Computational cognitive modeling: How to investigate child language acquisition using math

Lisa Pearl
University of California, Irvine

April 26, 2024
Linguistics Student Association
CSU Fullerton
About child language acquisition
child language acquisition: extraction of language information by young children
One question:
What kind of language information exactly?
Some examples of target language information
child language acquisition

how to identify words in fluent speech (speech segmentation)

what a pretty kitty!
child language acquisition

how to pronounce words (phonology)

✔ Kitty
❌ ki TTY
child language acquisition

what a pretty kitty!

speech segmentation

phonology

what a pretty ___!

penguin

kitty

owl

certain words behave like other words (syntactic categorization)

Noun

Countable
child language acquisition

speech segmentation

how to interpret words in context (syntax, semantics)

“Oh look — a kitty!”
“He’s such a pretty kitty!”

“Look — there’s another one!”
how to put words together to ask questions (syntax)

*This kitty was bought as a present for someone.*

*Lily thinks this kitty is pretty.*

“Who does Lily think the kitty for is pretty?”
child language acquisition

“Oh look — a pretty kitty!”
“He’s such a pretty kitty!”
“Look — there’s another one!”

syntax, semantics

“Who does Lily think the kitty for is pretty?”

syntax

how to identify the right interpretation in context (pragmastics)

“I think I saw all the kitties on the stairs.”

“No - every kitty didn’t sit on the stairs”
child language acquisition

speech segmentation

non-canonical phonology

speech segmentation

Noun

Noun

phonology

penguin

owl

syntactic categorization

syntax, semantics

syntax

how to identify the right interpretation in context (pragmatics)

“Who does Lily think the kitty for is pretty?”

“Look — there’s another one!”

“No - every kitty didn’t sit on the stairs”

Not all kitties sat on the stairs.

Not all kitties sat on the stairs.
child language acquisition

"Oh look — a pretty kitty!"
"He's such a pretty kitty!"
"Look — there's another one!"

"Who does Lily think the kitty for is pretty?"

This is just some of the target knowledge for children's acquisition. (There's much more.)

"Every kitty didn't sit on the stairs"

Not all kitties sat on the stairs.
child language acquisition

“Oh look — a pretty kitty!”
“He’s such a pretty kitty!”
“Look — there’s another one!”

“Who does Lily think the kitty for is pretty?”

Another important question: How does this magical acquisition process occur?

“Every kitty didn’t sit on the stairs”
Not all kitties sat on the stairs.
Children are amazing at learning language.
Much of the linguistic system is already known by age 4.
Also, children figure language out mostly without explicit instruction.
What they’re doing: Extracting patterns and making generalizations from the surrounding data mostly just by hearing examples of what’s allowed in the language.
This is pretty magical. But how does it work?
I primarily use quantitative methods like computational cognitive modeling to try to figure this out.
Computational cognitive modeling lets us explore theoretical ideas precisely, and evaluate how well any particular theory can explain empirical data on children’s language acquisition.

Math is at the heart of this tool.

Pearl, 2010, 2014, Pearl and Goldwater, 2016, Pearl, 2017, 2019, 2020, 2021,a,b, 2023a,b, under review
computational cognitive modeling

One main part: Counting things

Pearl, 2010, 2014, Pearl and Goldwater, 2016, Pearl, 2017, 2019, 2020, 2021,a,b, 2023a,b, under review
One main part: *Counting things* (sometimes we count a lot of things)
computational cognitive modeling

Another part: *principled reasoning* based on those counts

Pearl, 2010, 2014, Pearl and Goldwater, 2016, Pearl, 2017, 2019, 2020, 2021,a,b, 2023a,b, under review
computational cognitive modeling

Another part: principled reasoning based on those counts

Bayesian inference

\[ p(\text{Generalization}|\text{Data}) \propto p(\text{Generalization}) \cdot p(\text{Data}|\text{Generalization}) \]

Pearl, 2010, 2014, Pearl and Goldwater, 2016, Pearl, 2017, 2019, 2020, 2021,a,b, 2023a,b, under review
But what do we count and reason over? How do we connect that information to language acquisition?
We typically use computational cognitive modeling to encode a child’s acquisition process very precisely.
We think the child is learning by counting different parts of her input and reasoning over those counts in a sensible way.

