Language acquisition and Kolmogorov complexity: Why is language acquisition possible?

Nick Chater
OVERVIEW

• A brain adapted for language…

• …or language shaped by the brain?

• What can be learned from positive data I
  – asymptotic results

• What can be learned from positive data II:
  – a recipe
1. A brain adapted for language?
COULD BRAINS BE ADAPTED FOR LANGUAGE?

• Language seems extremely complex

• And to have many highly specific and incredibly subtle properties

• How can children figure it out, while linguists can’t?

• That is, how is language acquisition possible?

• Perhaps the triggering of a genetically coded language-specific faculty?
  – language instinct
  – language organ
  – language acquisition device
  – language module
A LANGUAGE-SPECIFIC FACULTY IMPLIES THE BRAIN IS ADAPTED FOR LANGUAGE, JUST AS FOR VISION

• The visual environment today

• Visual environment of evolutionary adaptation

Strangely similar…
• The linguistic environment today

“…the cat sat on the mat…”

• Linguistic environment of evolutionary adaptation

Roar!

Strangely dissimilar…
BUT PERHAPS LANGUAGE AND THE LANGUAGE FACULTY CO-EVOLVED VIA THE BALDWIN EFFECT (PINKER & BLOOM, 1990)

• Driving acquired traits into the genes--
  – It may work for ostrich calluses
  – perhaps it works for language
THE BALDWIN EFFECT: A VERY SIMPLE SIMULATION

- “Language” is a string of features

- Genes can express bias or neutrality on each feature:
  - “fixed” .95 bias to red:  
  - “fixed” .95 bias to blue:  
  - “learnable”: unbiased

- “Genome”:

- Trial and error learning

- Only the fastest learners “reproduce”

- And create the next generation by sexual recombination and mutation

- Do the genes begin to adapt to the language???
THE BALDWIN EFFECT IN ACTION
BUT CO-EVOLUTION REQUIRES GENETIC ADAPTATION TO A VARYING LANGUAGE

• Can language change _lead_ language genes?

• Potential problem:
  – Language changes very fast, in relation to genetic change

• So what happens when language and genes can _both_ change?
SAME SPEED FOR LANGUAGE AND GENETIC MUTATION RATE
LANGUAGES CHANGES TWICE AS FAST

![Graph showing the number of learnable and initially correct bits over generations.](image)
LANGUAGE CHANGES 10 TIMES AS FAST

No Baldwin effect

No coevolution

“Learnable” genes win out
GENES CANNOT CATCH A LINGUISTIC “MOVING TARGET”

DIVERGING HUMAN POPULATIONS

Tracing Human History Through Genetic Mutations

By examining DNA patterns that are inherited maternally or paternally, scientists can trace human lineages back to the original branches, or sons and daughters, of a genetic Adam and an Eve.

Europe:
- **EVE** (Y-chrom)
  - The nine European lineages are lettered A through I and K. One of the lineages, K, diverges to America, but its route is not known.

America:
- **EVE** (Y-chrom)
  - All American lineages are lettered variants of African and Asian branches.

Africa:
- **EVE** (mitoch)
  - The three African branches are named L1 through L3, and L3 separates into all the other branches.

Asia:
- **EVE** (mitoch)
  - The five Asian branches are named A through D and F and G.

The seven Asian branches are 4 through 10, and these groups branch off into Oceania, Europe, and America.

Sources:
- Dr. Douglas C. Northrup, Marie H. Loo, Emory University
- Dr. Peter A. Underhill, Stanford University
- "Genes, Peoples, and Languages," by Dr. Luca Cavalli-Sforza

Joint work with Andrea Baronchelli, Romualdo Pastor-Satorras, Morten Christiansen, in preparation
Once populations are split, co-evolution will be specific to the local linguistic environment

If language-gene coevolution occurred, it had better stop, once populations diverge

But wide geographical separation occurred early, w.r.t., to presumed time-scale for language

(And even geographically nearby groups show very fast linguistic change)

Test with population splitting simulations...

No co-evolution beyond this point!

