Today’s Plan: Computational models of language acquisition

I. Why

II. How

III. What we can learn

Who does... is pretty?

another one

Every kitty didn’t...
Why do you want to model language acquisition?
Why do you want to model language acquisition?

*What does it mean to model something?*
What does it mean to model something?

It’s a scientific technique, like running an experiment. So saying “I want to model $thing” is just like saying “I want to run an experiment about $thing.” Basically, it’s a fine plan, but the important question is why you’re doing it. That is, what question are you trying to answer?

Once you know what question you’re trying to answer, you can design the right test of it — whether that’s an experiment or a model or something else entirely.
So what questions should we be using models for?

“...these questions tend to concern the process of acquisition that yields adult knowledge – that is, how exactly acquisition proceeds, using particular learning strategies.” - Pearl 2017

The importance of theory

“...an informative model of acquisition is the embodiment of a specific theory about acquisition.” - Pearl 2017
The importance of theory

“...you need to first have a theory about how acquisition works. Then, the model can be used to

(1) **make all the components** of that acquisition theory **explicit,**

(2) **evaluate whether it actually works,** and

(3) **determine precisely what makes it work** (or not work).”

- Pearl 2017
Making the components explicit

“It often turns out that the acquisition theories that seem explicit to humans don’t actually specify all the details necessary to implement the strategies these theories describe.”
- Pearl 2017

Example: Learning linguistic parameter values from triggers in the input

Specific example:
The trigger for wh-movement is seeing a wh-word in a position different from where it’s understood (e.g., what in the question What did the penguin do ___what?)
Making the components explicit

The trigger for *wh*-movement is seeing a *wh*-word in a position different from where it’s understood (e.g., *what* in the question *What did the penguin do ___what?*)

What do children need to **know** or be able to do in order to recognize the appropriate *wh*-movement trigger in their input?

- **Know**: a certain word is one of these special *wh*-words
- **Do**: reliable segmentation of words in the utterance in order to recognize a *wh*-word not appearing where it’s understood
- **Do**: remember the fronted *wh*-word in the utterance reliably enough to update the internal parameter value
- **Know**: ignore utterances where the *wh*-word doesn’t move (e.g., echo questions like *The penguin did what?!*)
Making the components explicit

The trigger for **wh-movement** is seeing a *wh*-word in a position different from where it’s understood (e.g., *what* in the question *What did the penguin do ___what?*)

Now, what about the **wh-in-situ** option (for languages like Mandarin Chinese and Japanese)?

- Does this have a trigger too? What is it?
- If not, is **wh-in-situ** the default option that gets overridden by the presence of **wh**-movement triggers? If so, **how many** does it take?
- If there are no defaults but **wh-in-situ** also has no trigger, does the child use **indirect negative evidence** to decide her language is **wh-in-situ**? **How much** indirect negative evidence does it take?
The importance of theory

“...you need to first have a theory about how acquisition works. Then, the model can be used to

(1) make all the components of that acquisition theory explicit,

(2) evaluate whether it actually works, and

(3) determine precisely what makes it work (or not work).”

- Pearl 2017
Evaluating whether the theory works and determining what makes it work

“Once an acquisition theory is specified enough to implement in a computational model, we can then evaluate it by comparing the predictions it generates against the empirical data available from children.”
- Pearl 2017

Two basic outcomes:

- the model predictions **match** children’s data
- the model predictions **don’t match** children’s data
Evaluating whether the theory works and determining what makes it work

The model predictions match children’s data

This is an existence proof that the acquisition theory, as implemented in the model, is a way acquisition could proceed.

Note: Doesn’t rule out alternative acquisition theories

Two basic outcomes:
• the model predictions don’t match children’s data
Evaluating whether the theory works and determining what makes it work

The model predictions **match** children’s data

This is an **existence proof** that the acquisition theory, *as implemented in the model*, is a way acquisition *could* proceed.

The model predictions **don’t match** children’s data

This is then evidence **against that acquisition theory, *as implemented by the model***.

Remember: A model often specifies components of a theory that the original theory didn’t. So, if this particular theory implementation doesn’t work, maybe it’s a problem with those components, and not the theory more broadly.
Evaluating whether the theory works and determining what makes it work

The model predictions match children’s data

This is an existence proof that the acquisition theory, as implemented in the model, is a way acquisition could proceed.

The model predictions don’t match children’s data

This is then evidence against that acquisition theory, as implemented by the model.

If you have an implemented model (whether it succeeds or fails), you can look inside it to determine what exactly makes it work or not work. This is something that’s much more difficult to do with children’s minds.
Evaluating whether the theory works and determining what makes it work

Suppose we have a successful model of the acquisition of wh-movement from triggers.

What did the penguin do ___what?
Evaluating whether the theory works and determining what makes it work

What did the penguin do ___what?

We can see if it’s important for English children to ignore wh-echo questions where there’s no wh-movement, or how necessary a Mandarin Chinese default wh-in-situ value is.

The penguin did what?!

default = don’t move
Evaluating whether the theory works and determining what makes it work

What did the penguin do? __what?

The penguin did what?! ✓

default = don’t move

If the model’s predictions don’t match children’s behavior without these, we can say they’re necessary components of the learning strategy this theory describes and we can explain why (e.g., they filter the input or help the child navigate the hypothesis space).

This is **useful**!
Modeling as a useful tool

Modeling can be used as a tool for both developing and refining acquisition theories.