So, the modeled learner will count those same things and learn about language by doing principled reasoning over those counts.
Let’s see how this works for a type of syntactic knowledge known as “syntactic islands”.
Syntactic islands involve wh-dependencies.

This kitty was bought as a present for someone.

Lily thinks this kitty is pretty.

Who does Lily think the kitty for is pretty?

What does Lily think is pretty, and who does she think it’s for?
Syntactic islands
involve *wh*-dependencies.

What's going on here?

There's a dependency between the *wh*-word *who* and where it's understood (the gap)

*Who does Lily think the kitty for* who *is pretty?*
Syntactic islands involve *wh*-dependencies.

What's going on here?

There's a dependency between the *wh*-word *who* and where it's understood (the gap)

*Who does Lily think the kitty for ___ who is pretty?*

This dependency is strongly dispreferred in English.
Syntactic islands involve *wh*-dependencies.

What's going on here?

There's a dependency between the *wh*-word *who* and where it's understood (the gap)

*Who does Lily think the kitty for* [understood as who] *is pretty?*

This dependency is **strongly dispreferred** in English.

One explanation: The dependency crosses a “syntactic island” (Ross 1967)
Syntactic islands

Who does Lily think the kitty for __who is pretty? Subject island

Ross 1967
Syntactic islands

Who does Lily think the kitty for \_who\_ is pretty?  

Ross 1967

Jack is somewhat tricksy.

He claimed he bought something.

What did Jack make the claim that he bought \_what\_?
Who does Lily think the kitty for __who is pretty? [Subject island]

What did Jack make the claim that he bought __what? [Complex NP island]

Jack is somewhat tricksy.

He claimed he bought something.

Elizabeth wondered if he actually did and what it was.

What did Elizabeth wonder whether Jack bought __what?
Who does Lily think the kitty for \_\_who is pretty? \underline{Subject island}

What did Jack make the claim that he bought \_\_what? \underline{Complex NP island}

What did Elizabeth wonder whether Jack bought \_\_what? \underline{Whether island}

Jack is somewhat tricksy.

He claimed he bought something.

Elizabeth worried it was something dangerous.

What did Elizabeth worry if Jack bought \_\_what?
Syntactic islands

*Who* does Lily think the kitty for *who* is pretty?  **Subject island**
*What* did Jack make the claim that he bought *what*?  **Complex NP island**
*What* did Elizabeth wonder whether Jack bought *what*?  **Whether island**
*What* did Elizabeth worry if Jack bought *what*?  **Adjunct island**

Important: It’s not about the length of the dependency.

(Chomsky 1965, Ross 1967)
Who does Lily think the kitty for __who is pretty?  [Subject island]

What did Jack make the claim that he bought __what?  [Complex NP island]

What did Elizabeth wonder whether Jack bought __what?  [Whether island]

What did Elizabeth worry if Jack bought __what?  [Adjunct island]

Important: It’s not about the length of the dependency.  (Chomsky 1965, Ross 1967)

What did Elizabeth think __what?
Who does Lily think the kitty for __who is pretty?  
What did Jack make the claim that he bought __what?  
What did Elizabeth wonder whether Jack bought __what?  
What did Elizabeth worry if Jack bought __what?  

