6%	0.3%
Who did he think stole it?	5	24	52	12
IP-VP-CP$_{null}$-IP-VP	0.9%	1.1%	0.4%	0.2%
What did he think she stole?	39	236	30	8
IP-VP-CP$_{null}$-IP-VP-IP-VP	<0.1%	0.1%	<0.1%	0.0%
What did he think she wanted to steal?	3	28	3	0
IP-VP-CP$_{null}$-IP-VP-IP-VP-IP-VP	0.0%	<0.1%	0.0%	0.0%
What did he think she wanted to pretend to steal?	0	2	0	0
IP-VP-CP$_{null}$-IP-VP-IP-VP-IP-VP-PP	0.0%	0.0%	<0.1%	0.0%
Who did he think she wanted to pretend to steal from?	0	0	1	0
IP-VP-CP$_{null}$-IP-VP-IP-VP-PP	0.0%	<0.1%	0.0%	0.0%
Who did he think she wanted to steal from?	0	1	0	0
IP-VP-CP$_{null}$-IP-VP-NP	0.0%	<0.1%	<0.1%	<0.1%
What did he think she said about it?	0	1	5	1

Distribution of *wh*-dependencies in the input				
Syntactic path and example utterance	L-CDS	H-CDS	H-ADS	H-ADT
IP-VP-CP$_{null}$-IP-VP-NP-PP	<0.1%	0.0%	0.0%	0.0%
What did he think it was a movie of?	3	0	0	0
IP-VP-CP$_{null}$-IP-VP-PP	0.1%	0.1%	<0.1%	<0.1%
What did he think she wanted it for?	4	28	5	1
IP-VP-CP$_{null}$-IP-VP-PP-PP	0.0%	<0.1%	0.0%	0.0%
What did he think she wanted out of?	0	1	0	0
IP-VP-CP$_{that}$-IP	<0.1%	0.0%	0.0%	0.0%
What do you think that happens?	2	0	0	0
IP-VP-CP$_{that}$-IP-VP	0.0%	<0.1%	<0.1%	<0.1%
What did he think that she stole?	0	2	5	2
IP-VP-CP$_{that}$-IP-VP-IP-VP	0.0%	0.0%	<0.1%	0.0%
What did he think that she wanted to steal?	0	0	1	0
IP-VP-CP$_{that}$-IP-VP-PP	0.0%	0.0%	<0.1%	0.0%
Who did he think that she wanted to steal from?	0	0	1	0
IP-VP-IP	0.0%	<0.1%	<0.1%	0.0%
Who did he want to steal the necklace?	0	9	2	0
IP-VP-IP-VP	7.5%	5.6%	3.4%	1.3%
What did he want her to steal?	296	1167	287	57
IP-VP-IP-VP-IP-VP	<0.1%	<0.1%	<0.1%	<0.1%
What did he want her to pretend to steal?	2	11	6	1
IP-VP-IP-VP-IP-VP-PP	0.0%	0.2%	<0.1%	0.0%
Who did he want her to pretend to steal from?	0	43	6	0
IP-VP-IP-VP-IP-VP-PP-IP-VP	<0.1%	0.0%	0.0%	0.0%
What did you want to try to plan on doing?	1	0	0	0
IP-VP-IP-VP-NP	0.0%	<0.1%	0.0%	0.0%
What did he want to say about it?	0	6	0	0
IP-VP-IP-VP-NP-IP-VP	0.0%	0.0%	0.0%	<0.1%
What did he have to give her the opportunity to steal?	0	0	0	1
IP-VP-IP-VP-NP-PP	0.0%	<0.1%	<0.1%	0.0%
What did she want to steal more of?	0	1	1	0
IP-VP-IP-VP-PP	0.8%	0.4%	0.4%	<0.1%
What did she want to steal from?	35	74	33	4
IP-VP-IP-VP-PP-PP	0.0%	0.0%	0.0%	<0.1%
What did she want to get out from under?	0	0	0	1
IP-VP-NP	0.0%	0.2%	0.1%	0.1%
What did she say about the necklace?	0	52	10	5

Distribution of *wh*-dependencies in the input				
Syntactic path and example utterance	L-CDS	H-CDS	H-ADS	H-ADT
IP-VP-NP-IP-VP	0.0%	0.0%	<0.1%	<0.1%
What did he give her the opportunity to steal?	0	0	1	2
IP-VP-NP-PP	<0.1%	<0.1%	<0.1%	0.0%
What was she a member of?	1	7	6	0
IP-VP-PP	4.0%	2.5%	4.3%	1.3%
Who did she steal from?	159	524	369	57
IP-VP-PP-CP$_{null}$-IP	0.0%	0.0%	<0.1%	0.0%
What did she feel like was a very good place?	0	0	1	0
IP-VP-PP-CP$_{null}$-IP-VP	0.0%	<0.1%	0.0%	0.0%
What did she feel like he saw?	0	1	0	0
IP-VP-PP-IP-VP	<0.1%	0.0%	<0.1%	0.0%
What did she think about buying?	2	0	3	0
IP-VP-PP-NP	0.0%	0.0%	<0.1%	0.0%
Where was she at in the building?	0	0	2	0
IP-VP-PP-NP-PP	0.0%	<0.1%	0.0%	0.0%
What do you put it on top of?	0	2	0	0
IP-VP-PP-NP-PP-IP-VP	0.0%	0.0%	<0.1%	0.0%
What is she in the habit of doing?	0	0	1	0
IP-VP-PP-PP	0.5%	0.1%	0.0%	0.0%
What does he eat out of?	1	22	0	0
IP-VP-PP-IP-VP	0.0%	<0.1%	0.0%	0.0%
What did he think about stealing?	0	1	0	0