Formally defining the learning setting for child language acquisition

Jordan Kodner
Sarah Payne
Stony Brook University

The 101st LSA
January 2025
Philadelphia

CLA × CLT: What do we mean?

CLA × CLT: What do we mean?

Computational Learning Theory (CLT)

- A subfield of computational theory and AI
- Developed as the formal side of machine learning, but formally describes *all* learning, including by biological systems

CLA × CLT: What do we mean?

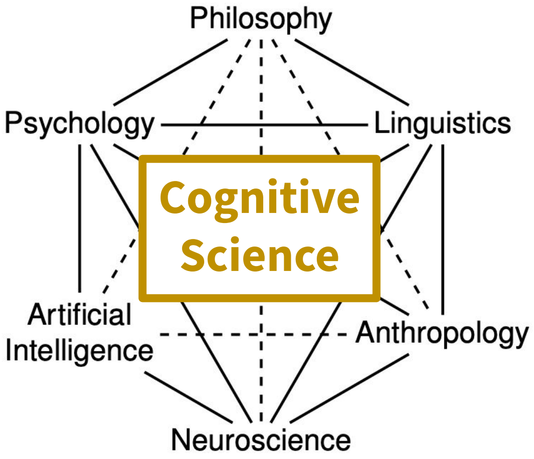
Computational Learning Theory (CLT)

- A subfield of computational theory and AI
- Developed as the formal side of machine learning, but formally describes *all* learning, including by biological systems

Child Language Acquisition (CLA)

- The (study of the) process by which children learn their native language(s)
- A unique learning task, subject to heavy study in linguistics and psychology, often with computational methods

CLA × CLT: What do we mean?



Computational Learning Theory (CLT)

- A subfield of **computational theory and AI**
- Developed as the formal side of machine learning, but **formally describes *all* learning, including by biological systems**

Child Language Acquisition (CLA)

- The (study of the) process by which children learn their native language(s)
- **A unique learning task**, subject to heavy study in **linguistics and psychology**, often with computational methods

Cognitive science combines these fields.

Can CLA and CLT be unified to elucidate linguistic cognition?

CLA × CLT: Why Combine Them?

From the perspective of Child Language Acquisition

- Reveal connections to other kinds of learning using shared formalisms
- Uncover explanatory gaps and new research pathways

CLA × CLT: Why Combine Them?

From the perspective of Child Language Acquisition

- Reveal connections to other kinds of learning using shared formalisms
- Uncover explanatory gaps and new research pathways

From the perspective of Computational Learning Theory

- Direct research effort at the heart of the problem of human language learning
- Offer a means to incorporate more empirical information into modeling

Our Proposal: Describe CLA in terms amenable to CLT

Easier said than done!

Our Proposal: Describe CLA in terms amenable to CLT

Easier said than done!

- CLA deals with real-world squishy biological entities
  → Well-described CLT frameworks do not neatly apply out-of-the-box

Our Proposal: Describe CLA in terms amenable to CLT

Easier said than done!

- CLA deals with real-world squishy biological entities
  → Well-described CLT frameworks do not neatly apply out-of-the-box
- CLA is really a constellation of semi-inter-dependent learning processes
  → Often not clear how to sensibly extract a self-contained problem for CLT

Our Proposal: Describe CLA in terms amenable to CLT

Easier said than done!

- CLA deals with real-world squishy biological entities
  → Well-described CLT frameworks do not neatly apply out-of-the-box
- CLA is really a constellation of semi-inter-dependent learning processes
  → Often not clear how to sensibly extract a self-contained problem for CLT

Today's goal:

We'll lay out some desiderata for a formalization of CLA

↔ CLA-able framework for CLT

What should be formally defined for CLA?

CLT defines the learning of some concept class
under some data presentation

What should be formally defined for CLA?

CLT defines the learning of some concept class under some data presentation

In terms more familiar to CLA and theoretical linguists:

- What does the input data look like?
  In terms of representations and distributions
- What does the output of learning look like?
  In terms of representation (the grammar) and how it manifests in behavior
  In terms of intermediate as well as final learner states

Desiderata:
The Input

The Primary Input: Language!

Language is the primary form of input to CLA

- Certainly there is other input as well (multi-sensory, world grounding…)
- But non-linguistic input is often apparently not necessary or even helpful

  Blind children follow nearly identical learning trajectories to seeing children[1]

  Blind adults demonstrate near-identical semantics for sight words[2]

  NLP/cog modeling systems don't necessarily benefit from multimodal input[3]

[1] Landau & Gleitman 1985, [2] Bedney et al. 2019, [3] Madasu & Lal 2023

The Primary Input: Language!

Language is the primary form of input to CLA

- Certainly there is other input as well (multi-sensory, world grounding…)
- But non-linguistic input is often apparently not necessary or even helpful

  Blind children follow nearly identical learning trajectories to seeing children[1]

  Blind adults demonstrate near-identical semantics for sight words[2]

  NLP/cog modeling systems don't necessarily benefit from multimodal input[3]

"The reports of my death are greatly exaggerated"

— The Primacy of Language Input

[1] Landau & Gleitman 1985, [2] Bedney et al. 2019, [3] Madasu & Lal 2023

The Primary Input: Language!

Language is the primary form of input to CLA

- Certainly there is other input as well (multi-sensory, world grounding…)
- But non-linguistic input is often apparently not necessary or even helpful
  Blind children follow nearly identical learning trajectories to seeing children[1]
  Blind adults demonstrate near-identical semantics for sight words[2]
  NLP/cog modeling systems don't necessarily benefit from multimodal input[3]

Implications for CLT

✔ (Except when demonstrated otherwise) language-only input is appropriate

[1] Landau & Gleitman 1985, [2] Bedney et al. 2019, [3] Madasu & Lal 2023

The Primary Input: Language!

Language is the primary form of input to CLA

- Certainly there is other input as well (multi-sensory, world grounding...)
- But non-linguistic input is often apparently not necessary or even helpful

  Blind children follow nearly identical learning trajectories to seeing children[1]

  Blind adults demonstrate near-identical semantics for sight words[2]

  NLP/cog modeling systems don't necessarily benefit from multimodal input[3]

Implications for CLT

✔ (Except when demonstrated otherwise) language-only input is appropriate

✔ Learning-focused math ling research is already on track for CLA 😀

[1] Landau & Gleitman 1985, [2] Bedney et al. 2019, [3] Madasu & Lal 2023

“The Poverty of the Stimulus”

Small, sparse, skewed input is a quantitative fact of CLA

- A particular challenge for any kind of naturalistic language learning
  → the argument from the Poverty of the Stimulus (Input Sparsity Problem)

More generally, for our purposes today:

✘ Simple, brute force, *tabula rasa* learning strategies will exhaust the input before successful learning

✔ Carefully selected representations + clever hypothesis generation or hypothesis search is necessary

✔ This is the kind of stuff CLT is great at!