Important: It’s not about the length of the dependency.  
(Chomsky 1965, Ross 1967)

What did Elizabeth think Jack said __what?
Syntactic islands

Who does Lily think the kitty for whom is pretty? [Subject island]
What did Jack make the claim that he bought what? [Complex NP island]
What did Elizabeth wonder whether Jack bought what? [Whether island]
What did Elizabeth worry if Jack bought what? [Adjunct island]

Important: It’s not about the length of the dependency. (Chomsky 1965, Ross 1967)

What did Elizabeth think Jack said Lily saw what?
Who does Lily think the kitty for __who is pretty?  Subject island
What did Jack make the claim that he bought __what?  Complex NP island
What did Elizabeth wonder whether Jack bought __what?  Whether island
What did Elizabeth worry if Jack bought __what?  Adjunct island

English adults judge these island-crossing dependencies to be far less acceptable than many others, including others that are very similar except that they don’t cross syntactic islands (Sprouse et al. 2012).
English-learning children strongly disprefer one of these island-crossing dependencies compared to others (de Villiers et al. 2008).
Who does Lily think the kitty for __who__ is pretty? **Subject island**

What did Jack make the claim that he bought __what__? **Complex NP island**

What did Elizabeth wonder whether Jack bought __what__? **Whether island**

What did Elizabeth worry if Jack bought __what__? **Adjunct island**

Additional *wh*-dependency knowledge:
The *frequency* of a lexical item can also affect adult *acceptability judgments* of potential syntactic islands.

What did Elizabeth say that Jack saw __what__?

What did Elizabeth whine that Jack saw __what__?
Who does Lily think the kitty for whom is pretty?  
What did Jack make the claim that he bought what?  
What did Elizabeth wonder whether Jack bought what?  
What did Elizabeth worry if Jack bought what?

These judgments and (dis)preferences are a measurable observable behavior that can signal the successful acquisition of syntactic island knowledge.
Who does Lily think the kitty for __who is pretty?  
What did Jack make the claim that he bought __what? 
What did Elizabeth wonder whether Jack bought __what? 
What did Elizabeth worry if Jack bought __what? 

measurable observable behavior
Syntactic islands

Who does Lily think the kitty for who is pretty? \textcolor{red}{Subject island}

What did Jack make the claim that he bought what? \textcolor{red}{Complex NP island}

What did Elizabeth wonder whether Jack bought what? \textcolor{red}{Whether island}

What did Elizabeth worry if Jack bought what? \textcolor{red}{Adjunct island}

So, these judgments and (dis)preferences can serve as a target for successful acquisition — an outcome we can measure.
This is the **target** of acquisition. But how could a child **learn** this?
Let’s use a computational cognitive model to help figure this out.
What’s a learning theory the modeled learner could encode?
one learning theory
Pearl & Sprouse 2013
Dickson, Pearl, & Futrell 2022, 2024, in prep.

Learn the right building blocks
Learn the right building blocks

Pearl & Sprouse 2013
Dickson, Pearl, & Futrell 2022, 2024, in prep.

View *wh*-dependencies in terms of their **building blocks** and **track (count)** those building blocks in the **input**.
building blocks of *wh*-dependencies

Dependencies represented as a sequence of container nodes

What phrases contain the gap (but not the *wh*-word)?

```
CP
  NP₁
    What
  NP
    Pro
    you
  VP
    V
    see
  IP
    did
```
building blocks of *wh*-dependencies

Dependencies represented as a sequence of container nodes

What phrases contain the gap (but not the *wh*-word)?

What did you see __?  
= What did [\_{IP} you [\_{VP} see __]]?  
= start-IP-VP-end
building blocks of *wh*-dependencies

Dependencies represented as a sequence of container nodes

What phrases contain the gap (but not the *wh*-word)?

What did you see __?
= What did [IP you [VP see __]]?
= start-IP-VP-end

What __ happened?
= What [IP __ happened]?
= start-IP-end
building blocks of *wh*-dependencies

Dependencies represented as a sequence of container nodes

What phrases contain the gap (but not the *wh*-word)?

