How Far Can Indirect Evidence Take Us? Anaphoric One Revisited
Lisa Pearl & Benjamin Mis, University of California, Irvine

Overview

Implication: Children have helpful learning biases.
Open question: The nature of those biases
Some dimensions of variation:
- Innate vs. Derived
- Domain-specific vs. Domain-general
- What to learn vs. How to learn

Universal Grammar (Chomsky 1965) = learning biases that are innate and domain-specific.

Main question
When induction problems exist, what does it take to solve them?
- What indirect evidence is available?
- Are the necessary biases innate and domain-specific?

Anaphoric one
One linguistic phenomenon argued to present an induction problem.
"Look - a red bottle!"

Adult representation:
Syntactic antecedent = "red bottle" (one = N).
Semantic referent = RED BOTTLE (modifier property is important).

Child behavior:
Same at 18 months as adults (Lude, Wexman, & Friedman 2003).

Assumption: 18-month-old representation is same as adult representation.

A potential source of indirect evidence
Pearl & Mis (P&M) Observation: Other pronouns can also be used anaphorically.
"Look: A cute penguin. I want to hug it!"

NP = [a [k cute [k [a penguin]]]] → [a NP]

They are always category NP (UNAM NP), as evidenced by their syntactic environment and antecedent. The referent is unambiguous w.r.t. to the mentioned property - i.e., the referent must have the property mentioned (e.g., cute). Note that one can also be category NP sometimes.

"Look: A cute penguin. I want one."
= "Look: A cute penguin. I want a cute penguin."

Why are these data useful?
They can help a child decide in general if the referent of an anaphoric pronoun should have a property mentioned in the potential syntactic antecedent.

P&M proposal:
Use UNAM NP data in addition to all the other data. To do this: The child must recognize one is similar to other anaphoric pronouns (look at syntactic and semantic distribution of one and other pronouns, using innate domain-general statistical learning abilities).

Modeling information in the data
Previous proposals for using the direct evidence: Input restrictions
Baker 1978 (Baker): Only UNAM data are informative. Children must have innate, domain-specific knowledge that one cannot be category NP because UNAM data rarely occur.

Regier & Gahl 2004 (R&G): Leverage SEM-SYN ambiguous data in addition to UNAM data. Children use innate domain-general statistical learning abilities to track suspicious coincidences in the properties that one's referents have.

Pearl & Lidz 2009 (P&L): Filter out SYN ambiguous data, even if using SEM-SYN ambiguous data - otherwise children will learn one is category NP. Children employ a domain-specific bias to ignore these data, which can be derived from an innate domain-general preference for learning local cases of uncertainty.

Online probabilistic framework
General form of update equations (Chow 1971)
\[ p_i = \alpha_i \times \text{data} \]
\[ \alpha_i = \frac{\beta_i}{\beta_i + \text{total data}}. \]

After every information data point encountered:
\[ \text{data} = \text{data} + s_i \] (incremented by probability that data point suggests i is true)
\[ \text{total data} = \text{total data} + s_i \]
One informative data point seen
\[ s_i = p_i \times \text{data} \times (1 - \text{data}) \]
\[ \alpha_i = \frac{p_i \times \text{data} \times (1 - \text{data})}{p_i \times \text{data} \times (1 - \text{data}) + (1 - p_i) \times (1 - \text{data})} \]

If property was mentioned: \( i \); in what the syntactic environment ic; if \( p_i \) the previous past mentioned a property ic; whether the object has the mentioned property.

Successful metrics
Want \( p_i \) near 1, \( p_{\text{not}} \) near 0; near one, and reproducing infant learner behavior (\( p_{\text{not}} \)) to be near 1. \( p_{\text{not}} = p_i - \text{uninformative} \times (1 - p_i) \times (1 - \alpha_i) \times (1 - \alpha_i) \times (1 - \text{data}) \times (1 - \text{data}) \)

Results & Implications
Averages over 1000 simulations, standard deviations in parentheses.

P&M learner: Correct behavior, even if the representation is incorrect (\( p_i \) is low). This is due to the additional indirect evidence data, since other learners produce the same qualitative results found previously.

Main Answers
- Child anaphoric one behavior can be reproduced without requiring innate domain-specific learning biases, provided the children learn from indirect evidence. No input filtering is required.
- However, this does not lead to the adult representation. That representation may be learned during a second stage of acquisition, and may require innate domain-specific learning biases to do so.