Using Computational Modeling to Understand Language Acquisition

UC Computational Social Science
Language acquisition: How humans learn language knowledge
Language acquisition: How humans learn language knowledge

First language acquisition = Learning native language(s)

Happens as a young child
Language acquisition: How humans learn language knowledge

Second language acquisition = Learning non-native/foreign language(s)

Happens as an older child or adult

First language acquisition
First language acquisition

How do children acquire the knowledge about language that they do from the language data they have?
Why first language acquisition?

Babies are amazing at learning language
Babies are amazing at learning language

And they learn *a lot!*
And they learn *a lot*!

Like what?
Like what?
Everything you know about your native language(s).
You know how to identify words in fluent speech (speech segmentation)

what a pretty kitty!
what a pretty kitty!

speech segmentation

You know how to pronounce words (metrical phonology)

✔ KI tty
× ki TTY
You know that certain words behave like other words (syntactic categorization)

what a pretty kitty!

speech segmentation

metrical phonology

Noun

what a pretty ___!

penguin

kitty

owl
You know how to interpret words in context (syntax, semantics)

“Oh look — a pretty kitty!”
“Look — there’s another one!”
This kitty was bought as a present for someone.

Lily thinks this kitty is pretty.

Who does Lily think the kitty for is pretty?

“Oh look — a pretty kitty!”
“Look — there’s another one!”
Who does Lily think the kitty for is pretty?

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

“Every kitty didn’t sit on the stairs”

“No kitties sat on the stairs.”

“Not all kitties sat on the stairs.”

You know how to identify the right interpretation in context (pragmatics)

You know how to identify the right interpretation in context (pragmatics)

“Every kitty didn’t sit on the stairs”

“No kitties sat on the stairs.”

“Not all kitties sat on the stairs.”

You know how to identify the right interpretation in context (pragmatics)
Who does Lily think the kitty for is pretty?

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

“Every kitty didn’t sit on the stairs”
Not all kitties sat on the stairs.

speech segmentation

metrical phonology

syntactic categorization

pragmatics
So how exactly do children learn all this?
So how exactly do children learn all this?

We know they do it relatively quickly.

Much of the linguistic system is already known by age 4.
So how exactly do children learn all this?

They also don’t seem to get a lot of explicit instruction. And when they do, they don’t really pay attention to things that don’t impact meaning.

(From Martin Braine)

Child: Want other one spoon, Daddy.
Father: You mean, you want the other spoon.
Child: Yes, I want other one spoon, please Daddy.
Father: Can you say “the other spoon”?
Child: Other...one...spoon.
Father: Say “other”.
Child: Other.
Father: “Spoon.”
Child: Spoon.
Father: “Other spoon.”
Child: Other...spoon. Now give me other one spoon?
So how exactly do children learn all this?

They also don’t seem to get a lot of *explicit instruction*. And when they do, they *don’t really pay attention* to things that don’t impact meaning.

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.
We can also think about this as an information processing task.
We can also think about this as an information processing task.

Given the available input,

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

**Input**

*Where did he hide?*
*What happened?*
We can also think about this as an information processing task.

Given the available input, information processing done by human minds

Look at that kitty! There’s another one.

Where did he hide? What happened?
We can also think about this as an information processing task.

Given the available input, information processing done by human minds to build a system of linguistic knowledge.
We can also think about this as 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.

Input

Where did he hide? What happened?

Where’s the kitty?

That one’s really cute.
To understand how children solve this acquisition task, we need to think more about all the components involved.

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

Adapted from Lidz & Gagliardi 2015
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 using current knowledge of the language (the developing grammar)…
Perceptual encoding:
Involves using current knowledge of the language (the developing grammar) deployed in real time to parse the input...
Perceptual encoding:

Involves using current knowledge of the language (the developing grammar) deployed in real time to parse the input, often drawing on extralinguistic systems (like working memory, auditory processing, etc.)
Perceptual encoding

High vs. Mid vs. Low
Main vs. relative pitch
secondary stress

syllables with stress
(Mom)

speaker identity
Generating observable behavior

Involves the current linguistic representations and the developing grammar being used by the production system.
Generating observable behavior

These are used in real time to generate linguistic behavior (utterances) and non-linguistic behavior (pointing, looking, etc.). These behaviors require linguistic systems (utterance generation) and extralinguistic systems (motor control, attention, decision-making, etc.)

Adapted from Lidz & Gagliardi 2015
Inference = learning

This is how children learn from the current data in order to update the developing grammar.

