Reminder: 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.

Legate & Yang (2002):
Poverty of the Stimulus Lives

Child Input

Very frequent
Is Hoggle talking to Jareth?

Very infrequent, if ever
Can someone who can solve the Labyrinth show someone who can't how?

Legate & Yang (2002):
Poverty of the Stimulus Lives

Hypotheses for frequent data type

**Structure-independent (linear)**
Front first auxiliary, Front last auxiliary, ...

**Structure-independent (hierarchical)**
Front the first auxiliary following the first noun phrase, Front the first auxiliary preceding a verb phrase, ...

**Structure-independent (creative)**
Front the auxiliary closest to a noun, Front the auxiliary that is an odd-numbered word position, ...

The Real Rule
Front the auxiliary following the subject noun phrase in the main clause.

But the unbiased child has to rule out all the other options, even ones that are simpler to compute. (For instance: front first auxiliary is much easier to compute.) We would expect to see errors of this type:

Is the dwarf who is talking to Jareth is going to give Sarah the peach?

Real Children
But kids don’t seem to make this error (Crain & Nakayama, 1987).

(Nativist) Implication: They’ve already ruled out that hypothesis, even though they’ve likely not seen much data (if any at all) incompatible with it. This is due to an innate bias to look for structure-dependent rules.

(Though see Perfors, Tenenbaum, & Regier 2006 for an ideal learner approach that shows structure-dependency can be learned from child-directed speech. See also Reali & Christiansen 2005, who show a simple statistical model can learn this particular sentence type from child-directed speech, and Kam et al. 2008 who show that that statistical model fails once it gets more varied child-directed speech and when it’s tested on different languages.)


Pullum & Scholz 2002 (P&S)
Claim: But there is enough disconfirming data available to children. So this situation is not true - poverty of the stimulus does not hold here.

Assumption: Only trying to rule out the front first auxiliary hypothesis, not all the other ones, too. (This isn’t necessarily true, and the PoS argument is based on the idea that the hypothesis space contains many more potential hypotheses.)


What kind of data?
One kind of disconfirming data: yes/no questions with two auxiliaries, where first auxiliary is not fronted
"Is the dwarf who is talking to Jareth is going to give Sarah the peach?"
(rare)

Another kind: wh-questions with complex subject, where first auxiliary is not fronted
"How could anyone who has watched Labyrinth before couldn’t not wince at this part?"
(how frequent?)


Pullum & Scholz 2002 (P&S): Corpus Hunt
Data set = 500 sentences of the Wall Street Journal
"How fundamental are the changes these events portend?"
"Is what I’m doing in the shareholders’ best interest?"
Not really a good sample of child-directed speech

Found that 1% are of this data type (5)


Child-directed speech (samples from Nina corpus of CHILDES)
"Where’s the little blue crib that was in the house before?"
"Where’s the other doll that was in here?"
"Where’s the other doll that goes in there?"

So data likely exists...
Estimate: 0.1%-1% of data are of this type


But Existence of Data ≠ Sufficiency of Data
We need to know if the amount of disconfirming (unambiguous data) is sufficient to learn the correct hypothesis by the time children seem to know it.

How much data is enough?

Gauging a threshold

Suppose we have two learning problems, Problem 1 and Problem 2.
Suppose both have only two hypotheses to choose from.
Suppose the frequency of unambiguous data for Problem 1 is Frequency 1 and the frequency of unambiguous data for Problem 2 is Frequency 2.

Idea: If children figure out Problem 1 and Problem 2 at the same time, and they're learning from the data alone, we would predict that Frequency 1 and Frequency 2 should be about equal.


Auxiliary-Fronting Threshold

Auxiliary-fronting is acquired by 3 years, 2 months (Crain & Nakayama 1987)

Is the girl who can solve the Labyrinth is going to save her brother?

* Can the girl who can solve the Labyrinth is going to save her brother?


Something else learned by about 3 years: Subject-drop (Valian 1991).
Except in special contexts, English speakers do not drop the subject.

She is going to eat the peach.
* Is going to eat the peach.

This is in contrast to languages like Spanish, which can optionally drop the subject.

Ella va a comer el melocotón.
she goes-3rd-sg to eat the peach

Va a comer el melocotón.
go-3rd-sg to eat the peach


Auxiliary-Fronting Threshold: Comparative

Auxiliary-fronting: acquired by 3 years, 2 months (Crain & Nakayama 1987)
Subject-drop: acquired by about 3 years (Valian 1991).

Unambiguous data for subject-drop: 1.2% of the data


Another bit of knowledge learned by about 3 years: Verb-Second movement in German and Dutch (German: Clahsen 1986, Yang 2000; Dutch: Lightfoot 1997, Yang 2000)

Sarah must solve the labyrinth.

German/Dutch:

Sarah must the labyrinth solve.
The labyrinth must Sarah solve.

Unambiguous evidence for Verb-Second movement: 1.2% of the data

Expectation: Auxiliary-fronting also needs 1.2% of the data to be unambiguous, in order for it to be learned by this age.


So how much data is there really?

Looking at the Nina corpus:
46,499 sentences
20,651 questions
14 unambiguous data examples (all of wh-question type)

Frequency of unambiguous data: 0.088% (much less than 1.2%)

So how much data is there really?

Looking at the Adam corpus:
- 20,372 sentences
- 8,889 questions
- 4 unambiguous data examples (all of wh-question type)

Frequency of unambiguous data: 0.045% (much less than 1.2%)

Data is not frequent enough for children to learn by the time they do.


A larger point about data-driven learning

Problem: “…wild statistical disparities between what is presented to children and how children actually learn”

Example: Subject-drop (lots of “data”, late generalization)

Almost all English sentences contain a subject, but children don’t get it till 3.


A larger point about data-driven learning

Problem: “…wild statistical disparities between what is presented to children and how children actually learn”

Example: Verb-Raising in French (little “data”, early generalization)

“She eats not the peach”

Only 7% of French sentences show this, but children acquire it by 1.5 years.

The point: Children come with innate biases that allow them to use data in specific ways to update their hypotheses.

Innate Bias = Domain-Specific?

Poverty of the Stimulus (the existence of an induction problem) is usually used as the motivation for Universal Grammar. But just because an induction problem exists doesn’t mean innate domain-specific knowledge like UG is required to solve it. The knowledge required could be derived from prior knowledge (domain-specific or domain-general) or simply be domain-general to begin with.

Exploring the Nature of the Necessary Bias(es): Computational Modeling Work

Domain-general biases explored:
- prefer subset hypothesis: Regier & Gahl 2004
- prefer simplicity: Perfors, Tenenbaum, & Regier 2006, submitted
- use only maximally informative data: Pearl & Weinberg 2007, Pearl 2008, Pearl & Lidz 2009

Domain-specific specific biases explored:
- ignore certain kinds of ambiguous data that are identified using domain-specific (linguistic) knowledge: Regier & Gahl 2004, Pearl & Lidz 2009
- ignore embedded clause data: Pearl & Weinberg 2007
- prefer syntactic information over semantic information: Foraker et al. 2007, forthcoming 2009