Notably, an acquisition theory actually includes two types of theories:

• theories of the learning process
• theories of the representations to be learned

An informative model incorporates both.
Today’s Plan:
Computational models of language acquisition

I. Why

II. How

III. What we can learn

Who does... is pretty?  
another one

Every kitty didn’t...
Today’s Plan:
Computational models of language acquisition

II. How
Language acquisition = An information processing task
Language acquisition = An information processing task

Given the available input,

*Look at that kitty!*
*There’s another one.*

**Input**

*Where did he hide?*
*What happened?*
Language acquisition = An information processing task

Given the available input, information processing done by human minds

Look at that kitty!
There’s another one.

Input
Where did he hide?
What happened?
Language acquisition = An information processing task

Given the available input, information processing done by human minds to build a **system of linguistic knowledge**

- Look at that kitty!
- There’s another one.
- Where did he hide?
- What happened?

**Input**

**Processing & generalization**

- words & morphemes
- metrical phonology
- syntactic categories
- syntax
- semantics
- pragmatics
Language acquisition = An information processing task

Given the available input, information processing done by human minds to build a system of linguistic knowledge whose output we observe

Look at that kitty!
There’s another one.

Where did he hide?
What happened?

Where’s the kitty?
That one’s really cute.
A framework that makes components of the acquisition task more explicit.
A framework that makes components of the acquisition task more explicit.

Distinguishes between things external to the child that we can observe (input signal, child’s behavior) vs. things internal to the child (everything else).
Perceptual encoding:

Turning the input signal into an internal linguistic representation = perceptual intake.
Perceptual encoding:
Involves current grammar
Perceptual encoding:
Involves current grammar being deployed in real time to parse the input.
Perceptual encoding:
Involves current grammar being deployed in real time to parse the input often drawing on extralinguistic systems
Generating observable behavior
Involves current linguistic representations being used by production systems.
Doing inference
Generalization happens
Doing inference

Generalization happens by using existing learning biases, (some of which may be innate and language-specific)
Doing inference

Generalization happens by using existing learning biases, (some of which may be innate and language-specific) operating over the acquisitional intake — what’s perceived as relevant for acquisition.
Doing inference

Generalization happens by using existing learning biases, (some of which may be innate and language-specific) operating over the acquisitional intake — what’s perceived as relevant for acquisition to produce the most up-to-date hypotheses about linguistic knowledge.
The current linguistic hypotheses are used in subsequent perceptual encoding.
This whole process **happens over and over again** throughout the **learning period**
This is language acquisition

An informative computational model of language acquisition captures these important pieces in an empirically-grounded way.
Informative computational models = informative about the learning strategies children use
A successful learning strategy is an existence proof that linguistic knowledge is attainable using the knowledge, learning biases, and capabilities comprising that strategy.
Learning strategies children use

Important learning strategy components include
• knowledge (= theories of representation)
Important learning strategy components include

- theories of representation
- biases & capabilities that must exist for that knowledge to be successfully deployed during acquisition (= theories of the learning process).
When building a specific model, it can be helpful to think about these different acquisition pieces in five main parts.
Initial state

What does the child start with? What knowledge, abilities, and learning biases does the child already have?
Initial state

What does the child start with? What **knowledge**, **abilities**, and learning **biases** does the child already have?

Example **knowledge**: syntactic categories exist and can be identified phrase structure exists and can be identified participant roles can be identified

---

Lidz & Gagliardi 2015
What does the child start with? What knowledge, abilities, and learning biases does the child already have?

Example abilities & biases: frequency information can be tracked, distributional information can be leveraged.

Initial state

Lidz & Gagliardi 2015
Initial state

What does the child start with? What knowledge, abilities, and learning biases does the child already have?

Example initial state: A strategy that depends on the frequency of certain syntactic structures would need the child to know about that syntactic structure via the developing grammar and/or Universal Grammar, recognize it in the input via the developing language processing abilities, and be able to track the frequency of that structure.
Initial state

What knowledge, abilities, and learning biases does the child start with?

Data intake

How does the modeled child perceive the input (=perceptual intake)? What part of the perceived data is used for acquisition (=acquisitional intake)?

Lidz & Gagliardi 2015
**Initial state**

What knowledge, abilities, and learning biases does the child start with?

---

**Data intake**

How does the modeled child perceive the input (=perceptual intake)? What part of the perceived data is used for acquisition (=acquisitional intake)?

---

Lidz & Gagliardi 2015

---

ex: all *wh*-utterances for learning about *wh*-dependencies
ex: all pronoun data when learning about anaphoric *one*
ex: syntactic and conceptual data for learning syntactic knowledge that links with conceptual knowledge

[defined by knowledge & biases/capabilities in the initial state]
**Initial state**

What knowledge, abilities, and learning biases does the child start with?

**Data intake**

What is the acquisitional intake?

**Inference**

How are updates made to the modeled child’s internal representations?

ex: probabilistic integration of available information (e.g., Bayesian inference)

ex: sequential hypothesis testing

[defined by knowledge & biases/capabilities in the initial state]
**Initial state**
What knowledge, abilities, and learning biases does the child start with?

**Data intake**
What is the **acquisitional intake**?

**Inference**
How are updates made?

**Learning period**
How long does the child have to learn?

---

**Diagram Description**

- **External Input**
  - Perceptual encoding
    - Developing grammar
    - Parsing procedures
    - Extralinguistic systems (audition, pattern recognition, memory, theory of mind, etc.)

- **Internal Behavior**
  - Perceptual intake (linguistic representations)
  - Production systems
  - Inference engine
    - Acquisitional intake
    - Universal grammar
  - Developing grammar

---

Lidz & Gagliardi 2015
What knowledge, abilities, and learning biases does the child start with?

What is the acquisitional intake?

How long does the child have to learn?

How are updates made?

Ex: 3 years, ~1,000,000 data points
Ex: 4 months, ~36,500 data points
What knowledge, abilities, and learning biases does the child start with?

What is the acquisitional intake?

How are updates made?

How long does the child have to learn?

