Syntactic Island and Learning Biases:

Combining Experimental Syntax and Computational Modeling to Investigate the Language Acquisition Problem

Lisa Pearl & Jon Sprouse

Outline of the Paper:

1. Intro: Framework of *Types* of Learning Biases
2. Background on Syntactic Islands
3. Creation/Analysis of Corpora with Syntax Annotations
4. A statistical algorithm for learning Syntactic Islands
5. Performance of the Algorithm
6. Rationales and Revelations of Algorithm's Biases
7. Impact of research on Syntactic Theory

Intro: Features of Learning Biases

1. Are they **domain-specific** or **domain-general**?
2. Are they **innate** or **derived** from prior experience?
3. Are they a constraint on the **hypothesis space** or a constraint on the **learning mechanism**?

	Domain-specificity	Innateness	Scope of constraint
Universal Grammar	Language-specific	Innate	Often hypothesis space, could be learning mechanism

Syntactic Island Effects

English *Wh*-word dependencies

- No adjacency between verbs and the pronoun (NP) the verb acts on

  Does Jack think that? vs.

  What does Jack think ___?
- However, there is still a syntactic/semantic dependency between the verb and the pronoun

- These dependencies are stable over long distances

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

Syntactic Island Effects

English *Wh*-word dependencies

- Syntactic rules do not prohibit dependencies from spanning long distances
- But *wh*- dependencies are prohibited from appearing in some syntactic constructions

Do you think the joke about <u>that</u> offended Jane? vs.

*<u>What</u> do you think the joke about ___ offended Jane?

Syntactic Island Effects

Syntactic constructions that disrupt dependencies are called *syntactic islands*.

- Other constructions in English that encounter island effects are
  - Relative-clause formation
  - Topicalization
  - Adjective-through constructions

Syntactic Island Effects

Sprouse, Wagers & Phillips (2012a) collected real adult acceptability judgments of sentences constructed with syntactic islands

- Four island types investigated
- Semantically intelligible but syntactically ungrammatical
- Ratings on a magnitude scale instead of binary judgments

- Most importantly: 2x2 design of sentences for each island type highlights if there is an island effect

Syntactic Island Effects

Sprouse, Wagers & Phillips (2012a)

(8) Whether islands

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

Syntactic Island Effects

Sprouse, Wagers & Phillips (2012a)

