**Title: Identifying if input differences are developmentally meaningful:
A look at complex syntactic input across socio-economic status**

Running Title: A look at complex syntactic input across socio-economic status

Authors: Alandi Bates and Lisa Pearl

Conflict of Interest Statement: We certify that there are no conflicts to report.

Acknowledgments: We are very grateful to the Quantitative Language Collective at UC Irvine for comments and discussion, as well as the audience at BUCLD 2018, Meredith Rowe, and three anonymous reviewers.

*Alandi Bates, University of California, Irvine, ajbates@uci.edu. Lisa Pearl, University of California, Irvine, lpearl@uci.edu.

Identifying if input differences are developmentally meaningful:
A look at complex syntactic input across socio-economic status

Research Highlights

- We describe a new way to identify developmentally-meaningful linguistic input differences, applying it to a case study in complex syntactic knowledge development. We focus on how children from different socio-economic status (**SES**) backgrounds use their input to learn constraints on *wh*-dependencies, called *syntactic islands*.
- Descriptive corpora analyses and quantitative analyses suggest that the relevant input distributions across SES are similar.
- Developmental computational modeling analyses suggest that the relevant input distributions are predicted to lead to the same learning outcomes for syntactic island knowledge; this result suggests any input differences across SES for this complex syntactic knowledge aren't developmentally meaningful.
- Our results also highlight how surface input differences may mask deeper input similarities, because a crucial structural building block for learning syntactic island knowledge is predicted to come from different parts of the input across SES.

Abstract

There's much naturally-occurring variation in children's input, but not all variation is developmentally meaningful – that is, variation that qualitatively impacts language development. While there are observed differences in input quantity and quality across socio-economic status (SES), it's unclear how often these input differences are developmentally meaningful. We describe a new way to identify developmentally-meaningful input differences that harnesses developmental computational modeling, which allows us to identify which aspects of the input are used by children to learn specific linguistic knowledge. We then investigate if there's developmentally-meaningful input variation across SES with respect to the complex syntactic knowledge called *syntactic islands*, which is used to form grammatical *wh*-questions.

Using quantitative analysis and cognitive modeling to assess low-SES child-directed speech samples, we find that the relevant data for learning about syntactic islands in low-SES children's input are quantitatively and qualitatively similar to those of high-SES children. In particular, low-SES children's input is predicted to also allow successful acquisition of syntactic island knowledge. Interestingly, at least one key building block for syntactic island knowledge comes from a different source in low-SES children's input, but is crucially still present. This highlights an important qualitative input similarity across SES. Our results suggest that the linguistic evidence for more complex syntactic knowledge like syntactic islands, in contrast with more foundational linguistic knowledge, may not differ by SES. We discuss implications for linguistic development and adult syntactic knowledge variability across SES.

Key Words: socioeconomic status, linguistic development, child-directed speech, syntactic islands constraints, computational modeling, quantitative approaches, input quantity, input quality

1 Identifying developmentally-meaningful input differences across socio-economic status

There's a lot of naturally-occurring variation in children's input, including how long they're talked to every day, which people talk to them (e.g., adults, other children), what environments they experience language interaction in (e.g., home, daycare, school), and what people talk to them about, among many other types of variation. Importantly, not all this input variation is *developmentally* *meaningful* – that is, not all input variation impacts language development in a way that deviates significantly from a typically-developing child's trajectory. However, some input variation does indeed impact language development – this variation is then developmentally meaningful; in particular, developmentally-meaningful input deficits lead to language development delays.

Language development delays appear across socio-economic status (**SES**), with lower-SES children behind their higher-SES peers with respect to different aspects of language development (e.g., vocabulary development (Hart & Risley 1995, Hoff 2003), language processing (Fernald et al. 2013, Weisleder & Fernald 2013)). But are these input-based delays? Certainly, there are observed input quantity and input quality differences across SES (though also within SES).

For instance, when it comes to input quantity at the word-level, some studies have found that lower-SES children may encounter 30 million fewer words of caretaker speech than their higher-SES peers (Hart & Risley 1995, Schwab & Lew-Williams 2016); other studies have found greater differences in input quantity within SES rather than across SES (Blum 2015, Sperry et al. 2018). At the clause-level, however, there appear to be fewer observed input quantity differences across SES; for instance, caretakers across SES produce approximately the same number of multi-clause utterances (e.g., *[He gave the book to the girl [who lived down the street]]* = 2 clauses) in their child-directed speech (Huttenlocher et al. 2002).

For input quality, differences have been observed at the lexical and foundational syntactic levels (Huttenlocher et al. 2010, Rowe 2012, Rowe et al. 2017), including the relative frequency of word types, word tokens, rare words, the diversity of syntactic constructions, and the relative frequency of decontexualized utterances like explanations (*Oh, we can't put them in the bus because the bus is full of blocks*), pretend (*I'll save you from the wicked sister*), and narrations (*He is going to look in your nose and your throat and your ears*).

Yet, how do we know if any observed input difference (whether about input quantity or input quality) is developmentally meaningful? A standard way to determine this is to observe some input difference, observe language development outcomes, and then see if the observed input difference is correlated with any observed outcome difference. If so, the language input difference *might* cause the language development outcome difference; so, targeting the input difference for intervention may lead to improved language development outcomes (e.g., input-based interventions allowing low-SES students to improve their language comprehension (Huttenlocher et al. 2002)). If intervention is indeed effective, this is more support that the language input difference caused the observed language outcome difference, and was therefore developmentally meaningful.

A new, complementary way to investigate if any observed input difference is developmentally meaningful uses developmental computational modeling (see Pearl (in press) for an overview of this technique applied to language acquisition more generally). A developmental computational model implements a specific learning theory about how children use their input to acquire particular linguistic knowledge; children then use that linguistic knowledge to generate observable outcomes (e.g., correctly comprehending a word or determining if a question is well-formed). This means that a developmental computational model will pinpoint what aspect of the child's input is relevant, and predict the expected language development outcome on the basis of that relevant input (e.g., if they will comprehend a particular word or believe a question is well-formed). That is, a developmental computational model identifies if any observed input difference is predicted to be developmentally meaningful.

Simply put, if the input difference is predicted to lead to an observable outcome difference, then the input difference is predicted to be developmentally meaningful. Any predicted language development differences can then be evaluated through standard behavioral measures of assessing children's linguistic knowledge. If the predicted language outcome differences do indeed appear, then we have strong support that the input aspect highlighted by the developmental computational model is developmentally meaningful. Developmentally-meaningful input differences can then be targeted for intervention, with the strong possibility of positively impacting language development outcomes. We note that because a developmental computational model describes exactly how the input causes the predicted developmental outcome, the model can also predict if an observed input difference is *not* developmentally meaningful (because the predicted outcome isn't qualitatively different). In this case, we would expect an input-based intervention to be ineffective at improving children's language development.