What does successful acquisition look like? What knowledge is the child trying to attain (often assessed in terms of observable behavior)?
**Initial state**

What knowledge, abilities, and learning biases does the child start with?

**Data intake**

What is the *acquisitional intake*?

**Inference**

How are updates made?

**Learning period**

How long does the child have to learn?

**Target state**

What does successful acquisition look like?

---

ex: *Where did Jack think the necklace from ___ was too expensive? 😞*

ex: Where did Jack buy a necklace from ___ for Lily for her birthday? 😊
What knowledge, abilities, and learning biases does the child start with?

What is the acquisitional intake?

How long does the child have to learn?

What does successful acquisition look like?

How are updates made?
Initial state
What knowledge, abilities, and learning biases does the child start with?

Target state
What does successful acquisition look like?

Data intake
What is the acquisitional intake?

Inference
How are updates made?

Learning period
How long does the child have to learn?

Defining each of these pieces for a model (as relevant) can help streamline the modeling process and make sure we’re building an informative model.
Building an informative model about...

Which learning strategies could children be using?
Building an informative model about...

Which learning strategies could children be using?

Which learning biases are necessary?
(Pearl & Sprouse in prep., Pearl, Ho, & Detrano in press, 2014; Pearl & Mis 2016, Pearl & Sprouse 2015, 2013a, 2013b, Pearl & Mis 2011, Pearl & Lidz 2009, Pearl 2008, Pearl & Weinberg 2007)
Which learning strategies could children be using?

Which learning biases are necessary?

Which knowledge representations are learnable — and which aren’t? (Pearl, Ho, & Detrano in press, 2014; Pearl 2017, Pearl 2011, Pearl 2009)
Building an informative model about...

Which learning strategies could children be using?

Which learning biases are necessary?

Which knowledge representations are learnable — and which aren’t?

When do children learn different aspects of the linguistic system?

(Bates, Pearl, & Braunwald in prep., Nguyen & Pearl in press, Caponigro, Pearl et al. 2012, Caponigro, Pearl et al. 2011)
Building an informative model about...

Which learning strategies could children be using?
Which learning biases are necessary?
Which knowledge representations are learnable — and which aren’t?
When do children learn different aspects of the linguistic system?

What factors affect children’s observable behavior?

(Savinelli, Scontras, & Pearl in prep., Nguyen & Pearl in press, Savinelli, Scontras, & Pearl 2017)
Today’s Plan:
Computational models of language acquisition

I. Why

II. How

III. What we can learn

Who does... is pretty?

another one

Every kitty didn’t...
Today’s Plan:
Computational models of language acquisition

II. How

A concrete example
How do we model language acquisition?

A concrete example with speech segmentation
How do we model language acquisition?
An example with speech segmentation

= ʍəɾəɾiɾiɾi
ʍə ə prir kir
what a pretty kitty!
How do we model language acquisition?
An example with speech segmentation

(1) Decide what kind of learner the model represents
   This depends on what task you’re modeling

For the first stages of speech segmentation:
Typically developing 6- to 8-month-old child learning first language
How do we model language acquisition?
An example with speech segmentation

(2) Decide what data the child learns from (input)

This depends on your acquisition theory and the empirical data available
How do we model language acquisition?
An example with speech segmentation

(2) Decide what data the child learns from (input)

Example empirical data: CHILDES database
http://childes.talkbank.org

Video/audio recordings of speech samples, along with transcriptions and some structural annotations.
How do we model language acquisition?
An example with speech segmentation

(3) Decide how the child perceives the data, and which data are relevant (intake)
This depends on your acquisition theory

what a pretty kitty!
How do we model language acquisition?

An example with speech segmentation

(3) Decide how the child perceives the data, and which data are relevant (intake)

what a pretty kitty!

syllables with stress

= ˈʌ ɾə ˈɪɾi ˈɪɾi ˈɪɾi
How do we model language acquisition?
An example with speech segmentation

Many models will try to make **cognitively plausible** assumptions about how the child is representing and processing input data.

\[
\text{what a pretty kitty!}
\]
How do we model language acquisition?
An example with speech segmentation

(4) Decide what hypotheses the child has and what information is being tracked in the input

This depends on your acquisition theory

= w'ʌ ɾə pɾɪɾ i k'ɪɾ i

what a pretty kitty!
How do we model language acquisition?
An example with speech segmentation

(4) Decide what hypotheses the child has and what information is being tracked in the input

Example hypotheses: what the words are
How do we model language acquisition?

An example with speech segmentation

Example information:
transitional probability between syllables,
stress on syllables

(4) Decide what hypotheses the child has and what information is being tracked in the input

```
what a pretty kitty!

w'ʌ ɾə prɪɾi k'ɾi
w'ʌ re prɪɾi k'ɾi
w'ʌ re prɪɾi k'ɾi
```

```
```
How do we model language acquisition?
An example with speech segmentation

(5) Decide how belief in different hypotheses is updated

This depends on your acquisition theory

Example: based on \textit{transitional probability} between syllables
How do we model language acquisition?

An example with speech segmentation

(5) Decide how belief in different hypotheses is updated

This depends on your acquisition theory

Example: based on **transitional probability** between syllables
How do we model language acquisition?

An example with speech segmentation

(6) Decide what the measure of success is

This can be based on your theory...
How do we model language acquisition?

An example with speech segmentation

\[ w'\Delta \quad r\quad p'\quad r\quad i \quad k'\quad r \]

\[ w'\Delta \quad r\quad p'\quad i \quad r \quad k'\quad i \quad r \]

= w'\Delta \quad r\quad p'\quad i \quad r \quad k'\quad i \quad r

what a pretty kitty!