What did you see ___?
= What did [IP you [VP see ___]]?
= start-IP-VP-end

What __ happened?
= What [IP __ happened]?
= start-IP-end

What did she want to do ___?
= What did [IP she [VP want [IP to [VP do ___]]]]?
= start-IP-VP-IP-VP-end
building blocks of *wh*-dependencies

What _ __ happened?  
= What [IP __ happened]?  
= start-IP-end

What did she want to do _ ?  
= What did [IP she [VP want [IP to [VP do ___]]]]?  
= start-IP-VP-iP-VP-end

(Much) less acceptable dependencies have **low probability segments**

```
[CP Who did [IP Lily [VP think [CP-that [IP [NP the kitty [PP for ___ ] was pretty ?]]]]]
```

```
start-IP-VP-CP-that-IP-NP-PP-end
```
building blocks of *wh*-dependencies

What ___ happened?
= What [IP ___ happened]?
= start-IP-end

What did she want to do ___?
= What did [IP she [VP want [IP to [VP do ___]]]]?
= start-IP-VP-IP-VP-end

What did you see ___?
= What did [IP you [VP see ___]]?
= start-IP-VP-end

(Much) less acceptable dependencies have low probability segments

[CP Who ___ did [IP Lily [VP think [CP that [IP NP the kitty [PP for ___ ] was pretty ?]]]]]

start-IP-VP-CP that-IP-NP-PP-end

So if children break these dependencies into smaller building blocks, they can identify if a dependency has bad segments (made up of one or more low probability building blocks).
Learn the right building blocks

Dickson, Pearl, & Futrell 2022, 2024, in prep.

Theory: The child tries to learn what the “best” building blocks are
Learn the right building blocks

Dickson, Pearl, & Futrell 2022, 2024, in prep.

Guiding intuition:
the “best” building blocks are the most “efficient” ones.
Learn the right building blocks

Efficient building blocks allow the representation of current and future *wh*-dependencies to be more probable.
Learn the right building blocks

Dickson, Pearl, & Futrell 2022, 2024, in prep.

Efficient building blocks allow the representation of current and future *wh*-dependencies to be more probable.

Why? One idea: Higher probability *wh*-dependencies are faster to process (comprehending or producing).
Learn the right building blocks

Dickson, Pearl, & Futrell 2022, 2024, in prep.

learning efficient building blocks

How? Look for building blocks that are a balance between
(1) how big they are
(2) how fast they are to put together to make a *wh*-dependency
Learning efficient building blocks

A balance between

1. how big they are
2. how fast they are to put together to make a wh-dependency

What did she say that he saw __?
Learn the right building blocks

learning efficient building blocks

a balance between

(1) how big they are

(2) how fast they are to put together to make a wh-dependency

$start$-$IP_{past}$-$VP_{say}$-$CP_{that}$-$IP_{past}$-$VP_{see}$-$end$
Learn the right building blocks

Dickson, Pearl, & Futrell 2022, 2024, in prep.

Learning efficient building blocks:

a balance between
(1) how big they are
(2) how fast they are to put together to
make a *wh*-dependency

\[start-\text{IP}_{\text{past}}-\text{VP}_{\text{say}}-\text{CP}_{\text{that}}-\text{IP}_{\text{past}}-\text{VP}_{\text{see}}-\text{end}\]

Pieces can be small, so that many of
them make up a *wh*-dependency
Learn the right building blocks

Dickson, Pearl, & Futrell 2022, 2024, in prep.

learning **efficient building blocks**

a balance between

1. **how big** they are
2. **how fast** they are to put together to make a *wh*-dependency

\[
\text{start-IP}_{\text{past}}-\text{VP}_{\text{say}}-\text{CP}_{\text{that}}-\text{IP}_{\text{past}}-\text{VP}_{\text{see}}-\text{end}
\]

It may be **slower** to put together many **small** pieces.
Learning efficient building blocks

A balance between

1. How big they are
2. How fast they are to put together to make a \textit{wh}-dependency

\textit{start-IP\textsubscript{past}-VP\textsubscript{say}-CP\textsubscript{that}-IP\textsubscript{past}-VP\textsubscript{see}-end}

Many smaller

Slower because many

But these pieces may get \textit{reused}, so that makes them \textit{faster} to access.
Learn the right building blocks

Dickson, Pearl, & Futrell 2022, 2024, in prep.