Here, we harness developmental computational modeling to identify input differences that are developmentally meaningful. We use this approach to investigate input differences across SES for a certain type of complex syntactic knowledge known as *syntactic islands* that concerns *wh*-questions (e.g., the grammatical *Who did Lily think the pretty kitty was for?* vs. the ungrammatical *Who did Lily think the kitty for was pretty?*). More specifically, syntactic islands are constraints on the permitted forms of *wh*-questions (among other linguistic forms). Knowledge of syntactic islands thus allows children to know which *wh*-questions are well-formed and which aren't.

We first briefly review what is currently known about the development of complex syntactic knowledge across SES, focusing on knowledge related to *wh*-questions. We then discuss the complex syntactic knowledge that syntactic islands involve, and review a developmental computational model for learning syntactic islands; this model implements a specific learning theory for how children use their input to yield syntactic island knowledge. The learning theory pinpoints that the relevant aspect of children's input for learning syntactic islands involves *wh*-dependencies, which rely on "*wh*-words" like *what* and *who* (among others).

We then investigate input variation for learning syntactic islands, looking at the distributions of *wh*-dependencies in American English child-directed speech (**CDS**) between high-SES populations and low-SES populations. In particular, we provide both a descriptive corpus analysis and a quantitative analysis comparing high-SES to low-SES input. We then provide a developmental computational modeling analysis of the input quality, where we predict the syntactic island knowledge that low-SES children would be able to attain on the basis of their *wh*-dependency input.

By assessing the *wh*-dependency distributions in the CDS of high- and low-SES children this way, we can determine whether the low-SES *wh*-dependency distribution supports the acquisition of syntactic island knowledge as well as the high-SES distribution has been shown to do (Pearl & Sprouse 2013*a*). That is, our developmental computational modeling approach allows us to predict whether there are developmentally-meaningful differences across SES for the input that supports the development of syntactic island knowledge.

We find that the low-SES input, in terms of *wh*-dependency distribution and the syntactic building blocks needed for syntactic islands, is both quantitatively and qualitatively similar to the high-SES CDS distribution. More specifically, our modeling results predict that low-SES input can support acquisition of syntactic islands as well as high-SES input does. Thus, our results suggest that there are no developmentally-meaningful differences across SES coming from children's input, with respect to the development of this complex syntactic knowledge.

Interestingly, a syntactic building block involving complementizer *that* is predicted to be crucial for successful knowledge development and comes from a different *wh*-dependency type in low-SES CDS, compared with high-SES CDS; this difference highlights that surface input quality differences may mask deeper input quality similarities. 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 and potential adult syntactic knowledge variation.

2 The development of *wh*-dependency knowledge across SES

Currently, far less is known about the development of complex syntactic knowledge across SES (especially with respect to *wh*-dependencies) than about the development of lexical and foundational syntactic knowledge. Still, we do know about the development of some *wh*-dependency knowledge across SES and a little about the *wh*-dependency input.

High-SES English-learning children are able to correctly repeat back well-formed *wh*-questions like *Who can Falkor save?* and generate new well-formed *wh*-questions by two and a half to three years old (Valian & Casey 2003). This suggests that these children have knowledge of core components of English *wh*-questions like fronting the *wh*-word (e.g., *who*) and moving the auxiliary (e.g., *can*) to the position before the subject (e.g., *Falkor*).

In addition, high-SES English-learning children appear to know a complex constraint on *wh*-questions (called a *relative clause island*) by three to four years old (de Villiers & Roeper 1995).

In particular, three- and four-year-olds can correctly interpret complex *wh*-questions like *"What is Jane drawing a monkey that is drinking milk with?"*. The only viable interpretation for adults is "*What is Jane drawing [a monkey that is drinking milk] with?*", where *with* refers to what Jane is using to draw (e.g., a pencil); this contrasts with an incorrect interpretation like "*What is Jane drawing [a monkey that is drinking milk with]?*", where *with* refers to what the monkey is using to drink the milk (e.g., a straw). The high-SES children showed adult-like interpretations for these complex *wh*-questions, correctly answering *a pencil* for this question and rarely answering *a straw*.

Interestingly, the evidence we have across SES indicates that learning outcomes for some *wh*-dependency knowledge are similar. In particular, there's suggestive evidence that by age four, children from diverse SES and linguistic backgrounds (i.e., both low- to high-SES children across dialects of American English) are capable of interpreting a variety of complex *wh*-questions, including the ones tested by de Villiers & Roeper (1995), similarly to adult speakers (de Villiers et al. 2008, Rombough & Thornton 2018). That is, there doesn't appear to be a difference in learning outcomes at four years old with respect to comprehension of many complex *wh*-questions across SES.

However, we know less about how children's input leads to the development of *wh*-dependency knowledge. We do know that the use of *wh*-questions in input to low-SES two-year-olds helps build their vocabulary and reasoning skills more generally (Rowe et al. 2017). However, it's unclear how the *wh*-questions in the input impact the development of complex *wh*-dependency knowledge (such as syntactic islands). More generally, much remains unknown about developmentally-meaningful input variation with respect to the development of complex syntactic knowledge, including how to form grammatical *wh*-questions.

3 Syntactic islands

A key component of human syntactic knowledge is the ability to have long-distance dependencies, where there's a relationship between two words that aren't 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). In (1), we can see that this dependency can stretch across one, two, three, or four clauses. In each case, *what* is understood as the thing Falkor ate, despite *what* not being adjacent to *eat*. This relationship is marked with $\_\_{what}$.

(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 (Chomsky 1965, Ross 1967, Chomsky 1973). Four examples of syntactic islands are in (2), 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 [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 bought $\_\_{what}$]?
d. **Adjunct island**
*What did Falkor worry [if Atreyu buys $\_\_{what}$]?

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. We note that this recognition is the measurable behavior of children's internalized knowledge – that is, distinguishing grammatical questions like (1) from ungrammatical questions like (2) is one way to indicate knowledge of the relevant syntactic island constraints (whatever form that knowledge may take).

4 Linking children's input to syntactic island knowledge development

Pearl & Sprouse (2013*a*) constructed a developmental computational model for learning these syntactic island constraints. This model relies on a specific learning theory that assumes 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 that make up *wh*-questions such as Verb Phrases (VP), Inflectional Phrases (IP), and Complementizer Phrases (CP), among others. Importantly, these are the structures that *wh*-dependencies could cross when forming *wh*-questions. Under this view, children simply need to learn which long-distance dependencies have licit syntactic paths and which don't.

To developmentally model this learning process, Pearl & Sprouse (2013*a*) implemented their learning theory as 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).[1]

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[2] and so generate the probability of any *wh*-dependency (as shown in (4)- (5)). More specifically, any *wh*-dependency's probability is the product of the individual trigram probabilities that comprise its syntactic path, as shown in (6).

