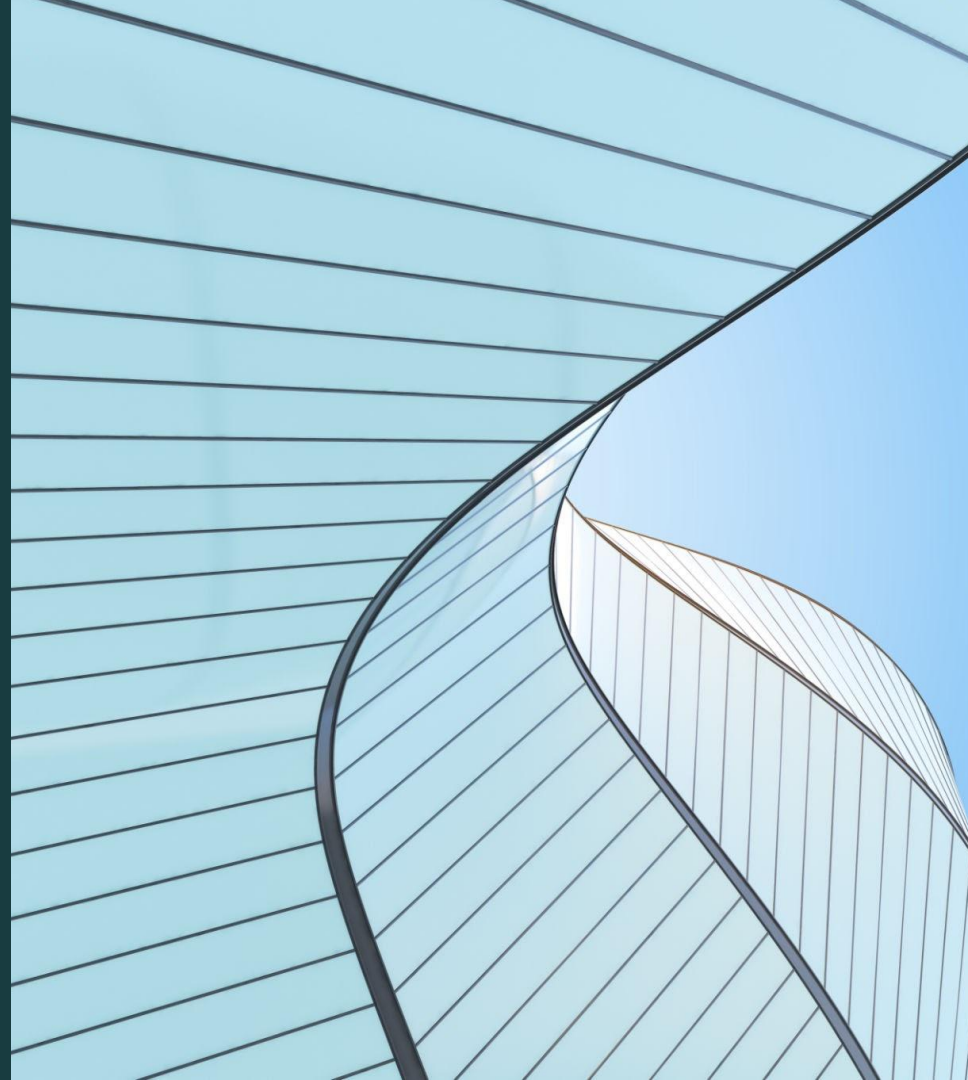


Maddux Lorge
04/17/2026

Pearl 2021 Chomsky handbook 1.2.1 + Pearl 2021 JoCL UG+Stats 2.2

Bridging the Gap: Statistical
Learning Meets Universal Grammar



For my discussion lead I am planning on discussing the overall information and details about the article. Then at the end I have some discussion questions to help us further elaborate about these topics.

Core Question: Are children "counting" their way to fluent language, or are they born with a blueprint?

What does this mean:

- Language is too complex to learn from scratch so babies must have a biological "cheat sheet" that tells them how sentences work (belief)
 - Kinda like a GPS
- Counting Argument (statistical learning): by counting how often certain words or sounds follow each other, babies can calculate the rules of grammar through probability

The Middle Ground

Hybrid:

→ Babies use both!!