(6) Decide what the measure of success is

This can be based on your theory or empirical data about behavior
How do we model language acquisition? An example with speech segmentation

(6) Decide what the measure of success is

Example developing knowledge

Proto-lexicon of word forms

This can be based on your theory or empirical data about behavior
How do we model language acquisition?

An example with speech segmentation

Example behavior indicating developed knowledge:
Recognizing useful units (such as words) in a fluent speech stream, as indicated by looking time behavior

(6) Decide what the measure of success is.

This can be based on your theory or empirical data about behavior.
How do we model language acquisition?

An example with speech segmentation

This is the heart of the model
How do we model language acquisition?
An example with speech segmentation

(7) Implement the model in a programming language of choice
How do we model language acquisition?

An example with speech segmentation

(8) See how well the model did w.r.t. the measure of success

Example developing knowledge

Proto-lexicon of word forms

\[
\begin{align*}
\text{what} & : w^l \Lambda \text{ra} \text{prri ri kiri} \\
\text{a} & : \text{a} \\
\text{pretty} & : \text{priri} \\
\text{kitty} & : \text{kiri}
\end{align*}
\]
How do we model language acquisition?

An example with speech segmentation

(8) See how well the model did w.r.t. the measure of success

Recognizing useful units (such as words) in a fluent speech stream, as indicated by looking time behavior

???
How do we model language acquisition?

An example with speech segmentation

(8) See how well the model did w.r.t. the measure of success

From this, we can determine how well the model did — and more importantly, how well the strategy implemented concretely in the model did.
How do we model language acquisition?

An example with speech segmentation

(9) Interpret the results for other people who aren’t you so they know why they should care

“The modeled child has the same developing knowledge as we think 8-month-olds do. This strategy can be what they’re using!”
How do we model language acquisition?

An example with speech segmentation

(9) Interpret the results for other people who aren’t you so they know why they should care

“The modeled child can reproduce the behavior we see in 8-month-olds. This strategy could be what they’re using to generate that behavior!”
Today’s Plan:
Computational models of language acquisition

II. How

Levels of explanation
How do we model language acquisition?

What **level** of model do you want to build?

A very basic question:
Is it possible for the child with a **specific initial state** to use the **acquisitional intake** to achieve the **target state**?

**Computational-level** (Marr 1982)

Is this the right conceptualization of the acquisition task? Do we have the right goal in mind?
How do we model language acquisition?

What **level** of model do you want to build?

**Computational-level**
A very basic question:
Is it possible for the child with a **specific initial state** to use the **acquisitional intake** to achieve the **target state**?

Helpful for determining **if this implementation of the acquisition task is the right one.**

Are these **useful** learning assumptions for children to have? Are these **useful** linguistic representations?
How do we model language acquisition?

What level of model do you want to build?

**Computational-level**
A very basic question:
Is it possible for the child with a specific initial state to use the acquisitional intake to achieve the target state?

This is typically implemented as an ideal learner model, which isn’t concerned with the cognitive limitations and incremental learning restrictions children have.

(That is, useful for children is different from useable by children in real life.)
How do we model language acquisition?

What **level** of model do you want to build?

**Computational-level**
A very basic question:
Is it possible for the child with a specific initial state to use the acquisitional intake to achieve the target state?

Practical note:
Doing a computational-level analysis is often a really good idea to make sure we’ve got the right conceptualization of the acquisition task (see Pearl 2011 for the trouble you can get into when you don’t do this first).
How do we model language acquisition?

What **level** of model do you want to build?

**Computational-level**
A very basic question:
Is it possible for the child with a **specific initial state** to use the **acquisitional intake** to achieve the **target state**?

*(What happened in a nutshell in Pearl 2011)*

Why do none of these learning strategies work?

Because they’re solving the wrong acquisition task... oops.
How do we model language acquisition?

What **level** of model do you want to build?

Another basic question:
Is it possible for the child with a specific initial state to use the acquisitional intake to achieve the target state in the amount of time children typically get to do it, given the incremental nature of learning and children’s cognitive constraints?

Computational-level

Lidz & Gagliardi 2015
How do we model language acquisition?

What level of model do you want to build?

Another basic question:
Is it possible for the child with a specific initial state to use the acquisitional intake to achieve the target state in the amount of time children typically get to do it, given the incremental nature of learning and children’s cognitive constraints?

Algorithmic-level (Marr 1982)
Is it possible for children to use this strategy? That is, once we know it’s useful for children, it’s important to make sure it’s also useable by children.
How do we model language acquisition?

What level of model do you want to build?

**Computational-level**

Another important (not so basic) question: If we have an algorithm that seems *useable* by children to *usefully* solve an acquisition task, how is it *implemented* in the brain?

**Algorithmic-level**

**Implementational-level**

Lidz & Gagliardi 2015
How do we model language acquisition?

What level of model do you want to build?

Another important (not so basic) question: If we have an algorithm that seems useable by children to usefully solve an acquisition task, how is it implemented in the brain?

Implementational-level

This isn’t easy to model yet. Advances in natural language processing: ways to encode complex information into distributed representations like what we think the brain uses.

How do we model language acquisition?

What level of model do you want to build?

The types I generally work with

- **Computational-level**
- **Algorithmic-level**
- **Implementational-level**
Today’s Plan:
Computational models of language acquisition

I. Why

II. How

III. What we can learn

Who does... is pretty?

Another one

Every kitty didn’t...

Noun
Today’s Plan:
Computational models of language acquisition

III. What we can learn

- speech segmentation
- syntax
- metrical phonology
- syntactic categorization
- pragmatics
- syntax, semantics

Who does... is pretty?