The generated probability corresponds to whether that dependency is allowed, with higher probabilities indicating grammatical dependencies and lower probabilities indicating ungrammatical dependencies. So, the modeled learner can generate judgments of *wh*-questions (e.g., grammatical vs. ungrammatical); if this learner can generate the same pattern of judgments that adults do, we can assume the learner has internalized some version of the knowledge adults use to make those judgments. In this case, that means the modeled learner has internalized knowledge (via the syntactic trigrams) that allow it to capture syntactic island constraints. In this way, we can say that it's learned those syntactic island constraints.

(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*

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*

[1]For discussion of the empirical motivation for the modeling choices, including using trigrams and the aggregation of trigrams into a dependency, please see Pearl & Sprouse (2013*a*).

[2]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 & Sprouse (2013*a*, 2015) for further discussion of this point.

(4) Smoothed probabilities of trigrams:
p(*start*-IP-VP) $\approx \frac{count(start-IP-VP)}{total\ count\ of\ all\ trigrams}$
...
p(IP-VP-*end*) $\approx \frac{count(IP-VP-end)}{total\ count\ of\ all\ trigrams}$

(5) Probability of new *wh*-dependency: What did Engywook tell Atreyu $\_\__{what}$?
Syntactic structures = What did [$_{IP}$ Engywook [$_{VP}$ tell Atreyu $\_\__{what}$?]]
Syntactic path = *start*-IP-VP-*end*
trigrams = *start*-IP-VP, IP-VP-*end*
Probability = p(*start*-IP-VP-*end*) = p(*start*-IP-VP)*p(IP-VP-*end*)

(6) General formula for generating a *wh*-dependency's probability:
$\prod_{trigrams \in T} p(trigram)$

We note that this developmental learning model requires children to have certain (potentially sophisticated) knowledge and abilities before they can utilize the learning strategy implemented by this model. Core assumptions of the model require that the child be able to (i) parse sentences into phrase structure trees, (ii) extract sequences of container nodes for the dependencies, (iii) track the frequency of trigram container nodes, and (iv) calculate the probability for the complete container node sequence of the dependency, based on its trigrams. It remains for future work to determine when these core pieces are in place in children – once they are, children would be able to harness the input the way this learning model does. [3]

5 High-SES input quality for syntactic islands

To evaluate high-SES syntactic input quality, Pearl & Sprouse (2013*a*) modeled a learner using this strategy and let the modeled learner use as input a realistic sample of high-SES American English CDS. These high-SES input data came 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 & Sprouse 2013*a*), comprising 102K utterances with 21K *wh*-dependencies. The modeled learner encountered a quantity of CDS equivalent to the quantity of data high-SES children typically encounter during the time when they're learning about syntactic island constraints (estimated to take three years), which was equivalent to ≈200K *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 for were those that American English adults had given acceptability judgments for in Sprouse et al. (2012), corresponding to the four islands from (2); a sample set for each island type is shown in (7)-(10), where island structures are 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). Each island stimuli set therefore had four dependency types: matrix+non-island, embedded+non-island, matrix+island, and embedded+island; the embedded+island stimulus in each case involved a *wh*-dependency that crossed a syntactic island, and so was ungrammatical. These experimental stimuli can be characterized by the syntactic paths shown in Table 1. Note that many of the grammatical dependencies for each island type (e.g., matrix+non-island and matrix+island) are characterized by the same syntactic path (e.g., *start*-IP-*end*).

[3] We also note that we're using this learning model as a novel way to assess input quality, on the basis of prior work (Pearl & Sprouse 2013*a*) that applies it to high-SES children's input. Because of this focus, we won't discuss the theoretical implications of this learning strategy for questions of innateness with respect to the knowledge needed and assumed by the model; we instead refer interested readers to the discussion in Pearl & Sprouse (2013*a*).

(7) 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 $\_\_{who}$]?

(8) Sample Subject island stimuli

a. matrix+non-island:
Who $\_\_{who}$ thinks the joke is hilarious?

b. embedded+non-island:
What does Falkor think $\_\_{what}$ is hilarious?

c. matrix+island:
Who $\_\_{who}$ thinks the joke about Atreyu is hilarious?

d. embedded+island:
*Who did Falkor think [[the joke about $\_\_{who}$] was hilarious]?

(9) Sample Whether island stimuli

a. matrix+non-island:
Who $\_\_{who}$ thinks Atreyu bought the medallion?

b. embedded+non-island:
What does Falkor think Atreyu bought $\_\_{what}$?

c. matrix+island:
Who $\_\_{who}$ wonders if Atreyu bought the medallion?

d. embedded+island:
*What did Falkor wonder [whether Atreyu bought $\_\_{what}$]?

(10) Sample Adjunct island stimuli

a. matrix+non-island:
Who $\_\_{who}$ thinks Atreyu bought the medallion?

b. embedded+non-island:
What does Falkor think that Atreyu bought $\_\_{what}$?

c. matrix+island:
Who $\_\_{who}$ worries if Atreyu bought the medallion?

d. embedded+island:
*What did Falkor worry [if Atreyu buys $\_\_{what}$]?

This design allows syntactic island knowledge to surface as a superadditive interaction of acceptability judgments; this superadditivity 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 (7). The matrix+non-island dependency of (7a) has a certain acceptability score – this is the top-lefthand point. There is a (slight) drop in acceptability when the matrix+island dependency of (7c) is judged in comparison to (7a) – this is 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 (7b) is judged in comparison to (7a) – 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 (7d) 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 (7d) 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 (7d) – i.e., implicit knowledge of syntactic

Table 1: 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*

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 high-SES adults. The right column of Figure 1 shows the log probability for the same stimuli for each of the four islands, as predicted by the developmental computational model in Pearl & Sprouse (2013*a*). Log probabilities are reported for each dependency because the probabilities are very small numbers (due to the multiplication of syntactic trigram probabilities).[4] The visible superadditive interactions indicate that the high-SES input was predicted to be sufficient to learn these syntactic island constraints.