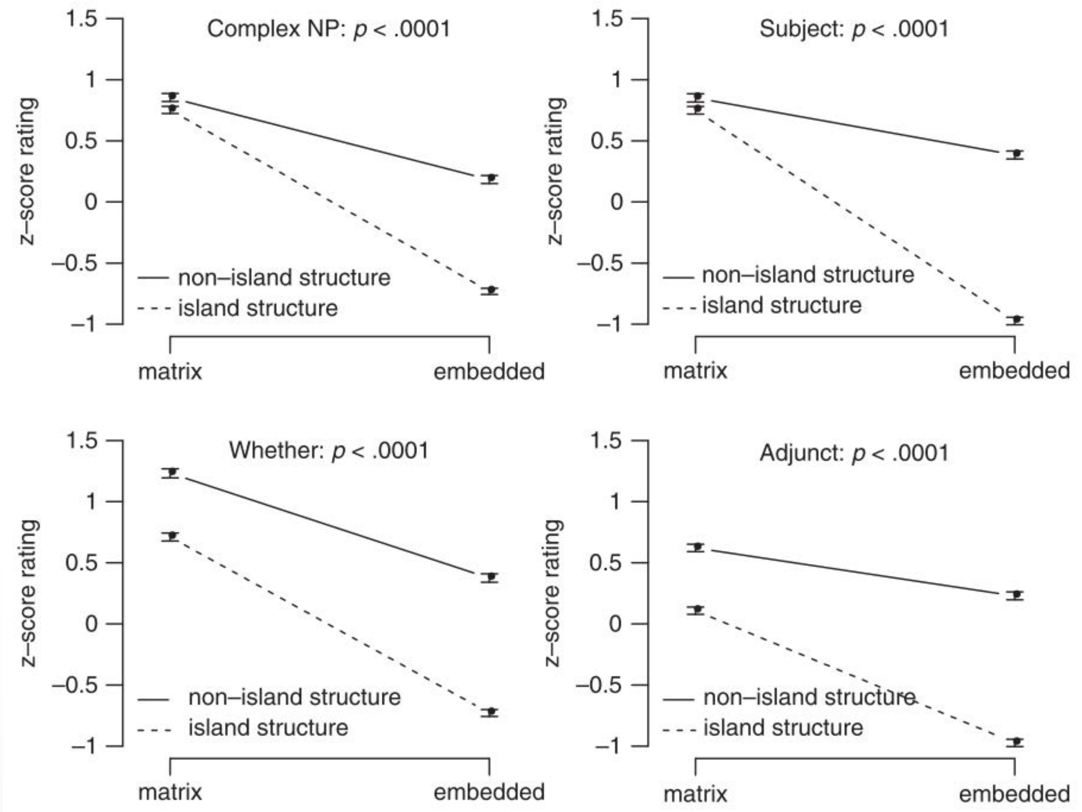

FIGURE 2 Experimentally derived acceptability judgments for the four island types from Sprouse, Wagers & Phillips (2012a) ($N = 173$).

Annotated Corpora Creation

Needed structurally annotated corpora to assess the frequency of the island structures in speech to children.

- 5 well known corpora
- Child-directed speech for children 1-5 years old
- Utterances with Sprouse (2012) type dependencies were quite rare

TABLE 1
The Corpus Analysis of the Child-Directed Speech Samples

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

Note. These are the child-directed speech samples from CHILDES, given the experimental stimuli used in Sprouse, Wagers & Phillips et al. (2012a) for the four island types examined. The syntactic island condition (which is Ungrammatical) is bolded.

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

Learning Algorithm

Pearl & Sprouse (2013) designed a (non-Bayesian) statistical learning algorithm to work over this type of structural input and attempt to learn that syntactic islands are ungrammatical.

1. Type and quantity of learning input for a child
2. The structure of the hypothesis space
3. The learning process

Algorithm: Input

The algorithm must have access to phrase structure information to track the dependency of a *Wh-* sentence.

Therefore, learning must take place at the same level of abstraction.

The algorithm works over the **container nodes** of a sentence.

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

b. IP VP CP IP VP PP

c. IP-VP-CP-IP-VP-PP

Algorithm: Input

The possible container nodes are specified.

- Eg. NP, VP, IP, CP

However, to differentiate between certain kinds of sentences used in Sprouse (where humans rate different grammaticality judgments) CP must be further subdivided

- Eg Cp_{that} vs $CP_{whether}$

Algorithm: Hypothesis Space

The algorithm operates in a hypothesis space that is made of sequences of trigrams of container nodes

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

b. IP VP CP_{null} IP VP PP

c. start-IP-VP-CP_{null}-IP-VP-PP-end =

start-IP-VP

IP-VP-CP_{null}

VP-CP_{null}-IP

CP_{null}-IP-VP

IP-VP-PP

VP-PP-end

Algorithm: Learning Process

The learner parses all the available data into sequences of trigrams.

Each trigram has a probability

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

Algorithm: Judgments

To calculate a grammaticality judgment, the algorithm simply multiplies the probability of each container node trigram in an utterance together.

(14) "Where does the reporter think Jack stole from?"