“The Poverty of the Stimulus”

Small, sparse, skewed input is a quantitative fact of CLA

- A particular challenge for any kind of naturalistic language learning
  → the argument from the Poverty of the Stimulus (Input Sparsity Problem)
- Sparsity and skew are banes on NLP
  → the perpetual search for more, more, more training data

  But children can’t just add more training data…

“The Poverty of the Stimulus”

Small, sparse, skewed input is a quantitative fact of CLA

- A particular challenge for any kind of naturalistic language learning
  → the argument from the Poverty of the Stimulus (Input Sparsity Problem)
- Sparsity and skew are banes on NLP
  → the perpetual search for more, more, more training data
  But children can’t just add more training data…

Consequence:

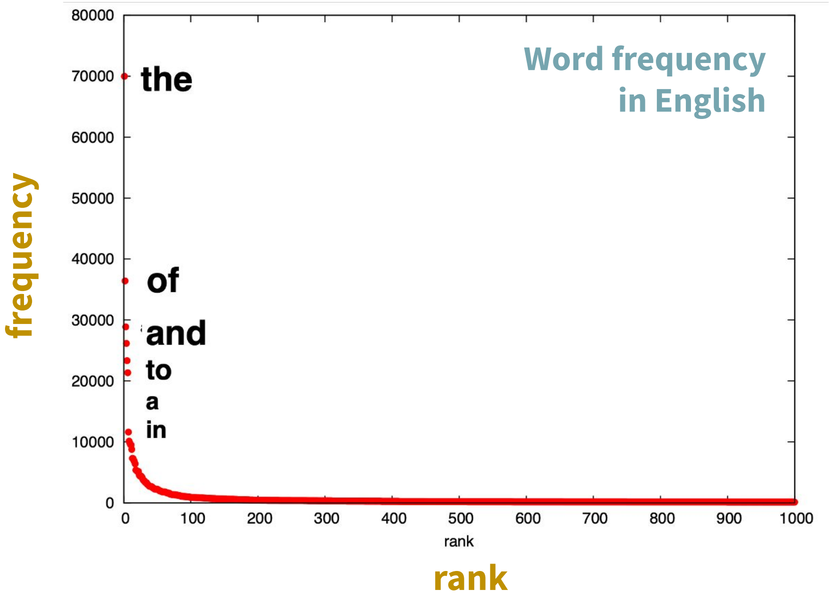
A CLA-relevant CLT framework should incorporate the Input Sparsity Problem

Input Distributions

Long-Tailed distributions are pervasive across language

- **The Zipfian (log-log) distribution is the most famous, but not the only one**

$$f \propto \frac{1}{r}$$

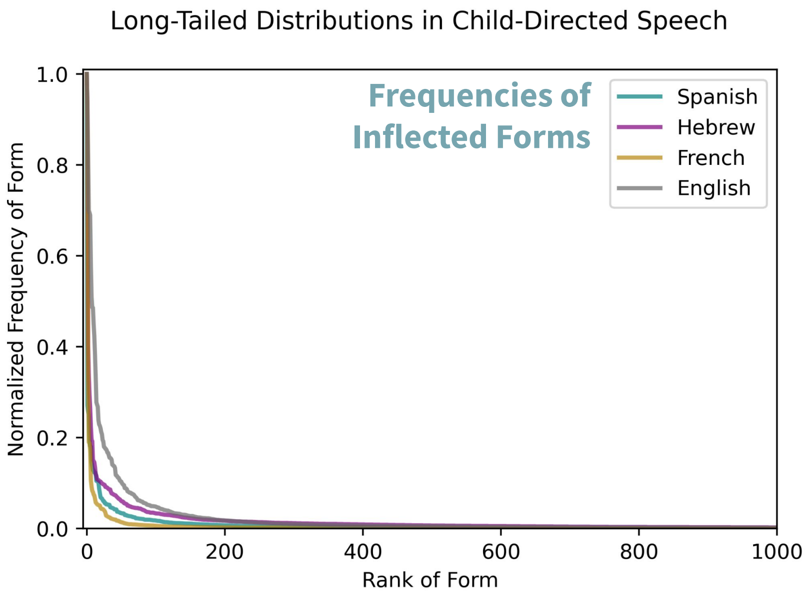


Input Distributions

Long-Tailed distributions are pervasive across language

- **The Zipfian (log-log) distribution is the most famous, but not the only one**

$$f \propto \frac{1}{r}$$



Input Distributions

Long-Tailed distributions are pervasive across language

- The Zipfian (log-log) distribution is the most famous, but not the only one
- It and similar distributions apply to many aspects of language:

Input Distributions

Long-Tailed distributions are pervasive across language

- The Zipfian (log-log) distribution is the most famous, but not the only one
- It and similar distributions apply to many aspects of language:

  Frequency distributions of:

    Words

    Parts of Speech

    Inflectional Categories

    Phonemes

    Syllables & their subcomponents…

Input Distributions

Long-Tailed distributions are pervasive across language

- The Zipfian (log-log) distribution is the most famous, but not the only one
- It and similar distributions apply to many aspects of language:

  Frequency distributions of:

    Words

    Parts of Speech

    Inflectional Categories

    Phonemes

    Syllables & their subcomponents…

  And of derived distributions:

    Morphological paradigm saturation, lexical attestation in constructions…

Input Distributions

Long-Tailed distributions are pervasive across language

- The Zipfian (log-log) distribution is the most famous, but not the only one
- It and similar distributions apply to many aspects of language
- They also apply outside of language:

SEARS Catalogue[1]

Number of items of like price

Number of items with a similar number of styles

Number of products per page

Number of pictures per product

...