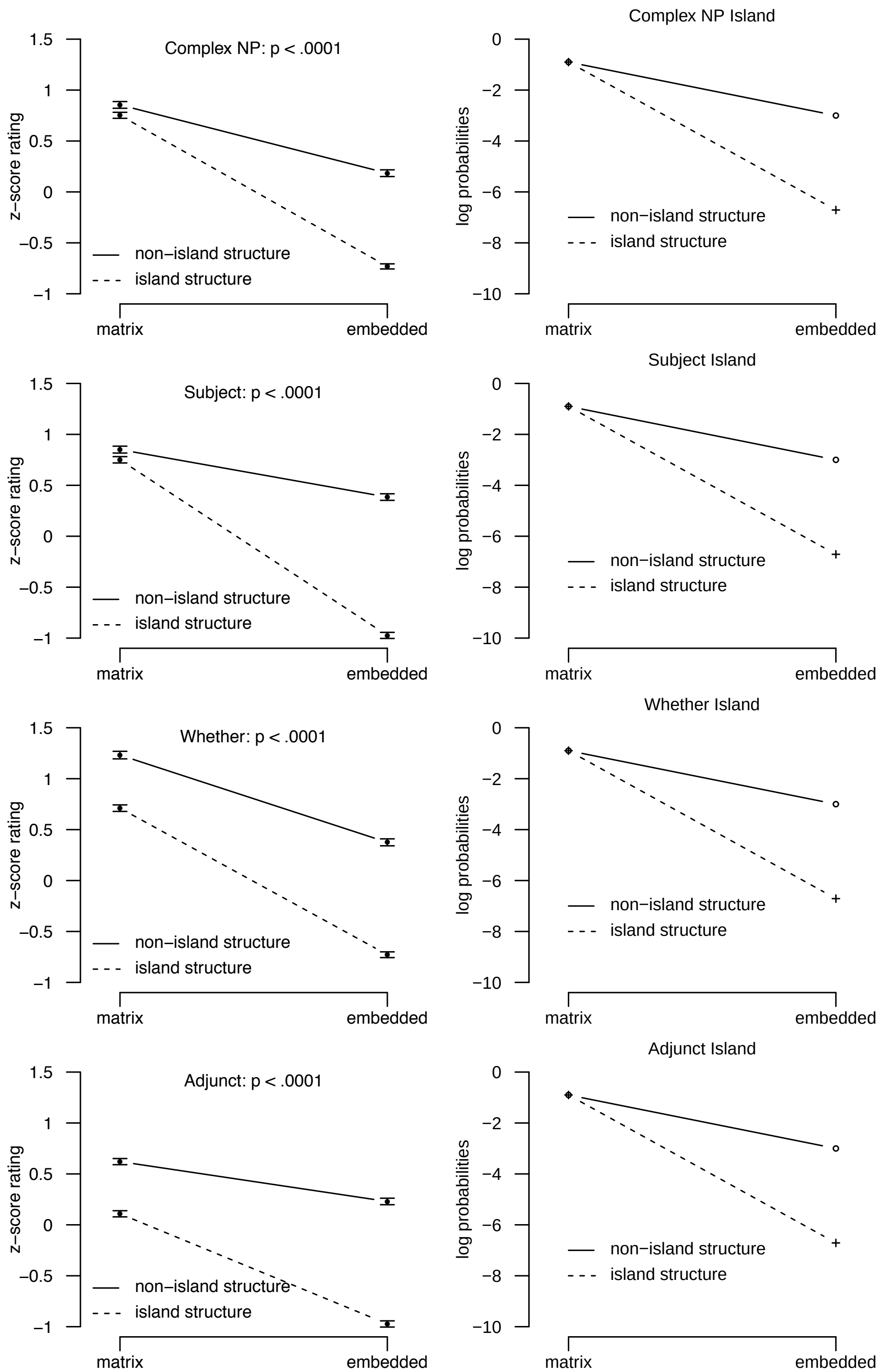
6 Low-SES input quality for syntactic islands

Here we assess low-SES input, focusing on the information necessary for the development of the implicit syntactic island knowledge that was previously assessed by Pearl & Sprouse (2013*a*) 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 and resulting syntactic trigrams available; recall that these dependencies and trigrams are the foundation of the development of syntactic island constraints, based on the learning theory in the model of Pearl & Sprouse (2013*b*). We'll identify quantitative input differences via quantitative analysis of the distribution of *wh*-dependencies and syntactic trigrams available.

We then want to identify differences between the high-SES and low-SES input in terms of how well the *wh*-dependencies and syntactic trigrams available scaffold the development of syntactic island constraints – these would be developmentally-meaningful differences. 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 & Sprouse (2013*a*) that allows successful acquisition of this knowledge from high-SES input. In particular, the modeled learner will learn from the same quantity of data a low-SES child would encounter, with the same input distributions, based on our low-SES CDS samples. If successful acquisition of island constraints occurs when learning from low-SES input, this would suggest low-SES input isn't qualitatively different from high-SES input in this respect; any input differences wouldn't be predicted to be developmentally meaningful. In

[4]For log probabilities, less negative numbers are equivalent to higher probabilities. For example, log(.001) = log(10^{-3}) = -3, while log(.000001) = log(10^{-6}) = -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.



contrast, if successful acquisition doesn't occur when learning from low-SES input, this would indicate a qualitative difference for complex syntactic acquisition between low-SES and high-SES input; so, any input differences would be predicted to be developmentally meaningful.

6.1 Low-SES CDS input samples

Our low-SES CDS input samples come from a subpart of the HSLLD corpus (Dickinson & Tabors 2001) in CHILDES (MacWhinney 2000), and SES was defined according to maternal education and annual income. Maternal education ranged from 6 years of schooling to some post-high school education. Annual income didn't have hard lower and upper bounds; instead, 70% of the families reported an annual income of $20,000 or less, while 21% of the families reported an income of over $25,000. The annual income of the remaining 9% was unreported. In this dataset, we focused on the Elicited Report, Mealtime, and Toy Play sections, which represent more naturalistic interactions. We also drew our samples from Home Visit 1, which recorded child language interactions involving children between the ages of three and five. Our sample contained 31,875 utterances and 3,904 *wh*-dependencies (12.2% of all utterances), directed at 78 children.

We extracted and syntactically annotated all *wh*-dependencies following the format of the CHILDES Treebank (Pearl & Sprouse 2013*b*), which indicates the syntactic structure necessary to characterize the syntactic paths of *wh*-dependencies. We then coded the syntactic paths of the dependencies (as in (3b) and shown below with a different example in (11)). Following Pearl & Sprouse (2013*b*), the *CP* phrase structure 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 & Sprouse (2013*b*) to distinguish dependencies judged by high-SES adults to be grammatical, like (11a), from those judged to be ungrammatical, like (11b). With these syntactic paths characterizing *wh*-dependencies, we can then assess the distribution of the *wh*-dependencies in the low-SES input sample.

(11) 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*

6.2 Descriptive corpus analysis

For *wh*-dependencies, 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 & Sprouse (2013*b*).[5] 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*?)

Interestingly, this last dependency type is an example of a "*that*-trace" violation and is judged ungrammatical by high-SES adults (Cowart 1997). This represents a difference across SES, with respect to adult knowledge of specific *wh*-dependencies. Additionally, when we compare the rate (and therefore quantity) of *wh*-dependencies across SES, we find another difference. The *wh*-dependency rate in the high-SES CDS sample of Pearl & Sprouse (2013*b*) was 20.5%, while the *wh*-dependency rate in our low-SES CDS sample was 12.2% – a lower rate.

[5] A more detailed description of the *wh*-dependency distribution across SES is available in Appendix A.1.

However, there's a striking similarity when we look at the most frequent *wh*-dependencies types across SES: 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 (shown in (12)). This suggests a high-level qualitative similarity in the *wh*-dependency input across SES, despite the individual *wh*-dependency differences.