Adapted from Lidz & Gagliardi 2015
Inference = learning

Constraints on children’s hypotheses and filters on their attention cause them to heed a subset of the perceptual intake — this is the acquisitional intake.
Adapted from Lidz & Gagliardi 2015

**perceptual intake**

- High vs. Mid vs. Low
- Main vs. relative pitch
- Secondary stress

**s**

wˈʌɾə rpˈɪɾ ˈɪɾ M M M

**syllables with stress**

(Mom)

**speaker identity**

**acquisitional intake**

**syllables with stress**

wˈʌɾə rpˈɪɾ ˈɪɾ kˈɪ ˈɪɾ

Parsing procedures

Extralinguistic systems

Developing grammar

Utterance generation

Extralinguistic systems

Inference

Constraints & filters

Extralinguistic systems

Perceptual intake
Inference = learning

Inference happens over the **acquisitional intake**, using **extralinguistic abilities** (statistical learning, probabilistic inference, hypothesis testing, etc.) ...
Inference = learning
Inference happens over the acquisitional intake, using extralinguistic abilities (statistical learning, probabilistic inference, hypothesis testing, etc.) to generate the most up-to-date ideas about the language’s grammar.
This whole process **happens over and over again** throughout the **learning period**
An informative computational model of language acquisition captures these important pieces in an *empirically-grounded* way.
When we have an informative computational model, it will connect the child’s input to the child’s output in just this way.
We can then look “under the hood” to see what internal pieces made that possible — this part is hard to do in real children’s minds!
Upshot: With computational modeling, we can understand more precisely how the learning strategies that children use work.
Some things we’ve learned by model-building this way

speech segmentation

= wəʔəpəɾiɾiɾi
what a pretty kitty!
Some things we’ve learned

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


$$P(s|u) \propto P(s)P(u|s)$$
Some things we’ve learned

The intuition of Bayesian inference (applied to speech segmentation)

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

The best answer (based on the utterance you just heard) ...

Some things we’ve learned

The intuitions of Bayesian inference (applied to speech segmentation)

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

The best answer (based on the utterance you just heard) depends on your prior beliefs about what good answers look like ...

Some things we’ve learned

The intuition of Bayesian inference (applied to speech segmentation)

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

The best answer (based on the utterance you just heard) depends on your prior beliefs about what good answers look like and how easily an answer explains the data observed in the utterance.

Some things we’ve learned

Bayesian inference

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

Strategy: Identify a list of word forms (= lexicon) that best generates the observable fluent speech utterances

Mathematically encoded preferences:

\[
\begin{array}{c}
\text{Mathematically encoded preferences:} \\
\text{\texttt{w\textsc{la}r\textsc{e}}}
\end{array}
\]

= \texttt{w\textsc{la}r\textsc{e}p\textsc{r}\textsc{i}r\textsc{i}r\textsc{i}}

= \texttt{w\textsc{la}r\textsc{e}p\textsc{r}\textsc{i}r\textsc{i}r\textsc{i}}

Some things we’ve learned

Bayesian inference

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

Strategy: Identify a list of word forms (= lexicon) that best generates the observable fluent speech utterances

Mathematically encoded preferences:

(1) Prefer shorter words

Some things we’ve learned

Bayesian inference

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

Strategy: Identify a list of word forms (= lexicon) that best generates the observable fluent speech utterances

Mathematically encoded preferences:

(1) Prefer shorter words
(2) Prefer lexicons with fewer words
Some things we’ve learned

Bayesian inference

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

Strategy: Identify a list of word forms (= lexicon) 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

Some things we’ve learned

Bayesian inference

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

Strategy: Identify a list of word forms (= lexicon) 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 preferences

Some things we’ve learned

Bayesian inference

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

Strategy: Identify a list of word forms (= lexicon) 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 preferences

and can generate the observable fluent speech utterances

Some things we’ve learned

Bayesian inference

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

✅ Is it **useful** for children?

Modeled learners without cognitive limitations on their inference and memory can use this strategy to segment fairly well when given realistic English child-directed speech data to learn from.

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

---

Some things we’ve learned

Bayesian inference

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

Is it useful?

Is it usable by children?

Modeled learners with cognitive limitations on their inference and memory can still use this strategy and segment English quite well.

Some things we’ve learned

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**

Some things we’ve learned

Bayesian inference

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

Is it useful? ✓

Is it useable? ✓

Does it work for different languages? ✓

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

Recap

Language acquisition is an interesting area of research in human cognition because it’s really hard and little humans are really good at it.
Recap

Language acquisition is interesting.

To understand how it works, we can build cognitive computational models that capture the important components of the process and then look inside to see exactly how they work.
Recap

Language acquisition is interesting

Models can capture important components and we can look inside

Some recent findings with this approach suggest Bayesian inference is a plausible early speech segmentation strategy that’s useful, useable, and works for many languages

\[ P(s|u) \propto P(s)P(u|s) \]
Thank you!

UC Computational Social Science

Who does... is pretty?

Every kitty didn’t...

another one

Noun

Kitty