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

IP VP CP_{null} IP VP PP

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

Trigrams: start-IP-VP

IP-VP-CP_{null}

VP-CP_{null}-IP

CP_{null}-IP-VP

IP-VP-PP

VP-PP-end

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

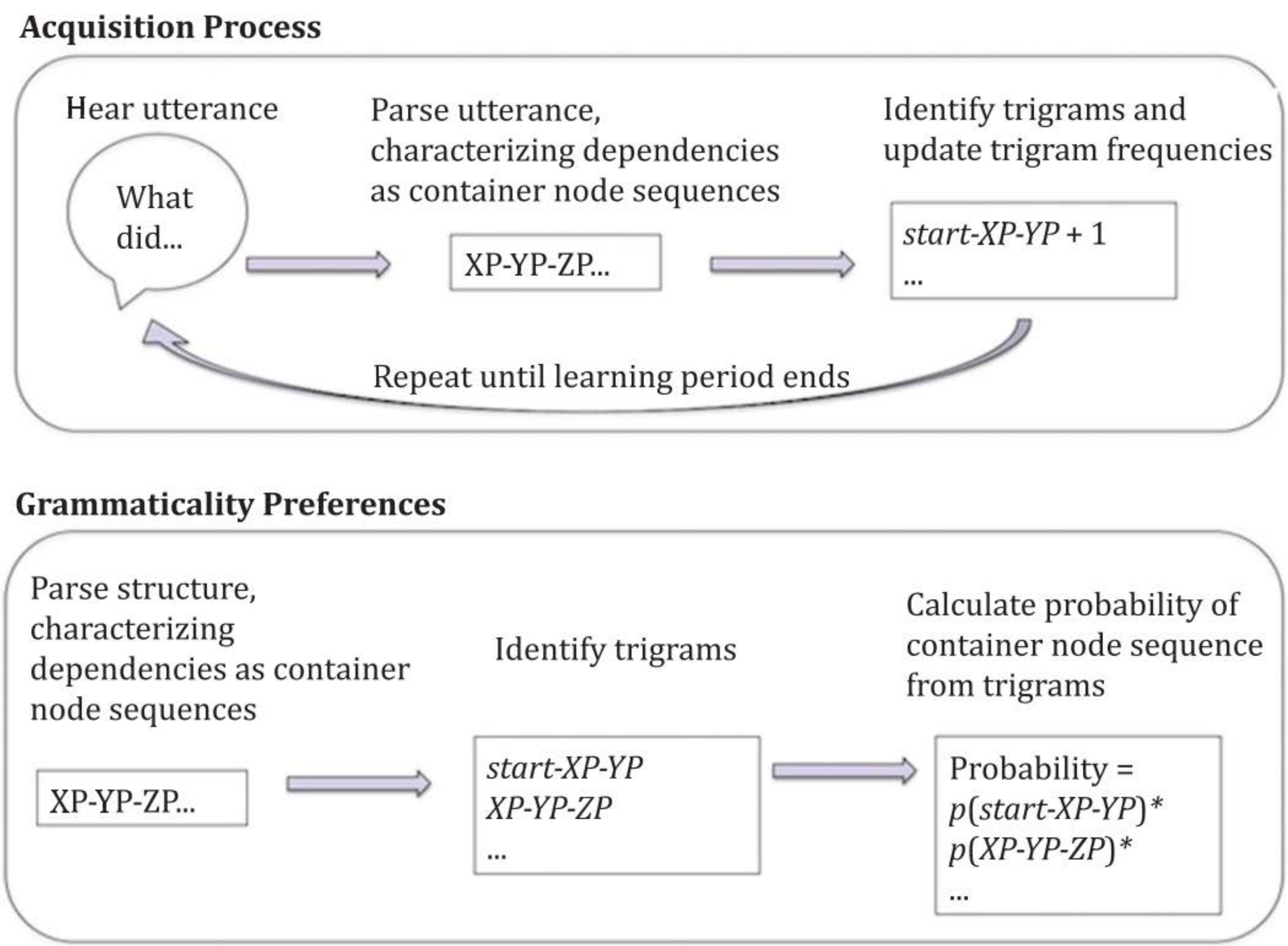

FIGURE 3 Steps in the acquisition process and calculation of grammaticality preferences (color figure available online).