(12) Proportions of the two most frequent *wh*-dependency types across SES
  a. 1^{st} most frequent: *start-IP-VP-end* (e.g., *What did Lily read* $\_\__{what}$*?*)
     75.5% low-SES, 76.7% high-SES
  b. 2^{nd} most frequent: *start-IP-end* (e.g., *What* $\_\__{what}$ *happened?*)
     10.3% low-SES, 12.8% high-SES

For syntactic trigrams, which serve as the building blocks of *wh*-dependencies under the Pearl & Sprouse learning strategy, our corpus analysis revealed 21 trigram types in the low-SES input, 14 of which also appeared in the high-SES corpora analyses of Pearl & Sprouse (2013*b*).[6] Additionally, the low-SES input contained 1 syntactic trigram not found in the high-SES input, which comes from one of the dependencies found only in the low-SES input:

- CP_{that}*-IP-end*
  (from *What do you think [that* $\_\__{what}$*)] happens*?)

Notably, just as with the *wh*-dependency analysis, the most frequent syntactic trigrams are very similar across SES. The three trigram types that account for the majority of the trigrams (85.0%) in the low-SES *wh*-dependency input are the same three that account for the majority of the trigrams (87.9%) in the high-SES *wh*-dependency input, and they occur in about the same proportions (shown in (13)). So, as with the *wh*-dependencies, this suggests a high-level qualitative similarity in the syntactic trigram input across SES, despite the individual syntactic trigram differences.

(13) Proportions of the three most frequent trigram types across SES
  a. 1^{st} most frequent: *start-IP-VP*
     41.4% low-SES, 41.8% high-SES
  b. 2^{nd} most frequent: *IP-VP-end*
     38.9% low-SES, 40.0% high-SES
  c. 3^{rd} most frequent: *start-IP-end*
     4.7% low-SES, 6.1% high-SES

6.3 Quantitative analysis

To more precisely quantify how similar the input distributions are for both *wh*-dependencies and syntactic trigrams across SES, we use the Jensen-Shannon divergence (**JSDiv**) (Endres & Schindelin 2003). JSDiv values range from 0 to 1, with 0 indicating identical distributions. That is, higher JSDiv values indicate greater divergence in the distributions, while values closer to 0 indicate distributions that are more similar. In this way, JSDiv analysis provides a way to quantify similarity between distributions; this makes it useful as a comparative measure, where different distributions are assessed for their relative similarity to each other.

With this in mind, we additionally use JSDiv to assess child-directed speech in comparison to adult-directed speech and text, in order to provide a comparison baseline for the similarity across input samples of both *wh*-dependencies and syntactic trigrams. In particular, we assess how similar the low-SES and high-SES CDS *wh*-dependency and trigram distributions are to those in high-SES adult-directed speech (**ADS**) and adult-directed text (**ADT**) samples from Pearl & Sprouse

[6] A more detailed description of the syntactic trigram distribution across SES is available in Appendix A.2.

(2013*b*). These adult-directed corpora are described in Table 2. This JSDiv analysis will reveal which factors impact *wh*-dependency and syntactic trigram distributions more: SES, whether the speech is directed at children or adults, or whether the input is speech-based vs. text-based.

Table 2: 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) samples.

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

Table 3: The nine *wh*-dependencies shared across all four corpora that are used in the JSDiv 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%

***Wh*-dependencies.** Figure 2 shows the results of the JSDiv analysis for *wh*-dependencies, calculated over the distribution of the 9 *wh*-dependencies (shown in Table 3) that these four corpora had in common. We see that low-SES CDS and high-SES CDS are the most similar in *wh*-dependency distribution (JS: 0.00445), and appear to be twice as similar as the next closest comparison, which is high-SES CDS vs. high-SES ADS (JS: 0.00948). This affirms a quantitative similarity across SES in child *wh*-dependency input, in terms of *wh*-dependency distribution. Moreover, these results highlight that CDS *across* SES is more similar than CDS vs. ADS *within* SES. That is, whether the speech is directed at children or adults matters more than whether speech is coming from a high-SES or low-SES population. We also note that these JSDivs 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 (JS: 0.03156) than it does from either high-SES CDS (JS: 0.00948) or low-SES CDS (JS: 0.01576).

Syntactic trigrams. Figure 3 shows the results of the JSDiv analysis for syntactic trigrams, calculated over the distribution of the 14 trigrams shown in Table 4 (see Table A2 in Appendix A.2 for the full list of trigrams) that these four corpora had in common across all *wh*-dependencies. These trigrams accounted for 99.5-99.8% of the total trigrams in these corpora. As with the analysis of the *wh*-dependencies, we see the same pattern emerge: (i) low-SES CDS is more similar to high-SES CDS (JSDiv: 0.00850) than any other input type, and (ii) all speech is more similar to other types of speech than to text (speech vs. speech: JSDiv=0.00850-0.02836; speech vs. text: JSDiv=0.07183-0.16279).

Figure 2: JSDiv analyses 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.

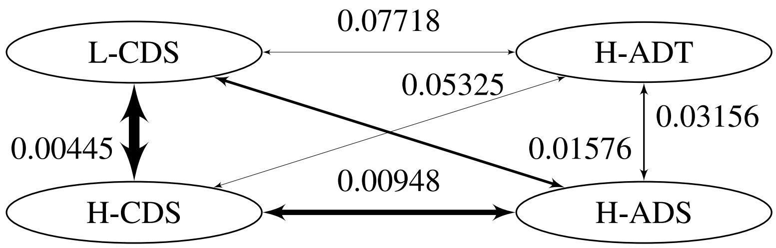
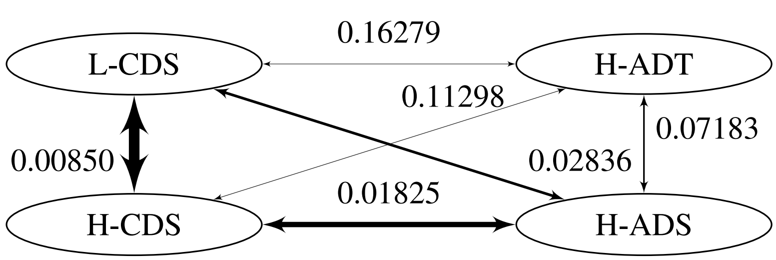


Table 4: Distribution of the 14 syntactic trigrams across child-directed Low-SES (L-CDS) and High-SES (H-CDS), as well as High-SES adult-directed speech (H-ADS) and text (H-ADT).

Syntactic trigrams	Syntactic trigram percentage
CP_{null}-IP-VP	0.1% - 0.7%
CP_{null}-IP-end	<0.1% - 0.3%
IP-VP-CP_{null}	0.3% - 0.7%
IP-VP-CP_{that}	<0.1%
IP-VP-IP	0.9% - 4.0%
IP-VP-NP	<0.1% - 0.1%
IP-VP-PP	0.8% - 2.5%
IP-VP-end	38.5% - 39.9%
VP-CP_{null}-IP	0.3 - 0.7%
VP-CP_{that}-IP	<0.1%
VP-IP-VP	0.9% - 4.0%
VP-PP-end	0.8% - 2.3%
start-IP-VP	38.6% - 41.7%
start-IP-end	4.7% - 19.0%

Figure 3: JSDiv analyses for low-SES CDS (L-CDS) trigrams, high-SES CDS (H-CDS) trigrams, high-SES adult-directed speech (H-ADS) trigrams, and high-SES adult-directed text (H-ADT) trigrams. Line thickness corresponds to similarity, with thicker lines indicating more similar distributions.



