Psych 245A: Computational Models of Language Learning

Lecture 8
Pullum & Scholz (2002)

Poverty of the Stimulus

Language
Can be thought of as the set of legal items in the language (sentences, strings, etc.). The child’s job: figure out the rules that generate that legal set and don’t generate illegal items.

Legal Items

A fairy who flies around the Labyrinth walls bites anyone who passes by.

Can the girl who can summon the Goblin King solve the Labyrinth?

Illegal Items

Hoggle is an ornery dwarf Can the girl who can summon the Goblin King solve the Labyrinth?

Poverty of the Stimulus

Standard Theory: Poverty of the Stimulus

The Logic of Poverty of the Stimulus (The Logical Problem of Language Acquisition)
1) Suppose there is some data.
2) Suppose there is an incorrect hypothesis compatible with the data.
3) Suppose children behave as if they never entertain the incorrect hypothesis.

Addendum (interpretation): Or children converge on the correct hypothesis much earlier than expected (Legate & Yang 2002).

Conclusion: Children possess innate knowledge ruling out the incorrect hypothesis from the hypothesis space considered.

Addendum (interpretation): The initial hypothesis space does not include all hypotheses. Specifically, the incorrect ones of a particular kind are not in the child’s hypothesis space.

Poverty of the Stimulus

The argument for having innate biases to guide language learning

Addendum: The data available to the child are compatible with a number of generalizations. However, children only seem to pick the right ones. Therefore, they must have some other constraints guiding their language learning.

The innate part: The guiding information must be available prior to learning.
Poverty of the Stimulus: More Details

Jareth can alter time. Can Jareth alter time? Rule?

Rule: Move first auxiliary?

Poverty of the Stimulus: More Details

Anyone who can wish away their brother would be tempted to do it. Would anyone who can wish away their brother be tempted to do it?
Jareth can alter time.

Can Jareth alter time?

Rule: Move first auxiliary?

Rule: Move last auxiliary?

Anyone who can wish away their brother would be tempted to do it. Would anyone who can wish away their brother be tempted to do it?

Rule: Move first auxiliary?

Rule: Move last auxiliary?

Someone who can solve the labyrinth can show someone who can't how. Can someone who can solve the labyrinth show someone who can't how?

Rule: Move first auxiliary?

Rule: Move last auxiliary?

Someone who is not easily fooled could trick someone who is. Could someone who is not easily fooled trick someone who is?
Jareth can alter time. Can Jareth alter time?

Anyone who can wish away their brother would be tempted to do it. Would anyone who can wish away their brother be tempted to do it?

Someone who can solve the labyrinth can show someone who can’t how. Can someone who can solve the labyrinth show someone who can’t how?

Someone who is not easily fooled could trick someone who is. Could someone who is not easily fooled trick someone who is?

Idea: Look at structure, not just linear order

Rule: Move main clause auxiliary

Learning bias: try structure-dependent rules

Poverty of the Stimulus: Data

Induction Problem: Logical Problem of Language Acquisition (Standard Theory)

Children don’t usually get access to all the data we just saw by the time they have the correct generalization (move main clause auxiliary). They learn from a subset of the legal items in the language. And still they seem to converge on the right generalizations… without trying out many (or all) of the wrong ones.

Items Encountered

Can the girl solve the Labyrinth?

Compatible with move first, move second, move last, move odd one counting from the beginning.

Can the girl who can summon the Goblin King solve the Labyrinth?
“…linguistic nativism is the view…that human infants have at least some linguistically specific innate knowledge”

“…issue is whether a full description of that predisposition incorporates anything that entails specific contingent facts about natural languages”

[poverty of the stimulus]

“…argument…turns on the claim that during the language acquisition process, children often come to know things about the language they are acquiring despite not having access to the crucial evidence that shows these things to be true of the language.”


