How Statistical Learning Complements UG

Pearl 2021, Chomsky Handbook, 1.3
Pearl 2021, JoCL UG+Stats, 3.5

Linking theories - what & why

- Linking theories explain how we figure out who does what to whom in a sentence using syntax.
  - Example: The little girl **blicked** the kitten on the stairs
    - even with a made-up verb, we know who acted
- Children use syntactic cues to infer thematic roles like AGENT and PATIENT, based on verb meaning and structure.
- Linking theories are crucial for learning how to generalize to new verbs, especially when the input is noisy or sparse.

UTAH vs. rUTAH

- Uniformity of Theta Assignment Hypothesis
- UTAH operates by grouping thematic roles into larger categories and then mapping these larger categories consistently to specific syntactic positions
  - AGENT = subject
- rUTAH assumes a relative ordering of thematic roles and maps the highest role available to the highest syntactic position
  - rUTAH is better at handling cases where not all roles are present, like *The ice broke.*
- Pearl & Sprouse's (2019) model favors rUTAH, especially when generalizing across diverse sentence structures and input variability

Learning linking theories - not innate…?

- Traditional view: children have innate linking knowledge
  - UG ‘tells’ them the rules
- BUT Pearl & Sprouse (2019) argue that kids learn linking theories using Bayesian inference over input.
- Children infer the most probable linking system from the data they receive, guided by UG constraints.
- This supports a gradual, input-driven model where statistical patterns lead to rule formation.

What is *movement*?

- Movement = when a word appears in one place in a sentence but is interpreted elsewhere.
- Common in raising, control, passives, and wh-questions.
- Example: Jack seemed to kiss Lily — Jack is the subject, but he's not the "seemer," he's the kisser.
- This involves movement from the embedded clause to the subject position in the main clause.

Raising vs. Control structures

- Raising: Subject or object has no role in the main verb but does in the embedded verb (seem, want).
- Control: The subject or object gets a role from both verbs, linked by a silent PRO.
  - Jack wanted to kiss Lily = control
  - Jack **seemed** to **kiss** Lily = raising → one person doing both actions
- Also applies to object-control vs. object-raising
  - Jack wanted Lily to laugh vs. Jack asked Lily to laugh

UG + Stats in verb learning

- Children need to learn which verbs allow which structures and how to interpret them.
- UG might guide attention to specific features like animacy or argument type.
- Statistical learning groups verbs based on shared properties using Bayesian clustering.
- Pearl & Sprouse tested whether kids treat verbs like want, like, and need as a class — some models said yes, others no.

Some key takeaways~

- UG helps children focus on meaningful features; statistical learning lets them generalize from data.
- This hybrid model explains how kids form linking theories and understand syntactic movement.
- Models can be tested with behavioral data
  - like whether 5-year-olds interpret verbs as predicted.
- Shows how structure-based knowledge can emerge from input without assuming it’s fully innate.