Joint work with Andrea Baronchelli, Romualdo Pastor-Satorras, Morten Christiansen (in prep)
Case 1: Language change is fast

- No coevolution
- Neutral “genes” dominate
- No UG
Case 2: Language change is slow

- Lots of local coevolution; few neutral genes
- Genetic divergence precisely mirrors linguistic divergence;
- No UG
1. A brain adapted for language?
2. ...or is language shaped by the brain?

Morten Christiansen
TWO PROBLEMS OF INDUCTION

N-induction
1, 2, 4…

Hard: We may have the wrong biases

C-induction
“1, 2, 4…”

Easy: we definitely have the right biases
SO LANGUAGE EMERGES FROM INTERACTING CONSTRAINTS...

- Perceptual-motor
  - (speech/auditory apparatus)
- General learning mechanisms
- Cognitive/processing constraints
  (e.g., heuristics and “good enough” parsing, Ferreira)
- Semantics
  (perhaps including embodiment, e.g., Casasanto & Lozano)
- Pragmatics (e.g., Dowty, 1980; Levinson, 2000; Reinhart, 1983)
  - All of which then can become “fossilized” in the language
  - (cf grammaticalization, Hopper, Bybee, etc.)
LANGUAGE IS OPTIMISED TO THESE FACTORS OVER GENERATIONS

Hence language will end up in local minimum in a “fitness landscape”
LEARNING LANGUAGE IS SEARCHING FOR LOST KEYS IN A VAST CITY, ON A DARK NIGHT...

...at least the keys are always right under the lamp-posts
3. What can be learned from positive data: I. Asymptotic results

Paul Vitányi
THE LOGICAL PROBLEM OF LANGUAGE ACQUISITION
(e.g., Hornstein & Lightfoot, 1981; Pinker, 1979)

• Children appear able to learn from *positive* evidence alone
• *Over*general grammars predict that bad sentences are actually ok
• “Mere” non-occurrence of sentences is not enough
• Backed-up by formal results
  – Gold, 1967
  – though Feldman, Horning et al

cf. Alex Clark – learning complex
grammars and time-complexity
Perfors, Tenenbaum & Regier on learning
that language is based on phrase structure not strings
John Goldsmith, learning phonological structure
SPECIFYING AN “IDEAL” LEARNING SET-UP

- Linguistic environment
- Measures of learning performance
- Learning method

- Positive evidence only; language generated by a computable-random process, $\mu$, with shortest code length $K(\mu)$
- Statistical, not exact (PAC-style)
- Simplicity
PREDICTION BY SIMPLICITY

- Find shortest ‘program/explanation’ for current ‘corpus’

- Predict using that program
  - Strictly, use ‘weighted sum’ of explanations, weighted by brevity
PREDICTION IS POSSIBLE! (SOLOMONOFF, 1978) 
SUMMED ERROR HAS FINITE BOUND

\[ \sum_{j=1}^{\infty} s_j^2 \leq \frac{\log_e 2}{2} K(\mu) \]

So prediction converges

[faster than \(1/n\log(n)\)], for corpus size \(n\)

An amazing, and fundamental, result; we assume only computability of the data

Admittedly, the method is uncomputable :-(
LOGICAL PROBLEM OF OVERGENERALIZATION IS SOLVABLE

• Suppose learner has probability $\Delta_j$ of erroneously guessing an ungrammatical $j$th word

$$\sum_{j=1}^{\infty} \langle \Delta_j \rangle \leq \frac{K(\mu)}{\log_e 2}$$

• Intuitive explanation:
  • overgeneralization underloads probabilities of grammatical sentences;
  • Small probabilities implies longer code lengths
ABSENCE AS IMPLICIT NEGATIVE EVIDENCE

- Overgeneral grammars predict missing sentences
- And their absence is a clue that the grammar is wrong

This overgeneralization theorem makes this intuition rigorous
EXTENSIONS

• Convergence to learning to *generate* language

• Convergence to relate form and meaning
  – both to a high level of accuracy

• If we assume i.i.d. (or stationary distribution?)
• Can we identify in the limit the true generative model, with overwhelming probability
• i.e., given a stream of sentences,
• Learner generates successive guesses, until, ultimately
  – The learner never changes its mind
  – And it has precisely identified the generative model (extensionally)
• The learner’s strategy is computable

Hsu, Chater & Vitanyi, under review
BUT ABSTRACT RESULTS ARE ONLY A START...