Children’s observable accomplishments

Speed: Children learn so fast.
Reliability: Children always succeed.
Productivity: Children learn a system.
Selectivity: Children pick the correct option from a bunch of incorrect (and “seductive”) alternatives.
Underdetermination: Children arrive at systems of knowledge underdetermined by the data.
Convergence: Children end up with the right system.
Universality: The system acquired has a lot of properties in common with other language systems of the world.


“Instead of clarifying the reasoning, each successive writer on this topic shakes together an idiosyncratic cocktail of claims about children’s learning of language, and concludes nativism is thereby supported. Most of the frequently encountered claims are about children’s observable accomplishments or aspects of the child’s environment.”


Children’s observable accomplishments

Speed: Children learn so fast.
Reliability: Children always succeed.
Productivity: Children learn a system.
Selectivity: Children pick the correct option from a bunch of incorrect (and “seductive”) alternatives.
Underdetermination: Children arrive at systems of knowledge underdetermined by the data.
Convergence: Children end up with the right system.
Universality: The system acquired has a lot of properties in common with other language systems of the world.

Children’s observable accomplishments

**Speed**: Children learn so fast.

  Relevant Interpretation: Faster than expected, given available data.

**Selectivity**: Children pick the correct option from a bunch of incorrect (and “seductive”) alternatives.

  Relevant Interpretation: Seductive because also compatible with data.

**Underdetermination**: Children arrive at systems of knowledge underdetermined by the data.

  Relevant Interpretation: Alternative hypotheses also compatible with data.


Aspects of Child’s Environment

**Ingratitude**: No explicit payoff for correct language usage.

**Finiteness**: Children don’t get infinite data to learn from.

**Idiosyncracy**: The subset of data children encounter varies from child to child.

**Incompleteness**: Children don’t hear everything in the language.

**Positivity**: No explicit instruction of what isn’t in the language.

**Degeneracy**: Input to children has noise.
Pullum & Scholz (2002): The Version Chosen To Attack

“People attain knowledge of the structure of their language for which no evidence is available in the data to which they were exposed as children.” - Hornstein & Lightfoot (1981)

“We replace total lack of evidence by lack of evidence that is adequate to the task…would not emerge in conversational data near often enough to guarantee that any particular child would ever encounter it.” - Pullum & Scholz

“…’the APS’ to stand for ‘the Argument Selected by Pullum & Scholz’”

Pullum & Scholz (2002): How to Support APS

Step 1: Describe in detail what is known.

Step 2a: Identify the crucial data that would lead a data-driven learner to that knowledge.
Step 2b: Given reason to believe that’s the crucial data.

Step 3: Show learners don’t have access to that crucial data.
One way: Look for really rare data types. These are likely to be close enough to absent.

Step 4: Show that learners nonetheless acquire the right knowledge.

Pullum & Scholz (2002): Case Studies

Case 1: Plurals in noun-noun compounds

3-6 yr olds behavior:
Irregular plural pattern (plural marker on first noun okay)
1 tooth-eater or 1 teeth-eater
1 mouse-eater or 1 mice-eater

Regular plural pattern (plural marker on first noun not okay)
1 toy-eater (but not 1 toys-eater)
1 rat-eater (but not 1 rats-eater)
Case 1: Plurals in noun-noun compounds

Knowledge of incomplete paradigm:
- tooth-eater
- teeth-eater
- toy-eater
- *toys-eater

Important point: No generalization to regular plural nouns.

Gordon (1986): Brown corpus (1,000,000 words):
- irregular sg compounds [tooth-eater] (153 tokens)
- irregular pl compounds [teeth-eater] (3 tokens)
- regular sg compounds [toy-eater] (…more…?)
- regular pl compounds [*toys-eater] (0 tokens)

Argument: Irregular pl compounds appear so rarely, they are similar in frequency to regular pl compounds (which never appear because they’re ungrammatical.) But children still produce the irregular pl compounds and do not produce the irregular sg compounds. This is hard to explain if they’re data-driven.

P&S rebuttals:
- Not clear 3-6 yr old behavior was really true outside the experimental setup (method flaws).
- Point: Not children’s behavior.
- Point: Not adult’s behavior either.