Corpora Results

TABLE 2
Basic Composition of the Child-Directed and Adult-Directed Input Corpora

	Child-Directed: Speech	*Adult-Directed:Speech*	*Adult-Directed:Text*
Total Utterances	101838	74576	24243
Total *wh*-Dependencies	20923	8508	4230

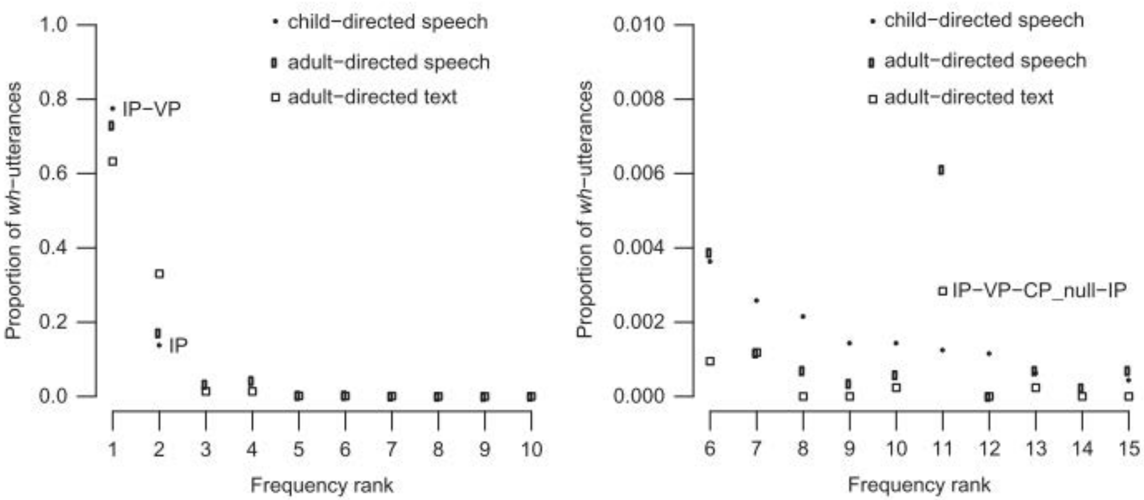

FIGURE 4 The 15 most frequent *wh*-dependency types in the three corpora types. The left panel displays the 10 most frequent *wh*-dependency types for each of the three corpora types, with IP-VP and IP dominating all three corpora types (IP-VP: rank 1, IP: rank 2). The right panel displays the 6th–15th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

Learner Probability Results

TABLE 3
Inferred Grammaticality of Different *Wh*-Dependencies from Sprouse, Wagers & Phillips (2012a), Represented with Log Probability

		Child-Directed Speech	*Adult-Directed Speech & Text*
Grammatical Dependencies			
Matrix Subject	IP	−1.21	−0.93
Embedded Subject	IP-VP-CP_{null}-IP	−7.89	−7.67
Embedded Object	IP-VP-CP_{that}-IP-VP	−13.84	−11.00
Island-Spanning Dependencies			
Complex NP	IP-VP-NP-CP_{that}-IP-VP	−19.81	−18.93
Subject	IP-VP-CP_{null}-IP-NP-PP	−20.17	−20.36
Whether	IP-VP-$CP_{whether}$-IP-VP	−18.54	−18.46
Adjunct	IP-VP-CP_{if}-IP-VP	−18.54	−18.46

Child Directed Speech Results:

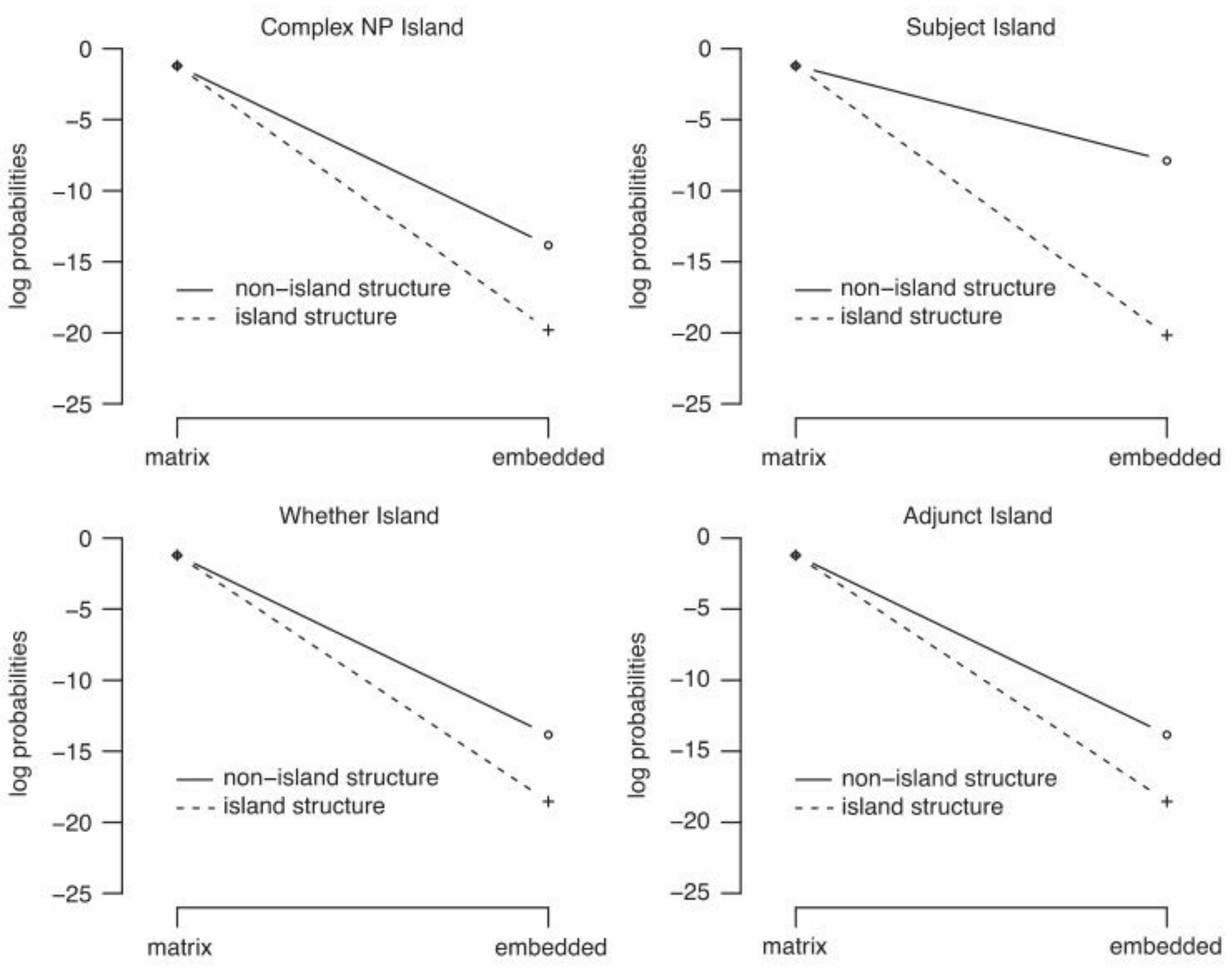

FIGURE 5 Log probabilities derived from a learner using child-directed speech.