[1] Zipf 1950, [2] Adamatzky 2022, [3] Newitz 2013, [4] Sorbaro et al. 2019, [5] Lin 2009

Input Distributions

Long-Tailed distributions are pervasive across language

- The Zipfian (log-log) distribution is the most famous, but not the only one
- It and similar distributions apply to many aspects of language
- They also apply outside of language:

  Sears Catalogue[1]

  Electrical Spiking among Fungi[2] 🍄⚡

  - Intervals between spikes
  - Average spike amplitude
  - Number of spikes in trains
  - …

[1] Zipf 1950, [2] Adamatzky 2022, [3] Newitz 2013, [4] Sorbaro et al. 2019, [5] Lin 2009

Input Distributions

Long-Tailed distributions are pervasive across language

- The Zipfian (log-log) distribution is the most famous, but not the only one
- It and similar distributions apply to many aspects of language
- They also apply outside of language:

  Sears Catalogue[1]

  Electrical Spiking among Fungi[2]

  Populations of Cities[3]

[1] Zipf 1950, [2] Adamatzky 2022, [3] Newitz 2013, [4] Sorbaro et al. 2019, [5] Lin 2009

Input Distributions

Long-Tailed distributions are pervasive across language

- The Zipfian (log-log) distribution is the most famous, but not the only one
- It and similar distributions apply to many aspects of language
- They also apply outside of language:

  Sears Catalogue[1]

  Electrical Spiking among Fungi[2]

  Populations of Cities[3]

  Neuronal Avalanches[4]

[1] Zipf 1950, [2] Adamatzky 2022, [3] Newitz 2013, [4] Sorbaro et al. 2019, [5] Lin 2009

Input Distributions

Long-Tailed distributions are pervasive across language

- The Zipfian (log-log) distribution is the most famous, but not the only one
- It and similar distributions apply to many aspects of language
- They also apply outside of language:

  Sears Catalogue[1]

  Electrical Spiking among Fungi[2]

  Populations of Cities[3]

  Neuronal Avalanches[4]

  Document Frequencies in Databases[5]

[1] Zipf 1950, [2] Adamatzky 2022, [3] Newitz 2013, [4] Sorbaro et al. 2019, [5] Lin 2009

Input Distributions

Long-Tailed distributions are pervasive across language

- The Zipfian (log-log) distribution is the most famous, but not the only one
- It and similar distributions apply to many aspects of language
- They also apply outside of language[1-5]

Consequence: Most input instances are ultimately redundant, uninformative, maybe even distracting?

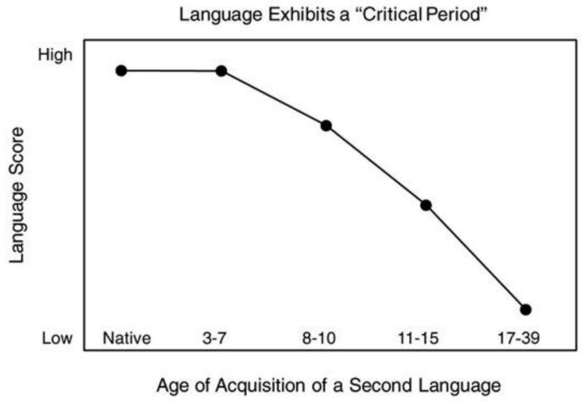
…But at least they aren't deliberately adversarial

[1] Zipf 1950, [2] Adamatzky 2022, [3] Newitz 2013, [4] Sorbaro et al. 2019, [5] Lin 2009

Small Data

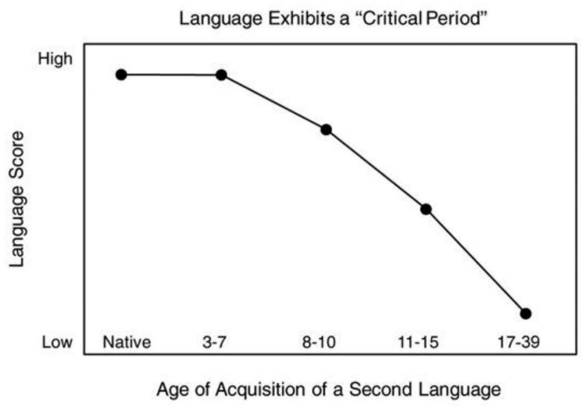
CLA “ends” eventually

- The Critical Period[1]: the grammar is acquired during CLA then remains nearly fixed
- There is no “cliff.” CLA trails off over time.
  We still learn certain aspects of language, like vocabulary, well in adulthood
- Most aspects of the grammar are acquired well before the end of the period



[1] following Lenneberg 1967

Small Data



CLA "ends" eventually

- **The Critical Period**[1]: the grammar is acquired during CLA then remains nearly fixed
- There is no "cliff." CLA trails off over time. We still learn certain aspects of language, like vocabulary, well in adulthood
- Most aspects of the grammar are acquired well before the end of the period

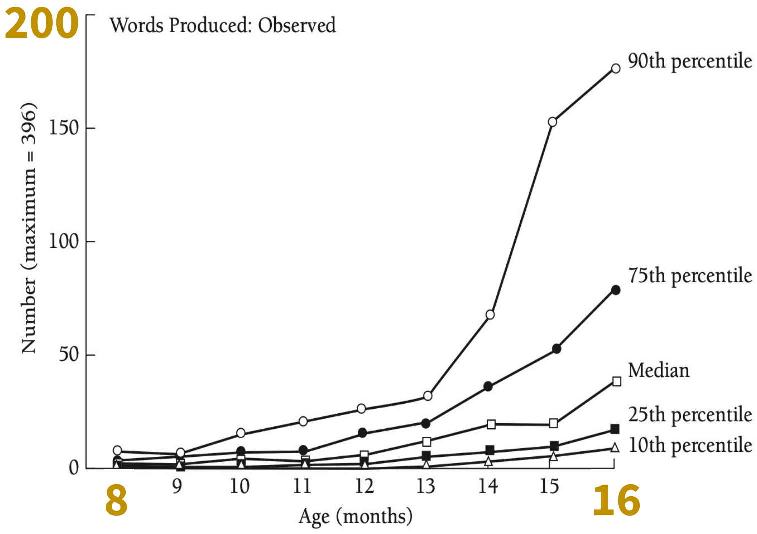
Consequences:

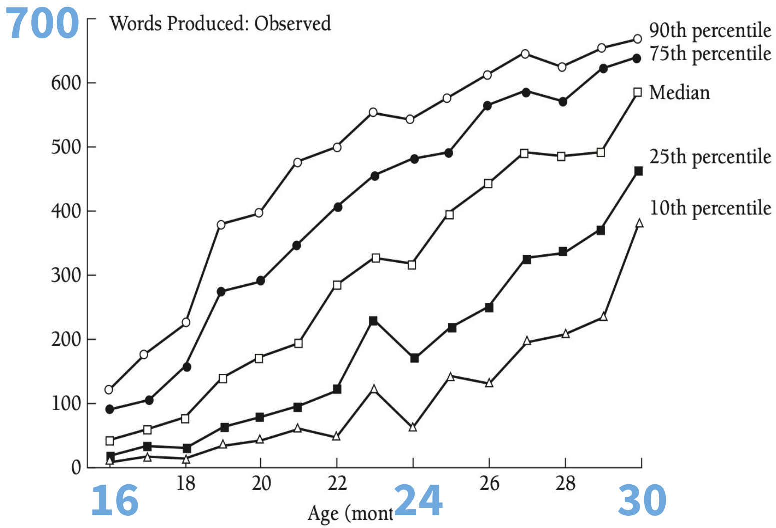
The input is not only finite but "small" by modern NLP standards.
Pieces of the grammar are acquired on different "small" input sizes

[1] following Lenneberg 1967

Small Data

- Learners receive on the order of 10 million tokens per year[1]
- Individual learner vocabularies grow over the course of development[2]



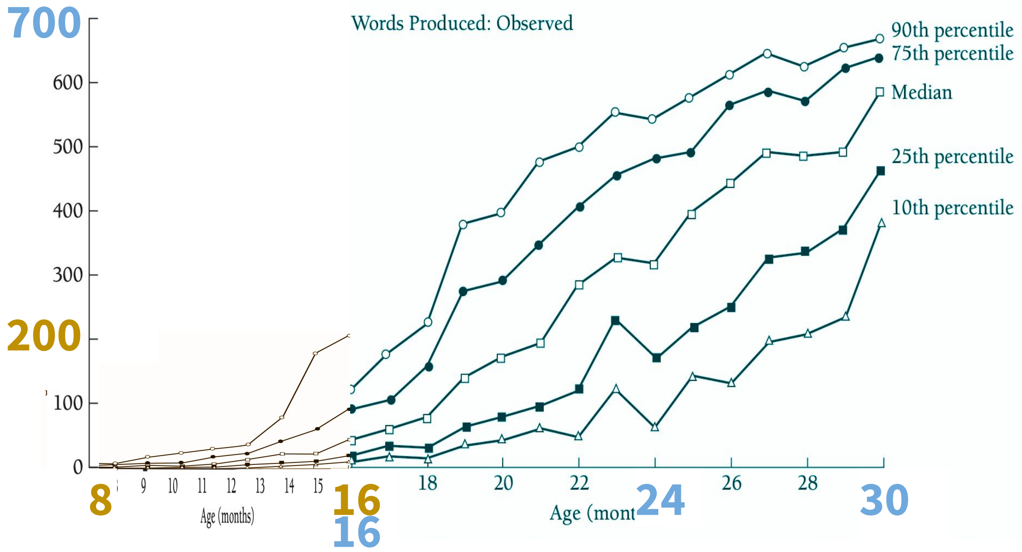


[1] Gilkerson et al. 2017, [2] Fenson et al 1994, Hart & Risley 2003, Plots from Fenson et al 1994

Small Data

- Learners receive on the order of 10 million tokens per year[1]
- Individual learner vocabularies grow over the course of development[2]

Our attempt to rescale these plots to match (the 90th percentile lines are inconsistent)



[1] Gilkerson et al. 2017, [2] Fenson et al 1994, Hart & Risley 2003, Plots from Fenson et al 1994

Variability in the Input

Children do not receive the same input instances

- Not even children in the same environment receive identical input

Variability in the Input

Children do not receive the same input instances

- **Not even children in the same environment receive identical input**

Measuring lexical overlap with Jaccard similarity [0-none, 1-complete]

The four children's subcorpora in the Providence Corpus range from

0.25 (Naima-William) to 0.37 (Ethan-William)[1]

[1] Richter 2021

Variability in the Input

Children do not receive the same input instances

- **Not even children in the same environment receive identical input**

Measuring lexical overlap with Jaccard similarity [0-none, 1-complete]

The four children's subcorpora in the Providence Corpus range from

0.25 (Naima-William) to 0.37 (Ethan-William)[1]

Comparable range to the Brown, Brent, and MacWhinney CDS corpora vs the ~1000 most frequent words in COCA genre corpora

0.21 (Academic'92-Brent) to 0.44 (Fiction'04-MacWhinney)[2]

[1] Richter 2021, [2] Kodner 2019

Variability in the Input

Children do not receive the same input instances

- Not even children in the same environment receive identical input[1,2]
- Though given input skew, high frequency instances are more likely to appear to more children and are more likely to appear early in the input sequence
- Despite all this, children exhibit stark uniformity in learning outcomes[3]

[1] Richter 2021, [2] Kodner 2019, [3] Labov 1972

Variability in the Input

Children do not receive the same input instances

- Not even children in the same environment receive identical input[1,2]
- Though given input skew, high frequency instances are more likely to appear to more children and are more likely to appear early in the input sequence
- Despite all this, children exhibit stark uniformity in learning outcomes[3]

Consequence:

CLA learners cannot generally assume specific instances will be present in the input, even though they can assume distributions. Outcomes should be robust to variation in the input

[1] Richter 2021, [2] Kodner 2019, [3] Labov 1972

Negative Evidence is a Non-Starter in CLA

Direct Negative Feedback

- Sparse and not consistently provided
- It's noisy — ever misunderstood a toddler?
- Famously ignored/misunderstood by children[1]

[1] Brown and Hanlon 1970, Braine 1971, Marcus 1992

Negative Evidence is a Non-Starter in CLA

Direct Negative Feedback

- Sparse and not consistently provided
- It's noisy — ever misunderstood a toddler?
- Famously ignored/misunderstood by children[1]

McNeill (1966)

Child: Nobody don't like me.
Mother: No, say "nobody likes me."
Child: Nobody don't like me.
[Eight repetitions of this dialogue follow.]
Mother: No, now listen carefully, say "NOBODY LIKES ME."
Child: Oh! Nobody don't likes me.

[1] Brown and Hanlon 1970, Braine 1971, Marcus 1992

Negative Evidence is a Non-Starter in CLA

Direct Negative Feedback

- Sparse and not consistently provided
- It's noisy — ever misunderstood a toddler?
- Famously ignored/misunderstood by children[1]

Child: My teacher holded the baby rabbits and we patted them.
Adult: Did you say your teacher held the baby rabbits?
Child: Yes.
Adult: What did you say she did?
Child: She holded the baby rabbits and we patted them.
Adult: Did you say she held them tightly?
Child: No, she holded them loosely.