Case 2: Auxiliary sequences

Kimball 1973:
- It rains, It may rain, It may have rained, It may be raining, It has rained, It has been raining, It is raining, It may have been raining,…

Rule: (...Aux Verb: {rains, may rain, may have rained, …})
- Aux --> Tense (Modal) (have +en) (be +ing)
- ({rains, may, might} have VERBed) (be VERBing)
Kimball 1973:
It rains, It may rain, It may have rained, It may be raining. It has rained, It has been raining, It is raining, It may have been raining, ...

Rule: (...) Aux Verb: {rains, may rain, may have rained, ...}
Aux → Tense (Modal) (have +en) (be +ing)
(present) (may, might) (have VERBen) (be VERBing)
(present) + rain = rains

Kimball 1973:
It rains, It may rain, It may have rained, It may be raining, It has rained, It has been raining, It is raining, It may have been raining, ...

Rule: (...) Aux Verb: {rains, may rain, may have rained, ...}
Aux → Tense (Modal) (have +en) (be +ing)
(present) (may, might) (have VERBen) (be VERBing)
(present) + (Modal) = may rain

Kimball 1973:
It rains, It may rain, It may have rained, It may be raining. It has rained, It has been raining, It is raining, It may have been raining, ...

Rule: (...) Aux Verb: {rains, may rain, may have rained, ...}
Aux → Tense (Modal) (have +en) (be +ing)
(present) (may, might) (have VERBen) (be VERBing)
(past) + (Modal) + (have + en) = may have rained

Kimball 1973:
It rains, It may rain, It may have rained, It may be raining, It has rained, It has been raining, It is raining, It may have been raining, ...

Rule: (...) Aux Verb: {rains, may rain, may have rained, ...}
Aux → Tense (Modal) (have +en) (be +ing)
(present) (may, might) (have VERBen) (be VERBing)
(past) + (have + en) = had rained
Pullum & Scholz (2002): Case Studies
Case 2: Auxiliary sequences

Kimball 1973:
It rains, It may rain, It may have rained, It may be raining. It has rained, It has been raining, It is raining, It may have been raining,…

Rule: (...Aux Verb: {rains, may rain, may have rained, …}  
Aux --> Tense (Modal)  (have +en) (be +ing)  
(presen[t], may, might) (have VERBen) (be VERBing)  
(past) + (Modal) + (be+ing) = may have been raining

Pullum & Scholz (2002): Case Studies
Case 2: Auxiliary sequences

Kimball 1973:
Aux --> Tense (Modal)  (have +en) (be +ing)

Crucial data to get proper rule sequence have all three optional components:  
“It may have been raining”  
No examples in 1,000,000 word corpus, vanishingly rare in conversation.

Pullum & Scholz (2002): Case Studies
Case 2: Auxiliary sequences

P&S rebuttal:
Is that rule really what children are acquiring? Instead, children may be able to abstract the necessary sequence from other sequences of not exactly that type.

Also, data not so vanishingly rare: hundreds of examples in adult literature (Moby Dick, Wuthering Heights) and many in children’s literature (Peter Pan, Alice in Wonderland, The Wizard of Oz)  
Estimate: 1 approximately every 3000-4000 sentences

The real question: How many is enough? Need a quantitative claim from the linguists.

Pullum & Scholz (2002): Case Studies
Case 3: Anaphoric One

Originally, Baker 1978
Recently described accessibly in Foraker et al. (2007)

“I liked the debate about learning. You liked the one about modeling.”  
“I’ll walk by the side of the road and you can walk by the one of the river.”