→ **Why this matters:**

→ If it's all counting, then anyone (or any animal) with a fast enough brain could learn language. If it's all blueprint, then language is a unique "human-only" biological superpower.



Statistical Learning Basics

Definition: statistical learning is counting things and updating hypotheses based on those counts

Domain General: it's not just for language as some evidence shows infants use this for both linguistic and non-linguistic patterns

The Mechanisms: observations become counts

- Counts become probabilities (ex// $3/10 \rightarrow 0.3$ probabilities)
- Higher probability = higher likelihood of that rule being correct

3 Key Learning Mechanisms

1) Reinforcement Learning :

- a) “reward/penalty system”
- b) if a linguistic option matches the input (e.g., wh-movement in English), its probability is "rewarded" and increases. If it doesn't match, it is penalized and decreased.

2) The Tolerance and Sufficiency Principles

- a) How many exceptions can a rule have before it's “broken”
- b) Explains how kids decide when a rule (like adding -ed for past tense) is productive vs when they just have to memorize exceptions

3) Bayesian Inference (most math heavy part)

- a) It uses Priors (how plausible a hypothesis is regardless of data) and Likelihoods (how well the data fits the hypothesis) to find the most probable grammar

The Synergy - UG + Statistics

- **The Integration** : UG defines the “menu” of possible languages (the hypothesis space), and Statistical Learning allows the child to “order” the correct one efficiently
- **Why it matters**: Without UG the search space is too big. Without Statistics, the child couldn't navigate the space effectively when faced with noisy or ambiguous data.
- **Observation From Reading**: "Statistical learning is often used to navigate through a hypothesis space to identify the correct hypothesis for the language."

The Bayesian Brain

- **The Goal:** To determine the *posterior probability* so basically, "How likely is this grammar rule given the talking I've heard?"
- **The Components:**
 - **Prior ($P(h)$):** How simple or "economical" is the rule? (UG might favor simpler rules).
 - **Likelihood ($P(D|h)$):** How well does the rule actually explain the sentences the child hears?
- **Key Insight:** Children aren't just robots; they weigh what they "know" (Priors) against what they "hear" (Likelihood).

The "Cut-off" – Tolerance & Sufficiency

- **The Problem:** Language is messy (e.g., "go" becomes "went," not "goed").

When does a child decide a rule is worth keeping?

- **The Efficiency Logic:** A rule is only useful if it helps you process language *faster* than just memorizing every word.

- **The Math of "Tolerance":** There is a specific threshold (a "tipping point") where the number of exceptions makes a rule too "expensive" for the brain to maintain

Overall

- **Domain Generality:** These mechanisms (Bayesian, Reinforcement) aren't "just for language", they are general tools the human brain uses for everything from walking to social cues.
- **The UG Synergy:** Statistical learning is the engine, but Universal Grammar provides the tracks. Without the tracks, the engine just spins its wheels in the mud.
- **Developmental Evidence:** The text cites evidence that infants as young as **6 months** are already capable of Bayesian inference.

Some Discussion Questions:

Reinforcement learning is "reward vs. penalty," while Bayesian is "probability of fit." Which one feels more like "real" human learning? Is language a series of rewards or a calculation of odds?

Some Discussion Questions:

The Tolerance Principle: How many exceptions can a rule have before it's "broken". This explains how kids decide when a rule (like adding *-ed* for past tense) is productive versus when they just have to memorize exceptions.

Why do you think children are so much better at letting go of "bad rules" than adults learning a second language?

Some Discussion Questions:

If kids are better at this at 6 months than adults are at 30 years old, what does that say about the "Critical Period" for language? Does our statistical "engine" get rustier as we get older?

Some Discussion Questions:

If a child is raised in a home with "messy" grammar or two languages, how would **Bayesian Inference** help them stay on track?

Some Discussion Questions:

If a child is exposed to a language with *too many* exceptions, would they eventually stop trying to find rules altogether? (Think about "creolization" where kids create new rules out of messy input).

Major Observation: Statistical learning is **domain-general**, but its *application* in language is constrained by **domain-specific** (linguistic) knowledge.

Final Takeaway: Language acquisition isn't just about being "born with it" or "learning it all from scratch." It is a sophisticated interaction where innate constraints make statistical learning possible in the first place.