Quantitative analysis summary. Our quantitative analyses suggest that the input children encounter for learning about syntactic islands is very similar across SES. In particular, both the *wh*-distributions and the syntactic trigram distributions appear quite similar, despite some individual *wh*-dependency and trigram differences. However, it's unclear if even these comparatively small differences may lead to different qualitative outcomes. This is because even small input differences could be developmentally meaningful. So, using the low-SES input distributions, do we predict children's acquisition of syntactic island knowledge will be the same as what high-SES children are predicted to learn? To assess the qualitative similarity of high-SES and low-SES CDS input with respect to predicted learning outcomes, we need to use the low-SES input distribution to predict what low-SES children could learn about syntactic islands.

6.4 Developmental computational modeling: Predicting what low-SES children could learn about syntactic islands

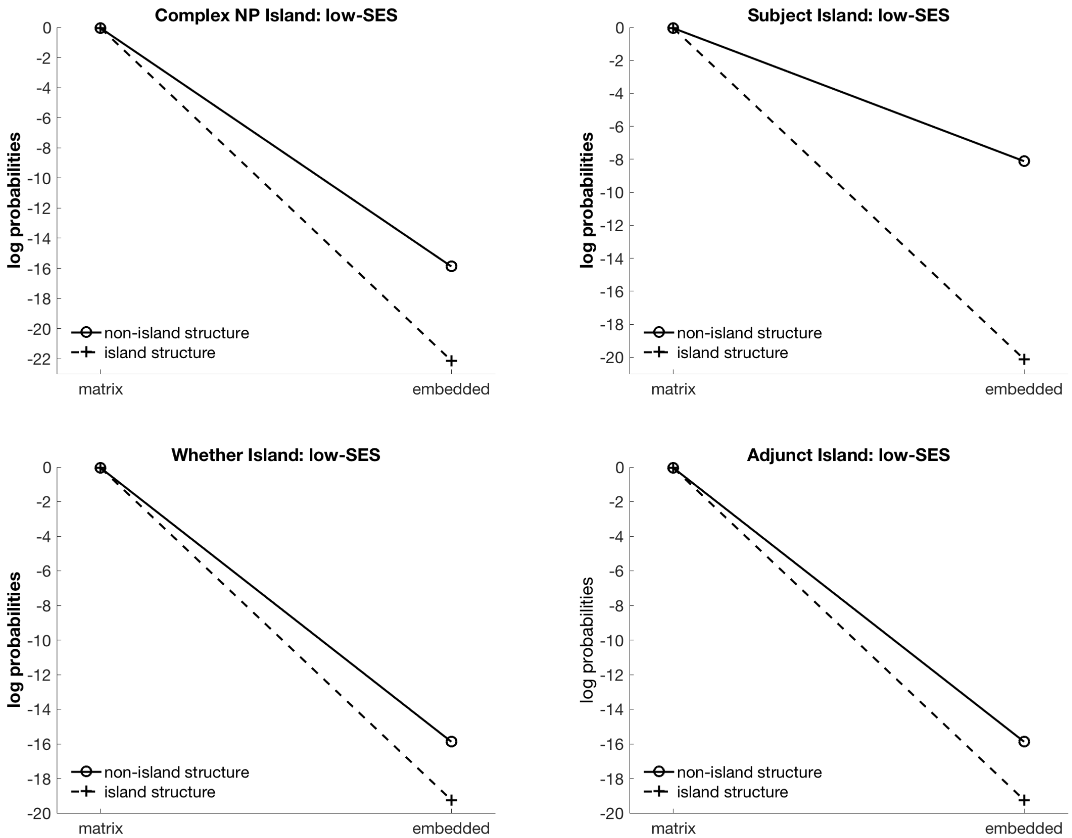
We use the same developmental computational learning model as Pearl & Sprouse (2013*b*); the modeled learner learns from the low-SES CDS input, encountering the same amount (≈200K) in the same distribution as low-SES children, and generates probabilities for the four sets of experimental stimuli of Sprouse et al. (2012), which correspond to Complex NP, Subject, Whether, and Adjunct islands. Recall that these experimental stimuli can be characterized by the syntactic paths shown in Table 1, where many of the grammatical dependencies are characterized by the same syntactic path (e.g., *start*-IP-*end* for both matrix+non-island and matrix+island); this is why Table 5, which shows the modeled learner's generated log probabilities of the relevant *wh*-dependencies, has only three grammatical dependency syntactic paths listed. Figure 4 shows the low-SES CDS log probabilities plotted on interaction plots for each of the four island types. To aid comparison of predicted learning outcomes across SES, Table 5 also shows the log probabilities generated by learners learning from the high-SES CDS, as well as the high-SES ADS and ADT reported in Pearl & Sprouse (2013*b*).

Table 5: 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. So, even the least acceptable grammatical dependency (with log probability -15.88) is predicted to be over 2000 times more acceptable than the most acceptable ungrammatical dependency (with log probability -19.25), because $\frac{10^{-15.88}}{10^{-19.25}} \approx 2344$. This is the same pattern which was found when learning from either high-SES child-directed or adult-directed input (high-SES grammatical: -0.93 to

Figure 4: Judgments derived from a modeled learner using low-SES CDS, demonstrating implicit knowledge of syntactic islands as indicated by superadditivity (which appears as non-parallel lines in these interaction plots).



-13.84; high-SES island-spanning: -18.46 to -20.36). Importantly, in Figure 4, 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 JSDiv analysis 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.

Additionally, although quantitative differences exist across ADS/ADT and CDS, these quantitative differences also don't impact predicted qualitative learning outcomes (i.e., they're not developmentally meaningful). For example, the low-SES CDS *wh*-dependency distribution is 17 times more similar to the high-SES CDS *wh*-dependency distribution than it is to high-SES ADT *wh*-dependency distribution, based on the JSDiv. Yet, all four *wh*-dependency distributions (low- and high-SES CDS, high-SES ADS, and high-SES ADT) contain the necessary information for the developmental computational model to learn syntactic island constraints. This indicates that even larger JSDiv differences lead to the same predicted learning outcomes, which in turn suggests that learning syntactic island knowledge may be fairly robust to input variation.

7 Discussion

Our results suggest that the *wh*-dependency input, and in turn the syntactic trigram input, that low-SES children receive is quantitatively and qualitatively similar to the input of high-SES children. In particular, any input differences across SES aren't predicted to be developmentally meaningful with respect to learning syntactic island knowledge. That is, our developmental computational modeling results serve as predictions of children's learning behavior for syntactic islands, and predict no learning outcome differences due to input differences across SES.