Syntactic distribution distinction: Difference between complement-taking nouns (side) and modifier-taking nouns (debate).  
Syntactic distinction: complement-taking nouns = Category N, modifier-taking nouns = Category N (larger than N)
Semantic distinction: Complement-taking nouns are conceptually different from modifier-taking nouns. (side = side of what?); debate can stand by itself.
Pullum & Scholz (2002): Case Studies

Case 3: Anaphoric One

Originally, Baker 1978

Necessary Data:
Baker 1978: rule out one = N

Need specific sentence & world situation:
“Jack wants a red ball. Lily doesn’t have one to give him.”
(Situation: Lily doesn’t have a red ball. She has a purple one.)
Case 3: Anaphoric One

Unambiguous data is pretty rare in child-directed speech: (Lidz, Waxman, & Freedman 2003; Pearl & Lidz, submitted)

Rebuttal of another kind: this is not the crucial evidence
"Can be learned from other available data": Reger & Gahl 2004, Foraker et al. 2007
"...Not without some other learning constraints/knowledge, too": Foraker et al. 2007, Pearl & Lidz, submitted

Case 4: Auxiliary Fronting

Chomsky 1971: Adult Knowledge

The girl is easily fooled.
Is the girl easily fooled?

The girl who can solve the labyrinth is easily fooled.
Is the girl who can solve the labyrinth easily fooled?

Someone who is not easily fooled could trick someone who is.
Could someone who is not easily fooled trick someone who is?

Chomsky 1971: Child Behavior

(Crain & Nakayama (1987))

The girl is easily fooled.
Is the girl easily fooled?

The girl who can solve the labyrinth is easily fooled.
Is the girl who can solve the labyrinth easily fooled?

* Can the girl who solve the labyrinth is easily fooled?
Rule: Move main-clause auxiliary

Chomsky 1971: Child Data

The girl is easily fooled. Very infrequent
Is the girl easily fooled?

The girl who can solve the labyrinth is easily fooled.
Is the girl who can solve the labyrinth easily fooled?

* Can the girl who solve the labyrinth is easily fooled?
Rule: Move main-clause auxiliary
Pullum & Scholz (2002): Case Studies

Case 4: Auxiliary Fronting

P&S rebuttal:

The girl is easily fooled. Is the girl easily fooled? **Very frequent**

Rules out “front first aux” hypothesis: Should be very frequent

I could borrow your pencil when you’re done. When you’re done, could I borrow your pencil? **Very infrequent**

The girl who can solve the labyrinth is easily fooled. Is the girl who can solve the labyrinth easily fooled? *Can the girl who solve the labyrinth is easily fooled? Very infrequent

The changes these events portend are how fundamental. How fundamental are the changes these events ported? **Very infrequent**

Rules out “front first aux” hypothesis, though not a yes/no question: 15th sentence in WSJ corpus

The other dolly that was in here is where. Where’s the other dolly that was in here? **Very infrequent**

Rules out “front first aux” hypothesis, though not yes/no question: Child-directed speech

* Can the girl who solve the labyrinth is easily fooled? Very infrequent
Pullum & Scholz (2002): Case Studies

Case 4: Auxiliary Fronting

P&S rebuttal:

Main point (similar to previous ones): How much data is enough?

And is it really the case that children are only trying to rule out one other hypothesis (“front the first auxiliary”)? If so, why would they only be considering that one as viable, and not considering others as well?

* The girl who can solve the labyrinth is easily fooled.
Is the girl who can solve the labyrinth easily fooled? Very infrequent

* Can the girl who solve the labyrinth is easily fooled?

Pullum & Scholz (2002): Summary

Linguists should be careful about what knowledge they think children are acquiring.

It’s not that there is no evidence for the child to learn from. It’s just that it’s rare. But since linguists haven’t determined how much is enough, then even rare data should be taken seriously. Stay tuned for next class’s discussion for linguists’ take on a quantitative boundary...

Additional Note: Larger Point about PoS (from Crain & Pietroski (2002))

“...it’s not enough to mention ways in which children could learn some things without Universal Grammar. To rebut poverty-of-the-stimulus arguments, one has to show how children could learn [everything] adults actually know, and as close investigation reveals, adults know a lot more than casual inspection suggests. That is the nativist’s main point.”

Example of linguistic knowledge that’s hard: restrictions on meaning (see Crain & Pietroski (2002) for details)