Cazden (1972)

[1] Brown and Hanlon 1970, Braine 1971, Marcus 1992

Negative Evidence is a Non-Starter in CLA

Direct Negative Feedback

- Sparse and not consistently provided
- It's noisy — ever misunderstood a toddler?
- Famously ignored/misunderstood by children[1]

Indirect Negative Feedback

- Indistinguishable from accidental gaps[2]
  esp. given small, sparse, skewed input
- Indistinguishable from noisy implicit feedback[3]

[1] Brown and Hanlon 1970, Braine 1971, Marcus 1992, [2] Yang 2016, [3] Marcus 1993

Negative Evidence is a Non-Starter in CLA

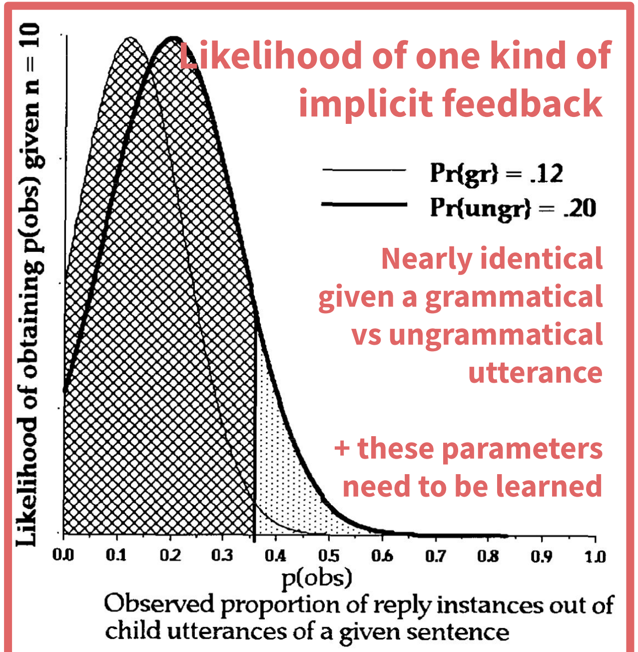
Direct Negative Feedback

- Sparse and not consistently provided
- It's noisy — ever misunderstood a toddler?
- Famously ignored/misunderstood by children[1]

Indirect Negative Feedback

- Indistinguishable from accidental gaps[2]
  esp. given small, sparse, skewed input
- Indistinguishable from noisy implicit feedback[3]

Marcus (1993)



[1] Brown and Hanlon 1970, Braine 1971, Marcus 1992, [2] Yang 2016, [3] Marcus 1993

Negative Evidence is a Non-Starter in CLA

Direct Negative Feedback

- Sparse and not consistently provided
- It's noisy — ever misunderstood a toddler?
- Famously ignored/misunderstood by children[1]

Indirect Negative Feedback

- Indistinguishable from accidental gaps[2]
  esp. given small, sparse, skewed input
- Indistinguishable from noisy implicit feedback[3]

Consequence:

CLA must proceed from positive evidence only

[1] Brown and Hanlon 1970, Braine 1971, Marcus 1992, [2] Yang 2016, [3] Marcus 1993

The Input to CLA: Summary

The input is small, sparse, and skewed

- Pieces of the grammar are acquired at different times
  → Specific problems call for specific input sizes (not learning in the limit)
- Long-tailed distributions are ubiquitous
  → Lots of redundant data. Informative inputs may be few and far between
- Specific input instances are highly variable between learners
  → CLA requires some substantial degree of robustness
- Negative evidence is hard to come by, unreliable, and ignored by learners

The Input to CLA: Summary

The input is small, sparse, and skewed

- Pieces of the grammar are acquired at different times
  → Specific problems call for specific input sizes (not learning in the limit)
- Long-tailed distributions are ubiquitous
  → Lots of redundant data. Informative inputs may be few and far between
- Specific input instances are highly variable between learners
  → CLA requires some substantial degree of robustness
- Negative evidence is hard to come by, unreliable, and ignored by learners

The big picture takeaway: Child language acquisition is fundamentally a game of generalization from positive examples

The Input to CLA: Summary

The input is small, sparse, and skewed

- Pieces o
  → Specif g in the limit)
- Long-tai
  → Lots o between
- Specific
  → CLA re
- Negativ learners

The Big Caveat:

None of this is to say that writing proofs about learning in the limit, guarantees about learning under any arbitrary distribution, etc., isn't worthwhile for linguists. Far from it.

It just isn't strictly CLA × CLT

The big picture takeaway: Child language acquisition is fundamentally a game of generalization from positive examples

Desiderata:
The Output

Representations Matter

CLA takes in language input and outputs a grammar. More formally,

$$h: L(G) \rightarrow G$$

Language acquisition is a function that takes
some language generated by a grammar as input and yields some grammar as output

Representations Matter

CLA takes in language input and outputs a grammar. More formally,

$$h: L(G) \rightarrow G$$

Language acquisition is a function that takes
some language generated by a grammar as input and yields **some grammar** as output

L(G) is what we've been talking about as "the input"

Representations Matter

CLA takes in language input and outputs a grammar. More formally,

$$h: L(G) \rightarrow G$$

Language acquisition is a function that takes some language generated by a grammar as input and yields some grammar as output

$$h(L(g_1)) = g_2, \quad g_1, g_2 \in G$$

A learner whose input was generated by a grammar g_1 should acquire grammar g_2

Representations Matter

CLA takes in language input and outputs a grammar. More formally,

$$h: L(G) \rightarrow G$$

Language acquisition is a function that takes some language generated by a grammar as input and yields some grammar as output

$$h(L(g_1)) = g_2, \quad g_1, g_2 \in G$$

A learner whose input was generated by a grammar g_1 should acquire grammar g_2

- **This characterization may be obvious, but it leaves far too much unsaid**
- **As we all know: the nature of h and G really matters**
- **Less thought about: So does relationship between g_1 and g_2!**

Representations Matter

What is G?: A shared interest

- Theoretical linguistics is mostly about cognitive representations
- Many CLA research cares about representations
- Formal linguistics & CLT demand careful thought about representations

Representations Matter

What is G?: A shared interest

- Theoretical linguistics is mostly about cognitive representations
- Many CLA research cares about representations
- Formal linguistics & CLT demand careful thought about representations

What is *h*?: A way forward

- CLT stands to direct CLA more towards representations
- CLA and CLT add the dimension of learnability to theoretical linguistics

What is Successful Learning?