Interestingly, there's a striking difference in the exact *wh*-dependency distribution across SES that's predicted 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 (11), the only distinction between certain dependencies judged grammatical and certain dependencies judged ungrammatical by high-SES adults is the complementizer. Example (11) 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* (14a) and ungrammatical dependencies with complementizers like *whether* (whether islands) or *if* (adjunct islands) (14b). Again, the only difference in the syntactic path of these dependencies is the CP building block, which is CP_{that} for the dependency judged grammatical and $CP_{whether}$ or CP_{if} for the dependencies judged ungrammatical.

(14) 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 we used here can leverage the CP_{that} building block to predict that (14a) should be judged as better than (14b). However, dependencies involving CP_{that} are actually fairly rare in naturalistic usage. Pearl & Sprouse (2013*b*) only found 2 of 20,923 (0.0096%) in high-SES CDS (along with 7 of 8,508 (0.082%) in high-SES ADS and 2 of 4,230 (0.048%) in high-SES ADT). For high-SES children, this would correspond to approximately one *wh*-dependency with CP_{that} every two months.[7]

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 (14a). 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 judged ungrammatical by high-SES adults: *start-IP-VP-*CP_{that}*-IP-end* instances like (15). For low-SES children, this would correspond to approximately one *wh*-dependency with CP_{that} every 0.35 months, or approximately 5-6 *wh*-dependencies with CP_{that} every two months (notably more frequent than what high-SES children would encounter).[8]

(15) What do you think that $\_\__{what}$ happens?
What do [$_{IP}$ you [$_{VP}$ think [$_{CP_{that}}$ that [$_{IP}$ $\_\__{what}$ [$_{VP}$ happens]]]]]?

[7]With an estimated learning period of 200K *wh*-dependencies over 3 years (36 months) from Pearl & Sprouse (2013*b*), this can be calculated as 200K *wh*-dependencies * .000096 CP_{that} *wh*-dependency rate = 19.2 CP_{that} *wh*-dependencies over 3 years (36 months). $\frac{19.2}{36}$ = 0.53 per month or 1 approximately every two months.

[8]With the same learning period of 200K *wh*-dependencies over 3 years (36 months) from Pearl & Sprouse (2013*b*), this can be calculated as 200K *wh*-dependencies * .00051 CP_{that} *wh*-dependency rate = 102 CP_{that} *wh*-dependencies over 3 years (36 months). $\frac{102}{36}$ = 2.83 per month or approximately 2-3 every month (or 5-6 every two months) or 1 every 0.35 months.

syntactic path: *start-IP-VP-CP_{that}-IP*

So, the presence of this *wh*-dependency type, which is ungrammatical in the high-SES dialect, is predicted to provide 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 be developmentally meaningful across SES, in contrast with lexical or more foundational syntactic knowledge. That is, there may not be a "complex syntax gap" across SES. For instance, a difference in input quantity across SES, as indicated by the relative rate of *wh*-dependencies, doesn't appear to be a meaningful one, as shown by our quantitative analyses using JSDiv. Moreover, there doesn't appear to be any difference with respect to input quality, based on our developmental computational modeling results. So, while there may be some surface-level quantitative differences in input across SES, there don't appear to be qualitative differences. That is, surface input differences mask deeper input similarities across SES for the development of this syntactic island knowledge.

For syntactic islands, we would predict 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. That is, the target knowledge low-SES children eventually achieve for these syntactic islands is predicted to be the same as that of high-SES children (and adults), even if the low-SES target knowledge may differ for other syntactic knowledge like *that*-trace violations. Current data on the development of other *wh*-dependency knowledge in children across SES suggests that by age four, low-SES children do indeed appear to have similar knowledge to high-SES children (de Villiers et al. 2008); however, the four syntactic islands we examined here have not yet been tested.

Another step for evaluating the model's predictions is to collect judgment data from low-SES adults for these syntactic islands and for *that*-trace violations specifically. We would expect low-SES adult judgments to be the same as high-SES adult judgments, except for the *that*-trace violation, which low-SES adults should find grammatical.

We note that the ability to leverage the *wh*-dependency and syntactic trigram 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 & Fernald 2013). These delays could lead to low-SES children being less able to harness the complex syntactic information available in their input, even if the information is in fact there. 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 (including accurately identifying where the *wh*-word is understood). 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 isn't predicted to cause them to lag behind their high-SES counterparts when it comes to learning about complex syntactic knowledge like syntactic islands.

8 Conclusion

We have aimed to provide a new way for identifying developmentally-meaningful input differences, harnessing developmental computational modeling. Developmental computational modeling can be used to assess input quality by predicting what children should be able to learn from their input. If input variation is developmentally meaningful, then the model predicts learning outcome differences; in contrast, the model predicts similar learning outcomes when the input variation isn't developmentally meaningful. To demonstrate this technique, we applied it to input variation across SES concerning the development of syntactic island knowledge; our model predicted that there were no developmentally-meaningful input differences. So, input quality for syntactic islands is predicted to be the same across SES. This result broadens the body of research on linguistic input variation across SES to include the nature of the input for more complex syntactic knowledge. To our knowledge, this is the first comparison across SES for these syntactic islands. Our results suggest that if we do see developmental differences in syntactic island knowledge across SES, it's not because of the information available in the input. Instead, children's ability to harness that information may differ. In short, the syntactic islands information is predicted to be there for children to use, no matter their SES – a key developmental step is for them to figure out how to use it.

References

Blum, S. (2015), 'Wordism": Is there a teacher in the house', *Journal of Linguistic Anthropology* **25**(1), 74–75.

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

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

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

Cowart, W. (1997), *Experimental Syntax: Applying Objective Methods to Sentence Judgements*, Thousand Oaks, CA: Sage.

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.

Dickinson, D. K. & Tabors, P. O. (2001), *Beginning literacy with language: Young children learning at home and school.*, Paul H Brookes Publishing.

Endres, D. M. & Schindelin, J. E. (2003), 'A new metric for probability distributions', *IEEE Transactions on Information theory* .

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

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

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

Huttenlocher, J., Vasilyeva, M., Cymerman, E. & Levine, S. (2002), 'Language input and child syntax', *Cognitive Psychology* **45**(3), 337–374.

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J. & Hedges, L. V. (2010), 'Sources of variability in children's language growth', *Cognitive Psychology* **61**(4), 343–365.

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

Pearl, L. (in press), Modeling syntactic acquisition, *in* J. Sprouse, ed., 'Oxford Handbook of Experimental Syntax', Oxford University Press.

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

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

Pearl, L. & Sprouse, J. (2015), 'Computational modeling for language acquisition: A tutorial with syntactic islands', *Journal of Speech, Language, and Hearing Research* **58**, 740–753.