• So far, our results show what an ‘ideal’ learner could do;
  – Uncomputable or intractable
  – Asymptotic
  – Not related to specific cognitive/linguistic phenomena

• How far can they be ‘scaled-down’ to deal with real language acquisition phenomena…?
4. What can be learned from positive data: II. A recipe

Anne Hsu
POVERTY OF THE STIMULUS: THE RECIPE

• Goal of the learner: find the shortest code for the linguistic data:
  – A “dual” of Bayesian inference

• Structure needs to pay its way

• Constraint $C$ is learnable if code which
  – 1. “invests” $I_c$ bits to encode $C$ can…
  – 2. …recoup its investment by saving more than $I_c$ bits in encoding the linguistic data
NATIVISM VS. EMPIRICISM, THE DEBATE

• Nativism
  – c is acquired
  – But key source of data X is not sufficient to recoup the cost of coding C
    » Empiricist: But there may be hidden sources of data

• Empiricism?
  – Ample supply of data to recoup the cost
    » Nativist: But the data may not be required or even registered

The recipe can be used for or against either side
EASY EXAMPLE: LEARNING SINGULAR-PLURAL

John loves _ tennis  x bits
They love _ tennis  y bits

John loves _ tennis
*John love _ tennis  x+1 bits
They love _ tennis
*They loves tennis  y+1 bits

If constraint applies to proportion \( p \) of \( n \) sentences,
constraint saves \( n \log_2 \frac{1}{p} \) bits
LEXICAL ALTERNATIONS

I snapped the pencil  I gave a book to the library
I made the pencil snap  I gave the library a book
I made the pencil disappear  I donated a book to the library
*I disappeared the pencil  *I donated the library a book

These *must* be learned, on any theory, of course…
CONTRACTIONS

• Gonna contraction
  I’m going to help her
  I’m gonna help her
  I’m going to the store
  *I’m gonna the store

• Wanna contraction
  Which team do you want to beat?
  Which team do you wanna beat?
  Which team do you want to win?
  *Which team do you wanna win?

These might arise from ‘deep’ syntactic principles, which might, or might not, be learned…
### PRELIMINARY RESULTS

<table>
<thead>
<tr>
<th></th>
<th>Investment (bits)</th>
<th>Encoding savings per occurrence (bits)</th>
<th>Occurrences in 1 year of child experience</th>
<th>Years to learn</th>
</tr>
</thead>
<tbody>
<tr>
<td>wanna</td>
<td>158.2</td>
<td>0.4</td>
<td>1.2</td>
<td>330</td>
</tr>
<tr>
<td>gonna</td>
<td>112.3</td>
<td>1.0</td>
<td>108.6</td>
<td>1</td>
</tr>
<tr>
<td>donate (the library)</td>
<td>44.9</td>
<td>4.9</td>
<td>7.2</td>
<td>1.4</td>
</tr>
<tr>
<td>disappear (a rabbit)</td>
<td>44.9</td>
<td>.2</td>
<td>61.8</td>
<td>3</td>
</tr>
</tbody>
</table>

∴ wanna contraction is not learnable in isolation, from distribution evidence alone

SUMMARY

• Why are learners and language aligned?
  – 1. Innate UG?; but UG is evolutionarily implausible?
  – 2. Language is shaped by the brain
    • And culture more broadly, including categories, rituals, social conventions, skills
    • Crucial to understand the inductive biases that have created these structures, in order to learn them (cf Tom Griffiths, on uncovering inductive biases experimentally)

• Is learning possible from positive evidence?
  – 3. Hopeful, but idealized, learnability results
  – 4. And a recipe for studying learnability of specific constraints