Kitty

Every kitty didn’t...
What we can learn

= \text{speech segmentation}

what a pretty kitty!
What we can learn

Investigating a Bayesian inference strategy for the very early stages of speech segmentation occurring around six months


\[ P(s|u) \propto P(s)P(u|s) \]
What we can learn

Bayesian inference

\[ P(s|u) \propto P(s)P(u|s) \]

Strategy: Identify a proto-lexicon of words that best generates the observable fluent speech utterances

Mathematically encoded preferences:

What we can learn

Bayesian inference

\[ P(s|u) \propto P(s)P(u|s) \]

Strategy: Identify a proto-lexicon of words that best generates the observable fluent speech utterances

Mathematically encoded preferences:
(1) Prefer shorter words

What we can learn

Bayesian inference

$$P(s|u) \propto P(s)P(u|s)$$

Strategy: Identify a proto-lexicon of words that best generates the observable fluent speech utterances

Mathematically encoded preferences:

1. Prefer shorter words
2. Prefer lexicons with fewer words

What we can learn

Bayesian inference

\[ P(s|u) \propto P(s)P(u|s) \]

Strategy: Identify a proto-lexicon of words that best generates the observable fluent speech utterances

Mathematically encoded preferences:

1. Prefer shorter words
2. Prefer lexicons with fewer words

Find the best segmentation

What we can learn

Bayesian inference

\[ P(s|u) \propto P(s)P(u|s) \]

Strategy: Identify a proto-lexicon of words that best generates the observable fluent speech utterances

Mathematically encoded preferences:
(1) Prefer shorter words
(2) Prefer lexicons with fewer words

Find the best segmentation that balances these proto-lexicon preferences

What we can learn

Bayesian inference

$$P(s|u) \propto P(s)P(u|s)$$

Strategy: Identify a proto-lexicon of words that best generates the observable fluent speech utterances

Mathematically encoded preferences:

1. Prefer shorter words
2. Prefer lexicons with fewer words

Find the best segmentation that balances these proto-lexicon preferences and can generate the observable fluent speech utterances

What we can learn

Bayesian inference

\[ P(s|u) \propto P(s)P(u|s) \]

Is it useful?

Computational-level modeled learners using this strategy segment fairly well, given realistic English child-directed speech data.

The inferred proto-lexicons, while not perfect, are very useful for subsequent stages of language acquisition.

What we can learn

Bayesian inference

\[ P(s|u) \propto P(s)P(u|s) \]

Is it useful?

Algorithmic-level modeled learners with cognitive constraints on their inference and memory can still use this strategy and segment English quite well.

Is it usable?

What we can learn

Bayesian inference

\[ P(s|u) \propto P(s)P(u|s) \]

- Is it **useful**?
- Is it **useable**?
- Does it work for **different languages**?

It segments well for languages with different morphology and syllable properties: Spanish, Italian, German, Hungarian, Japanese, Farsi

What we can learn

Bayesian inference

\[ P(s|u) \propto P(s)P(u|s) \]

Is it useful? ✓

Is it usable? ✓

Does it work for different languages? ✓

This kind of Bayesian inference seems to be a good proposal for a very early speech segmentation strategy.

What we can learn

speech segmentation

syntax

metrical phonology

Noun

syntactic categorization

pragmatics

syntax, semantics

Who does ... is pretty? another one
Every kitty didn’t ...
What we can learn

syntax, semantics

“Oh look — a pretty kitty!”

“Look — there’s another one!”
What we can learn syntax, semantics another one

“Oh look — a pretty kitty!”

“Look — there’s another one!”

Interpretation: another pretty kitty
same
syntactic category
???
What we can learn  

syntax, semantics

another one

“Oh look — a pretty kitty!”

“Look — there’s another one!”

Interpretation: another

same

syntactic category

???

bigger than a plain Noun

Noun

| pretty kitty
What we can learn

syntax, semantics

another one

“Look — there’s another one!”

Interpretation: another the pretty kitty

same syntactic category

smaller than a full Noun Phrase

Noun Phrase

the

Noun

| pretty kitty
What we can learn

“Look — there’s another one!”

Interpretation: another

In-between category **Noun’**
that includes strings with nouns
and modifiers+nouns

**Noun Phrase**

- the
- **Noun’**
- **Noun’**
- **Noun**
- **pretty kitty**
What we can learn

“Oh look — a pretty kitty!”

“Look — there’s another one!”

Interpretation: another

This is why we can also interpret one as just kitty.
What we can learn: syntax, semantics

“Oh look — a pretty kitty!”

“Do you see another one?”

Lidz, Waxman, & Freedman 2003: 18-month-old interpretations
What we can learn syntax, semantics another one

“Oh look — a pretty kitty!”

“Do you see another one?”

pretty kitty Noun’

Lidz, Waxman, & Freedman 2003: 18-month-old interpretations
What we can learn

“Oh look — a pretty kitty!”

“Do you see another kitty?”

Lidz, Waxman, & Freedman 2003: 18-month-old interpretations
What we can learn

“Oh look — a pretty kitty!”

“Do you see another kitty?”

Lidz, Waxman, & Freedman 2003: 18-month-old interpretations
What we can learn

syntax, semantics

another one

“Oh look — a pretty kitty!”

“Do you see another pretty kitty?”

Lidz, Waxman, & Freedman 2003: 18-month-old interpretations
What we can learn

“Oh look — a pretty kitty!”

“Do you see another pretty kitty?”

Lidz, Waxman, & Freedman 2003: 18-month-old interpretations
What we can learn

“Oh look — a pretty kitty!”

“Do you see another one?”

Several learning strategies implemented with algorithmic-level modeled learners, given realistic samples of English child-directed speech.

Pearl & Mis 2016
What we can learn

“Oh look — a pretty kitty!”

“Do you see another one?”