The obvious answer

$$h(L(g_t)) = g_h, \quad g_t = g_h$$

A learner whose input was generated by a grammar g_t should acquire grammar g_h

- **Learning succeeds when the grammar that the child hypothesizes is "the same" as the target grammar**

What is Successful Learning?

The obvious answer

$$h(L(g_t)) = g_h, \quad g_t = g_h$$

A learner whose input was generated by a grammar g_t should acquire grammar g_h

- Learning succeeds when the grammar that the child hypothesizes is "the same" as the target grammar

✔ Reasonable, practical, works under many definitions of "the same."

✘ Makes several unrealistic assumptions

What is Successful Learning?

- **Does "sameness" between g_1 and g_2 mean extensional equivalence, or a more intensional notion of equivalence?**

$$h(L(g_1)) = g_h, \quad L(g_t) = L(g_h) \quad \text{or} \quad g_t \equiv g_h \ ?$$

✔ **CLT is well-equipped to deal with notions of equivalence**

What is Successful Learning?

- **Which formal measure of "sameness" should be used?**

What is Successful Learning?

- Which formal measure of "sameness" should be used?
- Input is not really drawn from a single target grammar, but rather a mix

$h(L(g_1) \cup L(g_2) \ldots \cup L(g_n)) = \quad g_h \quad \text{or} \quad \{g_{h1}, g_{h2}, \ldots, g_{hm}\} \quad ?$

Variation is a ubiquitous fact about the input to CLA

From production and perception noise, individual differences, sociolinguistics

✔ Learning from noisy input is a heavily studied concept in CLT

What is Successful Learning?

- Which formal measure of "sameness" should be used?
- Input is not really drawn from a single target grammar, but rather a mix

$h(L(g_1) \cup L(g_2) \ldots \cup L(g_n)) = \quad g_h \quad \text{or} \quad \{g_{h1}, g_{h2}, \ldots, g_{hm}\} \quad ?$

Variation is a ubiquitous fact about the input to CLA

From production and perception noise, individual differences, sociolinguistics

✔ Learning from noisy input is a heavily studied concept in CLT

? Should the learner acquire one mega-grammar or many grammars?

Depending on G, what if no single grammar can cover the entire input?

What is Successful Learning?

- Which formal measure of "sameness" should be used?
- Input is not really drawn from a single target grammar, but rather a mix

$$h(L(g_1) \cup L(g_2) \ldots \cup L(g_n)) = \quad g_h \quad \text{or} \quad \{g_{h1}, g_{h2}, \ldots, g_{hm}\} \quad ?$$

Variation is a ubiquitous fact about the input to CLA

From production and perception noise, individual differences, sociolinguistics

✔ Learning from noisy input is a heavily studied concept in CLT

? Should the learner acquire one mega-grammar or many grammars?

Depending on G, what if no single grammar can cover the entire input?

? How does this direct our notions of appropriate measures of equivalence?

What is Successful Learning?

- Which formal measure of "sameness" should be used?
- Input is not really drawn from a single target grammar, but rather a mix
- Grammars that are not formally equivalent may be practically equivalent

$h(L(g_1)) = g_h, \quad L(g_t) \approx L(g_h)$ or $g_t \approxeq g_h$?

If the differences in their extensions virtually never appear in their outputs or g_1 and g_2 parse an utterance differently in a way that causes no confusion

✔ "Virtually never appear" could be handled with a "close enough" error term

✘ Semantic equivalence isn't good enough for "causes no confusion"

→ May be one situation where language-only isn't good enough…

What is Successful Learning?

Some recommendations for measuring success

- Characterize learning through noise
- Relax the idea of matching a single target grammar

→ Formalize “close enough” learning

What is Successful Learning?

Some recommendations for measuring success

- Characterize learning through noise
- Relax the idea of matching a single target grammar

→ Formalize “close enough” learning

✔ Consider abductive learning ← we like it 😉[1]

✔ Check how “close enough” manifests empirically through studies of language in use and of child development

[1] Belth et al. 2021, Payne 2022, Yang et al. *in prep*

Learning Trajectories

CLA is Online/Incremental

- As opposed to batch learning
- Intermediate hypotheses/learner states are important
- Much can be inferred about intermediate states from learner behavior

Across all levels of the grammar

Both observational/corpus and experimental methodologies

$$h(\mathrm{L}(g_t)) = \langle g_{h1}, g_{h2}, \ldots, g_{hn} \rangle$$

The learner should pass through a sequence of hypothesis grammars

Learning Trajectories

Remember French liaison just a bit ago??

CLA is Online/Incremental

- As opposed to batch learning
- Intermediate hypotheses/learner states are important
- Much can be inferred about intermediate states from learner behavior

Across all levels of the grammar

Both observational/corpus and experimental methodologies

$$h(L(g_t)) = \langle g_{h1}, g_{h2}, \ldots, g_{hn} \rangle$$

The learner should pass through a sequence of hypothesis grammars

Learning Trajectories

CLA is Online/Incremental

- As opposed to batch learning
- Intermediate hypotheses/learner states are important
- Much can be inferred about intermediate states from learner behavior

Consequence: **Online learning is most relevant for CLA × CLT.**
A fully successful CLT perspective should account for intermediate states, not just the final state

The Output of CLA: Summary

The output is a (sequence of) grammar(s), but evaluation is obscure

- The simplest success metric is not necessarily appropriate
- CLA is online and incremental

  → intermediate states should be taken into account

  → good news if you like algorithmic thinking! ← we do 😉
- Learning may be “good enough”

  → abductive learning is a good idea!

  → robustness to input noise (regardless of its origin) is crucial

The Output of CLA: Summary

The output is a (sequence of) grammar(s), but evaluation is obscure

- The simplest success metric is not necessarily appropriate
- CLA is o
  → interm
  → good r
- Learning
  → abduc
  → robustness to input noise (regardless of its origin) is crucial

The Big Caveat:

We aren't advocating against abstracting the problem of CLA. Rather, that it is important to sometimes revisit the consequences of abstraction

Too much takes us away from CLA x CLT

Conclusions

The Input to CLA: Pros and Cons for CLT Research

The Cons:

- Can't assume conveniently selected data presentations
  in terms of ordering (e.g., regulars first) or completeness (e.g., full paradigms)
- Must assume finite input that is probably smaller than we'd hope for
- Negative examples would be helpful, but we're out of luck

The Input to CLA: Pros and Cons for CLT Research

The Cons:

- Can't assume conveniently selected data presentations
  in terms of ordering (e.g., regulars first) or completeness (e.g., full paradigms)
- Must assume finite input that is probably smaller than we'd hope for
- Negative examples would be helpful, but we're out of luck