learning **efficient building blocks**

a **balance** between

(1) **how big** they are
(2) **how fast** they are to put together to
make a *wh*-dependency

\[ \text{start-IP}_{\text{past}}-\text{VP}_{\text{say}}-\text{CP}_{\text{that}}-\text{IP}_{\text{past}}-\text{VP}_{\text{see}}-\text{end} \]

many **smaller**

**slower** because many

But these pieces may get **reused**, so that
makes them **faster** to access.
Learn the right building blocks

Dickson, Pearl, & Futrell 2022, 2024, in prep.

learning efficient building blocks

a balance between

(1) how big they are

(2) how fast they are to put together to make a *wh*-dependency

\[
\text{start-IP}_\text{past-VP}_\text{say-CP}_\text{that-IP}_\text{past-VP}_\text{see-end}
\]

many smaller

many reused

Pieces can be big, so that only one makes up a *wh*-dependency
Learn the right building blocks

Dickson, Pearl, & Futrell 2022, 2024, in prep.

learning efficient building blocks

a balance between

(1) how big they are
(2) how fast they are to put together to make a wh-dependency

\[ \text{start-IP}_{\text{past-VP}_{\text{say-CP}_{\text{that-IP}_{\text{past-VP}_{\text{see-end}}}}}}} \]

many reused

many smaller

It may be faster to put together one big piece.
learning efficient building blocks

A balance between

1. how big they are
2. how fast they are to put together to make a $wh$-dependency

$start$-$IP_{\text{past}}$-$VP_{\text{say}}$-$CP_{\text{that}}$-$IP_{\text{past}}$-$VP_{\text{see}}$-$end$

many smaller
many reused

It may be slower if the piece is used rarely.
Learn the right building blocks

Dickson, Pearl, & Futrell 2022, 2024, in prep.

learning efficient building blocks

a balance between

(1) how big they are

(2) how fast they are to put together to

make a \textit{wh}-dependency

\[
\text{start-IP}_{\text{past}}-\text{VP}_{\text{say}}-\text{CP}_{\text{that}}-\text{IP}_{\text{past}}-\text{VP}_{\text{see}}-\text{end}
\]

many smaller

many reused

one big

faster because one

It may be slower if the piece is used rarely.
Learn the right building blocks

Dickson, Pearl, & Futrell 2022, 2024, in prep.

Learning efficient building blocks requires a balance between
1. how big they are
2. how fast they are to put together to make a wh-dependency

\[ \text{start-IP}_{\text{past}}-\text{VP}_{\text{say}}-\text{CP}_{\text{that}}-\text{IP}_{\text{past}}-\text{VP}_{\text{see}}-\text{end} \]

The most efficient option is probably a balance of bigger and smaller blocks that collectively are faster to access and put together.
Learn the right building blocks

learning efficient building blocks

a balance between

(1) how big they are
(2) how fast they are to put together to make a wh-dependency

\[ \text{start-IP}_{\text{past}}-\text{VP}_{\text{say}}-\text{CP}_{\text{that}}-\text{IP}_{\text{past}}-\text{VP}_{\text{see}}-\text{end} \]
Learn the right building blocks

Dickson, Pearl, & Futrell 2022, 2024, in prep.

learning efficient building blocks

How can children find the best balance?

many smaller
many reused
one big
one rare
Learn the right building blocks

Use Bayesian inference to search through the hypothesis space of all possible building blocks (O’Donnell 2015) and find an efficient set for children’s input.

Dickson, Pearl, & Futrell 2022, 2024, in prep.
Learn the right building blocks

Dickson, Pearl, & Futrell 2022, 2024, in prep.

So that’s what the modeled child will do
We’ll see if this modeled child can learn the target knowledge that real children learn.

Dickson, Pearl, & Futrell 2022, 2024, in prep.
We’ll have the modeled child learn from the same kind of input children encounter, for the same amount of time.
We’ll have the modeled child learn from the same kind of input children encounter, for the same amount of time.

\[ \approx 10,442,258 \text{ min} \]

(derived from De Villiers et al. 2008, Perkins & Lidz 2021, & Davis et al. 2004)

Estimated from the CHILDES Treebank (Pearl & Sprouse 2013)
We’ll have the modeled child learn from the same kind of input children encounter, for the same amount of time. 