Algorithmic-level
Evaluated on whether they matched 18-month-old looking preferences.

Pearl & Mis 2016
What we can learn

"Oh look — a pretty kitty!"

"Do you see another one?"

Algorithmic-level

Two strategies were successful at generating the 18-month-old behavior. We can then look inside the modeled learner and see what the underlying representations were.
What we can learn

“Oh look — a pretty kitty!”

Algorithmic-level

“Do you see another one?”

Strategy 1: Ignore some of the available one data in the input

Inference engine
Acquisitional intake
Developing grammar
What we can learn

“Oh look — a pretty kitty!”

Algorithmic-level

“Do you see another one?”

Strategy 1: Ignore some of the available one data in the input

Adult representations

Noun’

pretty kitty

But...required additional situational context to be present to succeed.

Less robust

Pearl & Mis 2016
What we can learn

"Oh look — a pretty kitty!"

Algorithmic-level

Strategy 1: Ignore

Less robust

Strategy 2: Include other pronoun data besides one data in the intake

Pearl & Mis 2016
What we can learn

**syntax, semantics**

“Oh look — a pretty kitty!”

**Algorithmic-level**

**Strategy 1: Ignore**

Less robust

“Do you see another one?”

**Strategy 2: Include other pronoun data besides one data in the intake**

Immature representations

- **Noun’ only in certain linguistic contexts**
- **pretty kitty**
- otherwise **Noun**

But...does this for pretty much any situational context.

More robust

Pearl & Mis 2016
What we can learn

“Oh look — a pretty kitty!”

Algorithmic-level

Strategy 1: Ignore

Less robust

Strategy 2: Include other

More robust

By modeling, we have two concrete proposals for how children learn the knowledge they do by 18 months.

This also motivates future experimental work to distinguish these two possibilities.
What we can learn

- **speech segmentation**
- **syntax**
- **metrical phonology**
- **syntactic categorization**
- **pragmatics**
- **syntax, semantics**

Examples:
- "Who does ... is pretty?"
- "Another one"
- "Every kitty didn’t ..."
What we can learn

This kitty was bought as a present for someone.

Lily thinks this kitty is pretty.

What’s going on here?

Who does Lily think the kitty for is pretty?

What does Lily think is pretty, and who does she think it’s for?
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 ___ is pretty?

This dependency is not allowed in English.

One explanation: The dependency crosses a “syntactic island” (Ross 1967)
What we can learn

What’s going on here? syntactic island

Who does Lily think the kitty for is pretty?

Jack is somewhat tricksy.
He claimed he bought something.

What did Jack make the claim that he bought ___?
What we can learn

What’s going on here? syntactic island

Who does Lily think the kitty for ___ is pretty?

What did Jack make the claim that he bought ___?

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 we can learn

What’s going on here? syntactic island

Who does Lily think the kitty for ___ is pretty?

What did Jack make the claim that he bought ___?

What did Elizabeth wonder whether Jack bought ___?

What did Elizabeth worry if Jack bought ___?

He claimed he bought something.

Elizabeth worried it was something dangerous.
What we can learn

Adults judge these dependencies to be far worse than many others, including others that are very similar except that they don’t cross syntactic islands (Sprouse et al. 2012).
What we can learn

Who does Lily think the kitty for is pretty?

Previous learning theories suggested children need syntactic-island-specific innate knowledge.
What we can learn

An alternative learning strategy suggests children need **less-specific linguistic prior knowledge** along with **probabilistic learning**.

Pearl & Sprouse (2013a, 2013b, 2015)
What we can learn

What we can learn

Who does Lily think the kitty for is pretty?

This alternative strategy was implemented in an algorithmic-level learning model that learned from realistic samples of child-directed speech. The modeled learner was able to reproduce the pattern of adult judgments.

Pearl & Sprouse (2013a, 2013b, 2015)
What we can learn

Lily think the kitty for ___ is pretty?

Who does Lily think the kitty for ___ is pretty?

Upshot: Children can learn these sophisticated restrictions without relying as much on very specific linguistic knowledge that’s necessarily innate.

Pearl & Sprouse (2013a, 2013b, 2015)
Today’s Plan:
Computational models of language acquisition

I. Why

II. How

III. What we can learn

Who does... is pretty?

Noun

another one

Every kitty didn’t...
Today’s Plan:
Computational models of language acquisition

I. Why: Because language acquisition is pretty amazing and we want to understand how it works

II. How

III. What we can learn

Who does... is pretty?
Another one
Every kitty didn’t...
Today’s Plan:
Computational models of language acquisition

I. Why: Because language acquisition is pretty amazing and we want to understand how it works

II. How: By building informative computational models

III. What we can learn

Who does... is pretty?

Every kitty didn’t...

Another one

Noun
Today’s Plan:
Computational models of language acquisition

I. Why: Because language acquisition is pretty amazing and we want to understand how it works

II. How: By building informative computational models

III. What we can learn: A lot about a lot

- speech segmentation
- syntax
- syntax, semantics
- metrical phonology
- syntactic categorization
- Who does... is pretty?
- Noun
- KI tty
- another one
- Every kitty didn’t...
Today’s Plan:
Computational models of language acquisition

I. Why: Because language acquisition is pretty amazing and we want to understand how it works

II. How: By building informative computational models

III. What we can learn: A lot about a lot

This is a great tool - so let’s use it to understand how linguistic representations develop!
Thank you!

This work was supported in part by NSF grants BCS-0843896 and BCS-1347028.
Extra material
What we can learn

speech segmentation

syntax

Who does... is pretty?

Another one

Every kitty didn’t...