The Pros:

- Don't have to prove learning under arbitrary data presentations
  Often assume long-tailed distributions that are not deliberately adversarial
- Language input (generally streams of symbols) is reasonable to assume 😅

The Output of CLA: Pros and Cons for CLT Research

The Cons:

- Noisy data, non-exact-match evaluation, and the possibility that no single grammar accounts for the entire the input all make life more complicated
- So do incremental learning and caring about learning trajectories
- ...and every piece of the grammar is going to have different requirements

The Output of CLA: Pros and Cons for CLT Research

The Cons:

- Noisy data, non-exact-match evaluation, and the possibility that no single grammar accounts for the entire the input all make life more complicated
- So do incremental learning and caring about learning trajectories
- ...and every piece of the grammar is going to have different requirements

The Pros:

- Incremental learning is good news if you like algorithmic thinking
- Intermediate states provide a wealth of evidence about the learner
  Both in terms of the hypothesis space and the learning strategy

A Blatant Non-Conclusion

We.DU.EXCL have made progress towards a formalization of CLA,
but we are still far from proposing a gold standard

→ We hope this talk inspires some discussion!

- What other desiderata are there for a formalization of CLA?
  e.g., Order of acquisition tells us what prior info learners can access
  The input doesn't contain underlying representations
  How does adult variation (French liaison again...) reflect imperfect convergence during learning?...
- What other important considerations are there for CLA x CLT?
- Where have we missed the mark?

Formally defining the learning setting for child language acquisition

iACS
INSTITUTE FOR ADVANCED COMPUTATIONAL SCIENCE

Stony Brook University

Formally defining the learning setting

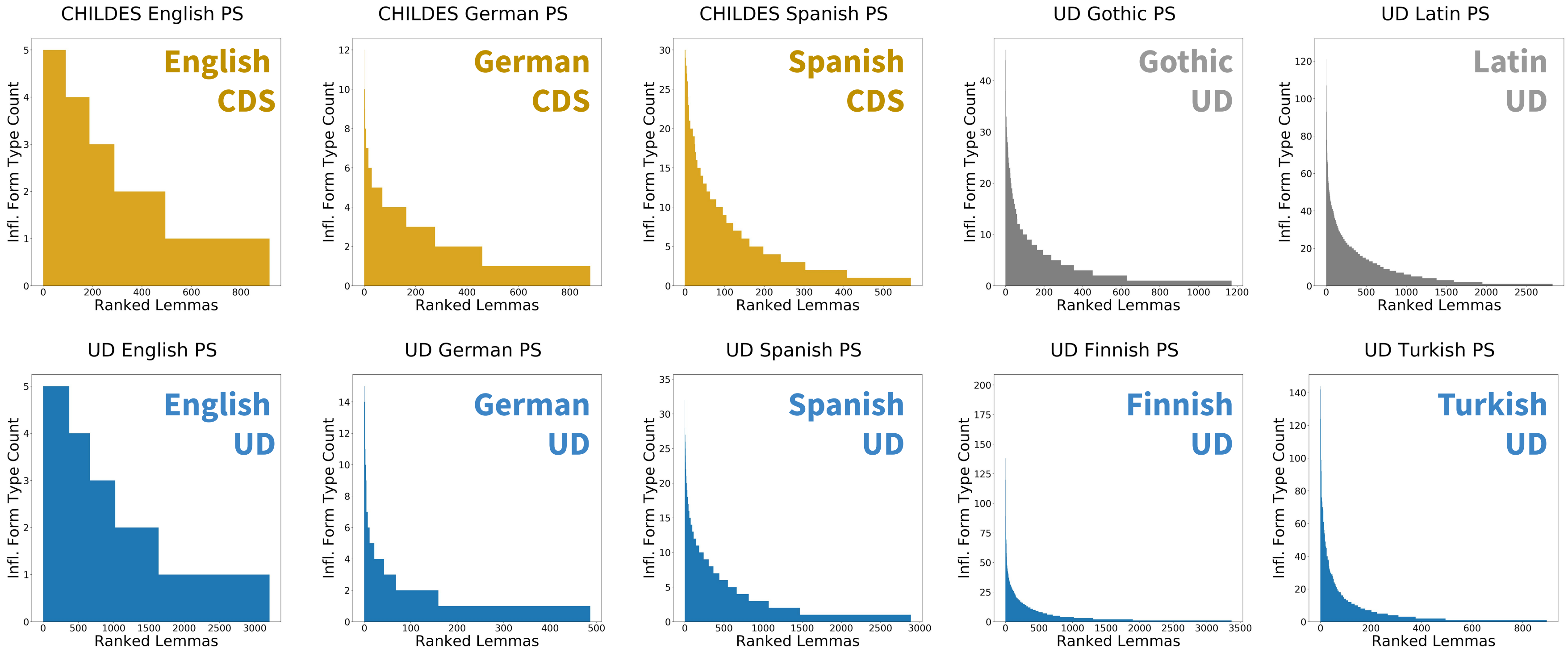
for child language acquisition

Jordan Kodner
Sarah Payne
Stony Brook University

The 101st LSA
January 2025
Philadelphia

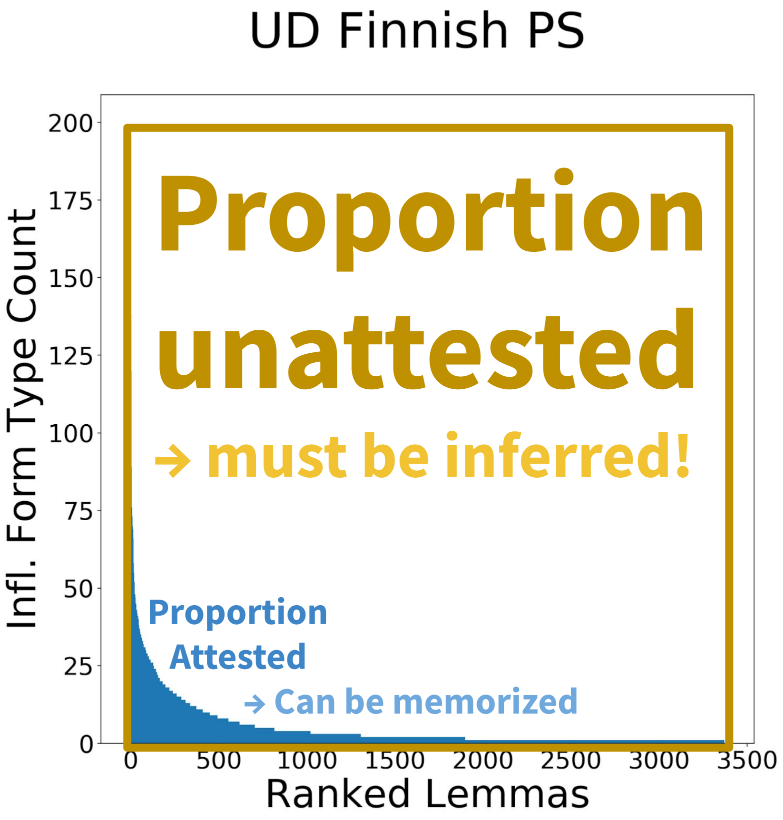
Desiderata:
The Output

A Concrete Example: Paradigm Saturation[1]

