Modeling Syntactic Acquisition

Cognitive Plausibility, Explanatory Levels, and Inference

By Austin Soe

Source: Pearl (2021)

Table of *contents*

Source: Pearl (2021)

01

Cognitive Plausibility

Section 2.1

Source: Pearl (2021)

Cognitive Plausibility (Section 2.1)

- Models must realistically reflect child language acquisition
- Use empirical data to inform assumptions:
  - **Theoretical work:** guides initial grammar and inference
  - **Corpus analysis:** shows real language input/output
  - **Experimental studies:** inform perceptual/processing abilities
- Modeling challenges:
  - Estimating realistic input (beyond just speech)
  - Choosing learning timelines and target states when data is limited

Source: Pearl (2021)

02

Levels of Explanation

Section 2.2

Source: Pearl (2021)

Levels of Explanation (Section 2.2)

From Marr (1982): Three levels for modeling cognition

- **Computational Level**
  - Defines the **goal** of acquisition (e.g., learn grammar from input)
  - Answers: Can this task be solved in principle?
- **Algorithmic Level**
  - Describes how children solve the task
  - Incorporates **limited memory, incremental learning,** and **realistic steps**
- **Implementational Level**
  - Focuses on **how learning occurs in the brain**
  - Explores **neural architecture and representations**

Source: Pearl (2021)

03

Inference

Section 2.3

Source: Pearl (2021)

Inference Overview (Section 2.3)

Inference = How the model updates its grammar hypotheses

- Types of inference mechanisms:
    - Counting
    - Reinforcement learning
    - Tolerance Principle
    - Bayesian updating

Source: Pearl (2021)

Inference Overview: *Counting as Inference*

- Counts are converted into probabilities
- Example: +wh-movement vs. -wh-movement
- Introduce smoothing (assigns small probability to unseen events)
- Simple, but powerful for measuring likelihood of grammatical rules

Source: Pearl (2021)

Inference Overview: *Reinforcement Learning*

- Inspired by Bush & Mosteller's linear reward-penalty model
- Data either rewards or penalizes a hypothesis
- Uses learning rate (γ) to adjust probabilities
- Works best with unambiguous data

Source: Pearl (2021)

Inference Overview: *The Tolerance Principle*

- Created by Yang (2005, 2016)
- Helps decide when to generalize
- Formula: A rule is adopted if exceptions $\leq N / \ln(N)$
- Focuses on efficiency in retrieval of knowledge

Source: Pearl (2021)

Inference Overview: *Bayesian Updating*

- Learner evaluates hypotheses based on prior knowledge + data
- Formula: Posterior ∝ Likelihood × Prior
- Can be done:
  - All at once (computational-level)
  - Incrementally (algorithmic-level)
- Hierarchical Bayesian Models introduce over hypotheses:
  - Abstract patterns that generalize beyond specific data

Source: Pearl (2021)

04

Key Takeaways

Source: Pearl (2021)

Key Takeaways

- Models must be cognitively grounded
- Important to distinguish between goal (computational) and process (algorithmic)
- Inference mechanisms allow models to simulate child learning
- These frameworks help us test linguistic theories using real-world data

Source: Pearl (2021)

Terminology

Cognitive Plausibility

How realistic a model is in approximating how children actually acquire language, based on theory, corpus data, and experiments.

Inference Process

The mechanism by which a child updates their grammar from data (e.g., counting, Bayesian reasoning).

Initial State

The learner’s starting knowledge (e.g., Universal Grammar, processing abilities).

Learning Period

The timeframe over which language learning happens in the model (e.g., infancy to early childhood).

Data Intake

The linguistic input that the modeled learner uses, derived from real-world child-directed speech.

Target State

The final grammar or knowledge state the child is expected to reach (adult-like or age-appropriate).

Source: Pearl (2021)

What is “Smoothing”

Smoothing is a statistical technique used in language acquisition models to handle the problem of unseen data. It prevents the model from assigning a probability of zero to any option just because it hasn’t been observed yet.

In the context of the modeled child:

- The child acts as if they’ve already seen each option a tiny bit — even before getting any actual data.
- This is done using pseudocounts (e.g., assigning 0.5 to each hypothesis).
- This way, no option starts with 0 probability — they all start with something like 0.5 out of 1.

Why it's useful:

- If an element had zero probability, the model would never be able to change its belief about it — no matter what future data came in.

  Smoothing gives the model “flexibility” to learn from future evidence, even about things it hasn’t seen yet.

Source: Pearl (2021)

References

- Pearl, L. S. (2021). Modeling syntactic acquisition. In J. Sprouse (Ed.), The Oxford Handbook of Experimental Syntax. Oxford University Press.

Thanks!