Noun

metrical phonology

syntactic categorization

pragmatics

syntax, semantics
What we can learn

metrical phonology

✓ Kitty

✗ ki TTY
What we can learn

- A DO ra ble
- A do RA ble
- a DO ra BLE
- Kitty
- KI tty
- ki TTY

metrical phonology
What we can learn

- a DO ra ble
- A do RA ble
- a DO ra BLE
- KI tty
- ki TTY

Our underlying knowledge representation of the metrical phonology system allows us to generate these metrical stress preferences.
What we can learn

knowledge representation options

parameters whose values must be set
What we can learn

knowledge representation options

parameters whose values must be set

English
What we can learn

knowledge representation options

parameters

Violable constraints that must be ranked

English
What we can learn

metrical phonology

knowledge representation options

parameters

Violable constraints that must be ranked

English
What we can learn

knowledge representation options

These representations have some similarities, but aren’t obviously using identical variables.

How do we choose among these representations and their English versions?
What we can learn

metrical phonology

What we can learn

knowledge representation options

parameters

constraints

English

How do we choose among these representations and their English versions?

Answer: Let’s see how learnable they are from the English data children typically encounter!

Pearl et al. 2014, Pearl 2017, Pearl et. al in press
What we can learn

**metrical phonology**

- a DO ra ble
- KI tty
- A do RA ble
- ki TTY
- a DO ra BLE

**knowledge representation options**

**parameters**

**constraints**

**how learnable they are**

**Computational-level analysis**

Modeled learners given realistic samples of English child-directed speech can identify parameter combinations or constraint rankings that are very good at accounting for the input especially if children use a data filter.

Pearl et al. 2014, Pearl 2017, Pearl et. al in press
What we can learn

Knowledge representation options

- Parameters
- Constraints

Metrical phonology

- a DO ra ble
- A do RA ble
- ki TTY

How learnable they are

Computational-level analysis

But the best options for English data aren’t the ones currently proposed for English.

Pearl et al. 2014, Pearl 2017, Pearl et. al in press
What we can learn

metrical phonology

knowledge representation options

parameters

constraints

how learnable they are

Computational-level analysis

Other options (differing very slightly) are much more easily learnable.

Pearl et al. 2014, Pearl 2017, Pearl et al in press
What we can learn

metrical phonology

 Goodman et al. 2014, Goodman 2017

knowledge representation options

parameters

constraints

Computational-level analysis

And two do particularly well when a data filter is in place.

Pearl et al. 2014, Pearl 2017, Pearl et. al in press
What we can learn

**metrical phonology**
- a DO ra ble ✓ KI tty ✓
- A do RA ble ✗
- a DO ra BLE ✗

knowledge representation options

parameters

constraints

By modeling acquisition, we provide support for these two theories of English representation.

Pearl et al. 2014, Pearl 2017, Pearl et. al in press
What we can learn

- **speech segmentation**
- **syntax**
- **pragmatics**
- **metrical phonology**
- **syntactic categorization**

Who does... is pretty?

*Every kitty didn’t...*
What we can learn

syntactic categorization

Noun

idea

penguin

sparkle

kitty

unicorn

owl
Nouns behave similarly:
They can combine with certain types of words to make larger units (like Noun Phrases).
What we can learn

Determiner + Noun ("the kitty")

\[ NP \rightarrow \text{Det} + \text{N} \]

Nouns behave similarly:

They can combine with certain types of words to make larger units (like Noun Phrases).
What we can learn

syntactic categorization

Rule with category Noun = new phrases with words of category Noun

- penguin
- glitter
- idea
- unicorn
- kitty
- owl

Determiner + Noun ("the")

[ NP \rightarrow Det + N ]

This is very handy for generating new expressions we haven’t heard before.
What we can learn

syntactic categorization

Determiner + Noun ("the dax")

[NP → Det + N]

Rule with category Noun = new phrases with words of category Noun

This is very handy for generating new expressions we haven’t heard before.
We have many categories in human language.

Some are open-class — it’s easy to add new words to them.
What we can learn

We have many categories in human language.

Some are open-class — it’s easy to add new words to them.

\[ VP \to \text{Negation} + V \]

It’s not {\texttt{daxing}}
- it’s dancing!

 Veronica

<table>
<thead>
<tr>
<th>Verb</th>
</tr>
</thead>
<tbody>
<tr>
<td>find</td>
</tr>
<tr>
<td>dance</td>
</tr>
<tr>
<td>adore</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>glitter</td>
</tr>
<tr>
<td>idea</td>
</tr>
<tr>
<td>unicorn</td>
</tr>
<tr>
<td>penguin</td>
</tr>
<tr>
<td>kitty</td>
</tr>
<tr>
<td>owl</td>
</tr>
</tbody>
</table>

| syntactic categorization |
What we can learn

We have many categories in human language.
Some are open-class — it’s easy to add new words to them.
What we can learn

We have many categories in human language.

Some are closed-class — the words in them are fixed.

[VP → Negation + V]

It’s not daxing
- it’s dancing!
What we can learn

We have many categories in human language.

Some are closed-class — the words in them are fixed.

\[
[VP \rightarrow \text{Auxiliary} + V] \\
\text{It would sing} \\
\text{if it could sing}
\]

Negation

\[\text{didn’t} \quad \text{not} \quad \text{wouldn’t} \quad \text{can’t} \quad \text{won’t}\]

Noun

\[\text{penguin} \quad \text{owl} \quad \text{glitter} \quad \text{idea} \quad \text{unicorn} \quad \text{kitty} \]

Verb

\[\text{stand} \quad \text{surprise} \quad \text{find} \quad \text{dance} \quad \text{adore} \]

Auxiliary

\[\text{might} \quad \text{would} \quad \text{could} \quad \text{will} \quad \text{should} \quad \text{can}\]
What we can learn

There’s significant debate on when these categories develop.
What we can learn

There’s significant debate on when these categories develop.