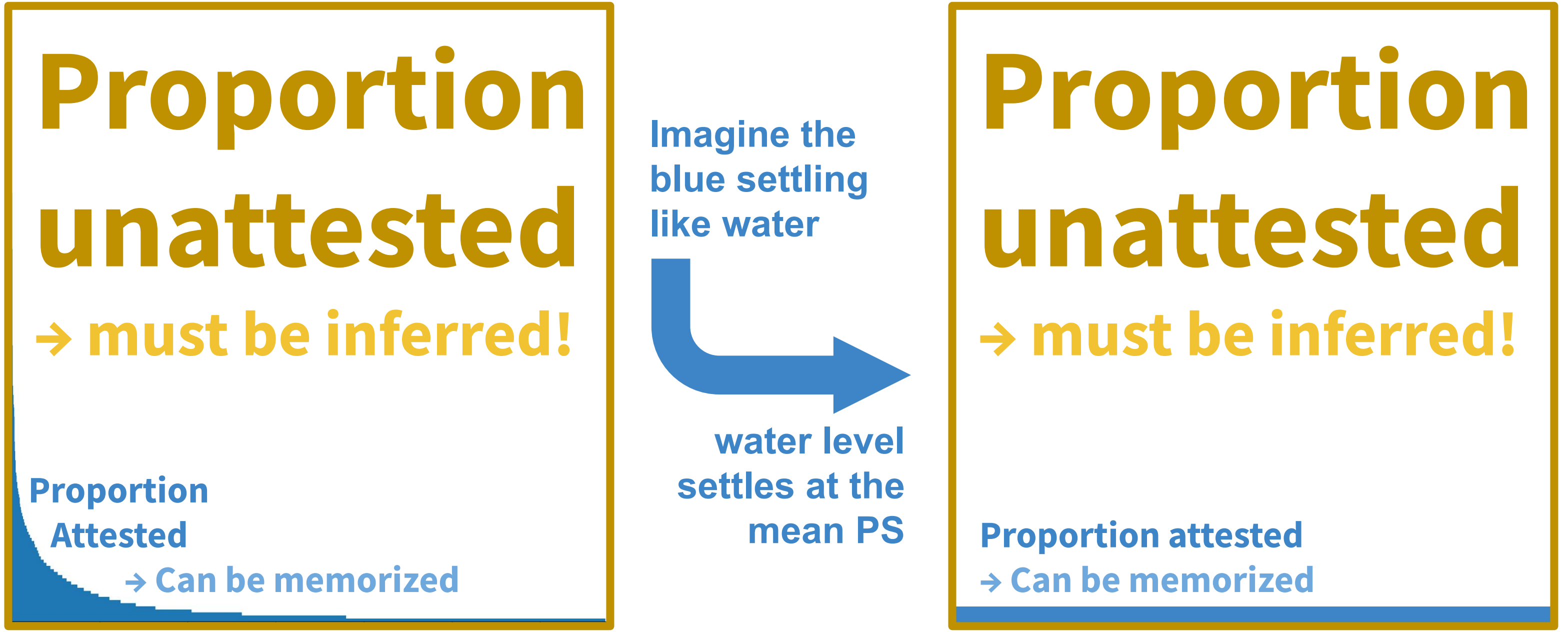


[1] Chan 2008, Lignos & Yang 2016, Payne 2022, Kodner 2023

A Concrete Example: Paradigm Saturation



A Concrete Example: Paradigm Saturation



A Concrete Example: U-Shaped Learning

English Past Tense Phase 1

- Past tense is marked inconsistently
- But is formed accurately when used

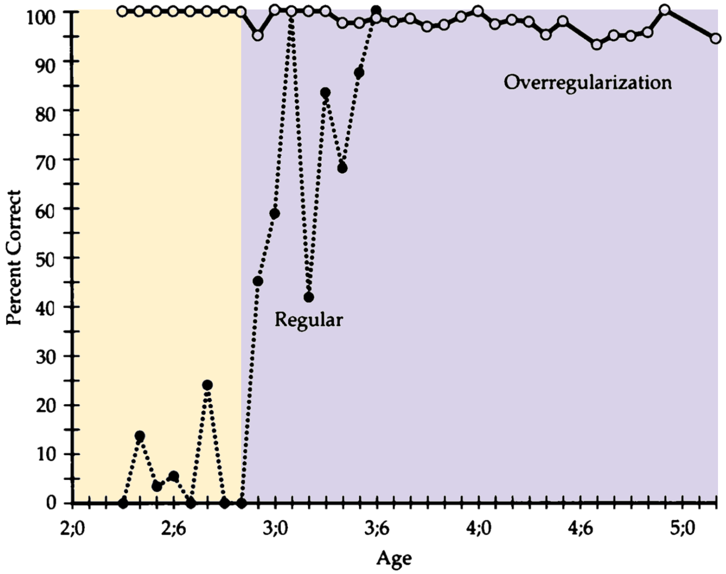
Phase 2 The U-Shape

- Past tense is marked consistently
- But irregulars are sometimes regularized
- Error rate declines over time

Interpretation: Phase 1 → Phase 2

Indicates a new hypothesis $g_{h1} \rightarrow g_{h2}$

Adam from the Brown Corpus
As visualized in Marcus (1993)



A Concrete Example: U-Shaped Learning

English Past Tense Phase 1

- Past tense is marked inconsistently
- But is formed accurately when used

Phase 2 The U-Shape

- Past tense is marked consistently
- But irregulars are sometimes regularized
- Error rate declines over time

Interpretation: Phase 1 → Phase 2

Indicates a new hypothesis g_{h1} → g_{h2}

Similar observation in Spanish verbs