Adult Directed Results:

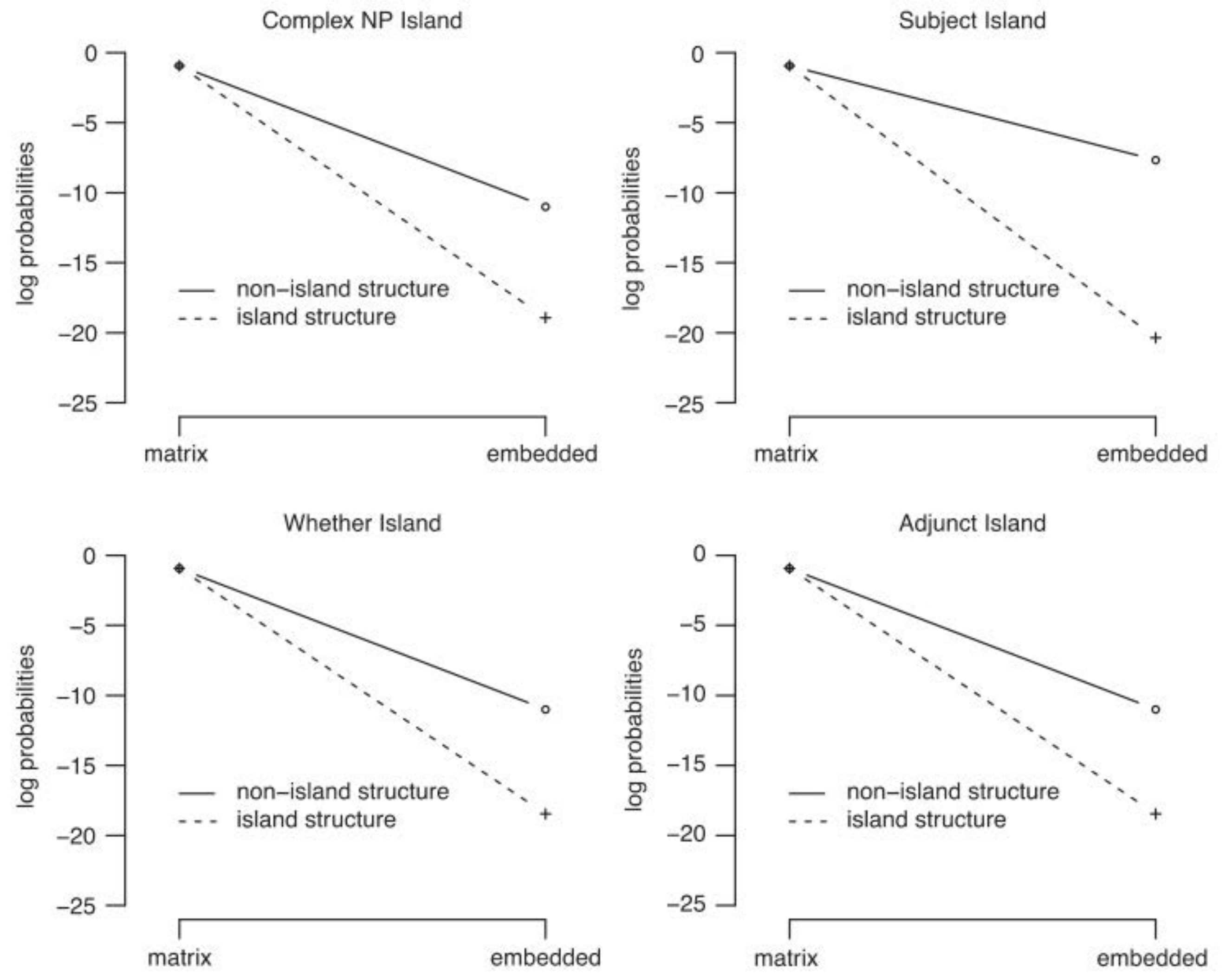

FIGURE 6 Log probabilities derived from a learner using adult-directed speech and text.

Relation to Syntactic Theory

The paper discusses a few scenarios that this algorithm cannot account for

These scenarios primarily arise due to the use of container node trigrams without further specificity

- Parasitic gaps & Across-the-Board
- Italian *Wh*- question dependencies
- Compartmentalizer *That* behavior

Relation to Syntactic Theory

In summary:

Everything is an empirical question