Easy to observe: When children know individual words.

it’s dancing

dance
What we can learn

**syntactic categorization**

Negation

didn't

not

can't

won't

Harder to observe: When children have recognized these words belong to categories.

There’s significant debate on when these categories develop.

**Auxiliary**

might

would

could

will

should

can

**Verb**

find
dance

**Noun**

penguin

owl

kitty

**Noun**

idea

unicorn

**Noun**

glitter

Harder to observe: When children have recognized these words belong to categories.

it’s dancing
What we can learn: **syntactic categorization**

What we can do: *Computational-level* analysis of children’s productions, using formal metrics that describe how children *generate their utterances* given their underlying representations.

Bates, Pearl & Braunwald, in prep.
Analyzing the utterances produced by a single American English child between the ages of 20 and 24 months.
Computational-level

Analyzing the utterances produced by a single American English child between the ages of 20 and 24 months

Utterances compatible with having adult-like closed-class categories, but not adult-like open-class categories.

Bates, Pearl & Braunwald, in prep.
This suggests that closed-class categories may develop into an adult-like state earlier than open-class categories and much earlier than previously thought.
What we can learn

- **speech segmentation**
- **syntax**
- **pragmatics**
- **metrical phonology**
- **syntactic categorization**

Examples:
- *Who does ... is pretty?*
- *another one*
- *Every kitty didn’t ...*
What we can learn

“Every kitty didn’t sit on the stairs”

✗ No kittens sat on the stairs.

✔ Not all kittens sat on the stairs.

Why are two interpretations available?

Quantifier scope
What we can learn

Quantifier scope

“Every kitty didn’t sit on the stairs”

\[ \forall \]

No kitties sat on the stairs.

Not all kitties sat on the stairs.
What we can learn

Quantifier scope

“Every kitty didn’t sit on the stairs”

\[ \forall k \text{ k sat on the stairs} \]

surface \[ \forall \text{kitties } k \text{ k sat on the stairs} \]

“For all kitties k, it’s not true that k sat on the stairs”

[X] No kitties sat on the stairs.

✓ Not all kitties sat on the stairs.
What we can learn

Quantifier scope

“Every kitty didn’t sit on the stairs”

∀

surface  ∀kitties k  k sat on the stairs

“For all kitties k, it’s not true that k sat on the stairs”

No kitties sat on the stairs.

inverse  ∀kitties k, k sat on the stairs

“It’s not true that for all kitties k, k sat on the stairs”

Not all kitties sat on the stairs.
What we can learn

Quantifier scope

✓ “Every kitty didn’t sit on the stairs”

surface ∀ No kitties sat on the stairs.

inverse ∀ Not all kitties sat on the stairs.

Adults
What we can learn

Quantifier scope

X “Every kitty didn’t sit on the stairs”

surface ∨ No kitties sat on the stairs.

inverse ∀ Not all kitties sat on the stairs.

5-year-olds

But why?
What we can learn

Quantifier scope

X “Every kitty didn’t sit on the stairs”

∀ ??? Not all kitties sat on the stairs.

inverse

∀

5-year-olds

One idea: grammatical processing problem
What we can learn

Quantifier scope

X “Every kitty didn’t sit on the stairs”

\( \forall \) ?? Not all kitties sat on the stairs.

5-year-olds

One idea: grammatical processing problem

The inverse scope is harder to get from the surface string.
What we can learn

Quantifier scope

❌ “Every kitty didn’t sit on the stairs”

∀ ?? Not all kitties sat on the stairs.

inverse ∀ 5-year-olds

One idea: grammatical processing problem

Another idea: pragmatic context management problem.
What we can learn

Quantifier scope

Every kitty didn’t sit on the stairs

Not all kitties sat on the stairs.

Did none of the kitties sit on the stairs?

Do kitties like stairs? QUD How many kitties sat on the stairs?

5-year-olds

One idea: grammatical processing problem

Another idea: pragmatic context management problem.

Children thought the topic of conversation (the implicit Question Under Discussion) was something else and this utterance doesn’t answer that QUD very well.
What we can learn

Quantifier scope

\[ \forall \quad \text{"Every kitty didn’t sit on the stairs"} \]

\[ \forall \quad \text{Not all kittens sat on the stairs.} \]

\[ \text{Kitties don’t like stairs} \]

\[ \text{expectations about the world} \]

Kitties love stairs. \quad Kitties don’t care about stairs.

5-year-olds

One idea: \textit{grammatical processing} problem

Another idea: \textit{pragmatic context} management problem.

\textit{QUD}

Children’s prior \textit{expectations about the world} make this utterance less informative.
What we can learn

Quantifier scope

≠ “Every kitty didn’t sit on the stairs”

inverse

∀ Not all kitties sat on the stairs.

grammatical processing expectations about the world

5-year-olds

It’s hard to manipulate only one of these factors in experimental research investigating children’s responses.
What we can learn

Quantifier scope

"Every kitty didn’t sit on the stairs"

∀

inverse

Not all kittens sat on the stairs.

QUD

grammatical processing

expectations about the world

5-year-olds

Using a computational-level model that formalizes the separate contribution of each factor, we can determine which ones have the largest impact on children’s observed behavior.

Savinelli, Scontras, & Pearl 2017
What we can learn

Quantifier scope

X “Every kitty didn’t sit on the stairs”

∀ ?? Not all kitties sat on the stairs.

inverse

∀

Grammatical processing

QUD

Expectations about the world

5-year-olds

The pragmatic factors seem to be the driving force behind children’s behavior. This suggests that 5-year-olds are still developing their ability to manage the pragmatic context of a conversation as well as adults do.

Savinelli, Scontras, & Pearl 2017