Rombough, K. & Thornton, R. (2018), 'Probing morphological, syntactic and pragmatic knowledge through answers to wh-questions in children with SLI', *International Journal of Speech-Language Pathology* **20**(2), 284–294.

Ross, J. (1967), Constraints on variables in syntax, PhD thesis, MIT, Cambridge, MA.

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

Rowe, M. L., Leech, K. A. & Cabrera, N. (2017), 'Going beyond input quantity: Wh-questions matter for toddlers' language and cognitive development', *Cognitive Science* **41**, 162–179.

Schwab, J. F. & Lew-Williams, C. (2016), 'Language learning, socioeconomic status, and child-directed speech', *Wiley Interdisciplinary Reviews: Cognitive Science* **7**, 264—275.

Sperry, D. E., Sperry, L. L. & Miller, P. J. (2018), 'Reexamining the verbal environments of children from different socioeconomic backgrounds', *Child development* .

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.

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

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

Valian, V. & Casey, L. (2003), 'Young children's acquisition of wh-questions: The role of structured input', *Journal of child language* **30**(1), 117–143.

Weisleder, A. & Fernald, A. (2013), 'Talking to children matters early language experience strengthens processing and builds vocabulary', *Psychological Science* **24**(11), 2143–2152.

A Appendices

A.1 *Wh*-dependency distribution across SES

Table A1 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 A1: 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 syntactic path is given below the path. Dependencies used in the Jensen-Shannon divergence (JSDiv) analysis are in teal. The dependency in the Low-SES dialect that's judged to be 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
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

Distribution of *wh*-dependencies in the input				
Syntactic path and example utterance	L-CDS	H-CDS	H-ADS	H-ADT
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
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%

Distribution of *wh*-dependencies in the input				
Syntactic path and example utterance	L-CDS	H-CDS	H-ADS	H-ADT
What do you put it on top of?	0	2	0	0
IP-VP-PP-NP-PP-IP-VP What is she in the habit of doing?	0.0% 0	0.0% 0	<0.1% 1	0.0% 0
IP-VP-PP-PP What does he eat out of?	0.5% 1	0.1% 22	0.0% 0	0.0% 0
IP-VP-PP-IP-VP What did he think about stealing?	0.0% 0	<0.1% 1	0.0% 0	0.0% 0

A.2 Syntactic trigram distribution across SES

Table A2 shows the distribution of the syntactic trigrams 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. The shared syntactic trigrams were used when calculating the Jensen-Shannon divergence (JSDiv) analyses.

Table A2: Distribution of the syntactic trigrams across 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). The 14 shared trigrams used in the JSDiv analysis are in teal.

Distribution of trigrams in the input				
Trigrams	L-CDS	H-CDS	H-ADS	H-ADT
AdjP-IP-VP	0.0% 0	0.0% 0	<0.1% 2	<0.1% 5
AdjP-PP-end	0.0% 0	0.0% 0	<0.1% 1	<0.1% 1
CP_{for}-IP-VP	0.0% 0	<0.1% 1	0.0% 0	0.0% 0
CP_{null}-IP-VP	0.6% 49	0.7% 298	0.2% 44	0.1% 10
CP_{null}-IP-end	<0.1% 5	<0.1% 24	0.3% 53	0.2% 12
CP_{that}-IP-VP	0.0% 0	<0.1% 2	<0.1% 7	<0.1% 2
CP_{that}-IP-end	<0.1% 2	0.0% 0	0.0% 0	0.0% 0
IP-VP-AdjP	0.0% 0	0.0% 0	<0.1% 3	<0.1% 6
IP-VP-CP_{for}	0.0% 0	<0.1% 1	0.0% 0	0.0% 0
IP-VP-CP_{null}	0.6% 54	0.7% 321	0.6% 96	0.3% 22
IP-VP-CP_{that}	<0.1% 2	<0.1% 2	<0.1% 7	<0.1% 2
IP-VP-IP	4.0% 340	3.2% 1398	2.1% 353	0.9% 65
IP-VP-NP	<0.1%	0.1%	0.1%	0.1%

Distribution of trigrams in the input				
Trigrams	L-CDS	H-CDS	H-ADS	H-ADT
	4	67	23	9
IP-VP-PP	2.4% 202	1.6% 698	2.5% 423	0.8% 63
IP-VP-end	38.9% 3292	39.9% 17487	38.5% 6553	37.4% 2753
NP-IP-VP	0.0% 0	0.0% 0	<0.1% 1	<0.1% 3
NP-PP-IP	0.0% 0	0.0% 0	<0.1% 1	0.0% 0
NP-PP-end	<0.1% 4	<0.1% 10	<0.1% 7	0.0% 0
PP-CP$_{null}$-IP	0.0% 0	<0.1% 1	<0.1% 1	0.0% 0
PP-IP-VP	<0.1% 3	<0.1% 1	<0.1% 4	0.0% 0
PP-NP-PP	0.0% 0	<0.1% 2	<0.1% 1	0.0% 0
PP-NP-end	0.0% 0	0.0% 0	<0.1% 2	0.0% 0
PP-PP-end	<0.1% 1	<0.1% 23	0.0% 0	<0.1% 1
VP-AdjP-IP	0.0% 0	0.0% 0	<0.1% 2	<0.1% 5
VP-AdjP-PP	0.0% 0	0.0% 0	<0.1% 1	<0.1% 1
VP-CP$_{for}$-IP	0.0% 0	<0.1% 1	0.0% 0	0.0% 0
VP-CP$_{null}$-IP	0.6% 54	0.7% 321	0.6% 96	0.3% 22
VP-CP$_{that}$-IP	<0.1% 2	<0.1 2	<0.1% 7	<0.1 2
VP-IP-VP	4.0% 340	3.2% 1389	2.1% 351	0.9% 65
VP-IP-end	0.0% 0	<0.1% 9	<0.1% 2	0.0% 0
VP-NP-IP	0.0% 0	0.0% 0	<0.1% 1	<0.1% 3
VP-NP-PP	<0.1 4	<0.1% 8	<0.1 7	0.0% 0
VP-NP-end	0.0% 0	0.1% 59	<0.1% 15	<0.1% 6
VP-PP-CP$_{null}$	0.0% 0	<0.1% 1	<0.1% 1	0.0% 0

Distribution of trigrams in the input				
Trigrams	L-CDS	H-CDS	H-ADS	H-ADT
VP-PP-IP	<0.1 3	<0.1% 1	<0.1 3	0.0% 0
VP-PP-NP	0.0% 0	<0.1% 2	<0.1% 3	0.0% 0
VP-PP-PP	<0.1 1	<0.1% 23	0.0% 0	<0.1 1
VP-PP-end	2.3% 198	1.5% 671	2.4% 416	0.8% 62
start-IP-VP	41.4% 3502	41.7% 18283	41.5% 7049	38.6% 2835
start-IP-end	4.7% 402	6.1% 2680	8.6% 1464	19.0% 1396