$\approx 2.15$ million $wh$-dependencies 

Estimated from the CHILDES Treebank (Pearl & Sprouse 2013)
So how did the modeled child do?
A reminder of the **measurable target behavior patterns**
Our modeled child can generate almost all of these target patterns.
This means the modeled child, who encoded the “efficient building blocks” learning theory, was able to generate almost all of children’s target behavior.
It turns out this is also true even if the modeled child has child-like memory constraints that cause it to forget some of its input.
Takeaway: This learning theory (implemented by the modeled child using math) is pretty good at capturing children’s target behavior. So, it may be a pretty good explanation for how children acquire syntactic island knowledge from their input.
Some other discoveries from my research group, using computational cognitive modeling
Children transform the input signal they encounter in order to learn.
Sometimes acquisition success may result only when the data intake for acquisition is a selective subset of the available input.

However, sometimes acquisition success may also occur by taking a broader perspective on what counts as relevant data.

(Pearl 2023a; *English anaphoric one*: Pearl and Mis 2011, 2016; *syntactic islands*: Pearl and Sprouse 2013a,b, Pearl 2014, Pearl and Sprouse 2015, Pearl 2017, Bates and Pearl 2019, Pearl and Bates 2022; *English passive*: Nguyen and Pearl 2018, 2019, 2021).
Children’s knowledge during language development may be different than we thought.
To learn some types of linguistic knowledge, children may need to have language-specific knowledge built into their minds — however the exact form of this knowledge may be different than what we previously thought.

(English anaphoric one: Pearl and Lidz 2009, Pearl and Mis 2011, 2016; syntactic islands: Pearl and Sprouse 2013a,b, Pearl 2014, Pearl and Sprouse 2015, Dickson, Pearl and Futrell 2022; linking theories: Pearl and Sprouse 2019, 2021)
Children’s linguistic knowledge may sometimes be far more well-developed than we realize earlier than we realize, with strong similarities between child and adult representations.

(closed-class syntactic categories: Bates, Pearl and Braunwald 2018; quantifier scope ambiguity resolution: Savinelli, Scontras and Pearl 2017, 2018, Scontras and Pearl 2021; adjective ordering preferences: Bar-Sever, Lee, Scontras and Pearl 2018; pronoun interpretation: Forsythe and Pearl 2019, Pearl and Forsythe under review)
“Different” doesn’t automatically mean “worse” when it comes to language development.
While there is certainly input variation across socio-economic status, the same language learning outcome could still occur, despite the variation. So, interventions targeted at “fixing” the input aren’t likely to be effective.

*syntactic islands*: Bates and Pearl 2019, Pearl and Bates 2022
Even if the representations of very young children may not match adult representations, they can still be “good enough” for helping other acquisition processes get started.

early speech segmentation: Phillips and Pearl 2012, 2014a,b, 2015a,b,c, Pearl and Phillips 2018; early syntactic categorization: Bar-Sever and Pearl 2016
Learners with processing constraints (like children) may sometimes learn better than learners with fewer limitations ("less is more").

Computational cognitive modeling is one way to use math to investigate the magical process of child language acquisition.
It allows us to implement learning theories concretely, evaluate them, and better understand how they (and potentially children) work.
So let’s keep using this wonderful tool to investigate child language acquisition!
Thank you!

- Niels Dickson
- Richard Futrell
- Jon Sprouse
- Alandi Bates

BUCLD 2018    UCSD Linguistics 2020    ForMA Group 2020
UMD Linguistics 2020    BUCLD 2021    SCiL 2022
UArizona Linguistics 2022    UChicago LEAP 2022
UPenn 2023    Pomona Acquisition Workshop 2024
UCI QuantLang Collective

Lisa S. Pearl
Professor
Department of Language Science
SSPB 2219
University of California, Irvine
lpearl@uci.edu

Computation of Language Laboratory
UC Irvine