# 3.1 Syntactic: Structure Dependence

**Poverty of the Stimulus Without Tears**
**Pearl (2023)**

Frankie Boren

# Structure dependence

## the correct hypothesis and the hypothesis space

- The structure of language is generally agreed to be hierarchical rather than linear.
- Linguistic rules rely on hierarchical structure (they are *structure dependent*).
    - Example: yes/no question formation in English:

        a. [CP Can the penguin [CP who is on the iceberg] (*t)* can find a fish]?

        b. [CP The penguin [CP who is on the iceberg] can find a fish].

## the correct hypothesis and the hypothesis space

- In the example of English yes/no question formation, the correct hypothesis involves inversion (i.e., movement) while the hypothesis space consists of all possible rules for transforming (a) into (b).

    a. [CP Can the penguin [CP who is on the iceberg] (*t*) can find a fish]?

    b. [CP The penguin [CP who is on the iceberg] can find a fish]

- So, the hypothesis space includes the inversion rule, other structure-dependent rules, **and** structure-independent rules.

- But rather than being confined to specific examples, structure-dependence is meant to apply to the whole linguistic system.

# The available data

## (or so we thought)

- Initial investigations into children's knowledge of structure dependence focused on individual phenomena (e.g., yes/no questions) rather than the system as a whole.
- This lead to the data being considered in isolation.
- As a result, various analyses suggested that most of the input data were ambiguous between the correct structure-dependent rule and other competing rules, including structure-independent ones.

# How children leverage the available data

- When it came to structure-dependence, it was initially assumed that:
    1. children only learned from direct positive evidence (i.e., by observation)
    2. ambiguous data were not informative
- These assumptions made it seem like the lack of sufficient unambiguous data for structure-dependence in complex yes/no questions resulted in poverty of the stimulus.

# Age of acquisition for constrained generalizations

- English children as young as age three seem to know that rules controlling complex yes/no question formation in English ought to be structure-dependent (Crain & Nakayama, 1987), despite their lack of exposure to positive evidence.

- Therefore, English children's early development of this structure-dependent knowledge appeared to be evidence for poverty of the stimulus.

# The investigations

## Reali and Christiansen (2005)

- An early computational investigation by Reali and Christiansen (2005) rejected the assumption that children were trying to learn a particular structure-dependent rule, given a mixed hypothesis space.
- The modeled learner was tuned to children's observable behavior and learned to distinguish grammatical complex yes/no questions from ungrammatical ones.

# Reali and Christiansen (2005)

- This investigation also rejected the assumption that available data was restricted to yes/no questions.
- Instead, the model learned from all available utterances. It leveraged the relative frequencies of 2-word and 3-word sequences, which allowed it to successfully distinguish between grammatical and ungrammatical complex yes/no questions.
    - (e.g., 2-word: Is-the, the-boy, boy-who; 3-word: Is-the-boy, the-boy-who)
- The result suggested that poverty of the stimulus was not taking place.

## Kam, Stoyneshka, Tornyova, Fodor, and Sakas (2008)

- However, Kam et al. (2008) demonstrated the the particular corpus used as input and the particular test sentences learned in the model by Reali and Christiansen (2005) resulted in a “lucky fluke.”

- After learning from a wider range of complex yes/no questions, the model failed to generate predictions consistent with empirical data.

- Therefore, poverty of the stimulus seemed to exist after all, despite the model’s predefined hypothesis space.

# Perfors, Tenenbaum, and Regier (2011)

- In a computational investigation by Perfors, Tenenbaum, and Regier (2011), the modeled learner considered a hypothesis space which included both structure-dependent and structure-independent representation types, some of which were not tied to complex yes/no questions. This data contained unresolvable ambiguity.

- The available data was representative of what children actually encounter, and presented a need to identify the sequences of syntactic categories within each utterance.

## Perfors, Tenenbaum, and Regier (2011)

- Their modeled learner used Bayesian inference to identify which representation was able to balance the representation's complexity with its ability to encode the data.
- It successfully identified the correct structure-dependent representation from those available in the hypothesis space, on the basis of the data children encounter.
- In principle, if children possess this simplicity bias (which is domain-general), they can use it over their input to arrive at the correct hypothesis.
    - This would support a non-linguistic nativist perspective.

## Abend et al. (2017)

- An investigation by Abend et al. (2017) also approached knowledge of structure dependence as it relates to a larger linguistic system rather than isolated phenomena.

- But rather than explicitly defining the hypothesis space as an assortment of structure-dependent and structure-independent representations, it was implicitly defined as infinite (via pre-defined structure-dependent building blocks and constraints on the combinatorial possibilities of those building blocks).

- So, while their modeled learner already assumed structure-dependence, it didn’t know which of infinitely many constructible structure-dependent representations was the correct hypothesis.

## Abend et al. (2017)

- To learn, the model would generate explicit structure-dependent hypotheses the basis of the pre-defined building blocks and constraints and then evaluate those hypotheses on the basis of the available data.

- Because the modeled learner saw syntactic structure as part of a larger linguistic system, with syntax and semantics **connected**, it learned from all the available data - including syntactic, semantic, and non-linguistic information.

- Even still, poverty of the stimulus seems to persist, considering the prior bias to use Bayesian inference, as well as an additional bias to minimize cost by reusing building blocks.

- Their solution (in principle) to the poverty of the stimulus relies on linguistic **and** non-linguistic biases.

# Fitz and Chang (2017)

- A recent investigation by Fitz and Chang (2017) also capitalized on the connection between **form and meaning**, but it rejected the assumption that children were trying to learn a particular structure-dependent rule for complex yes/no questions.
- This modeled learner was tuned to observable behavior. It knew that meaning representations were hierarchical, and observed that utterances were generated from those underlying meaning representations.
- A neural network was used to navigate the space of possible ways to generate observable sequences from underlying meaning representations, given the available data
- After it was trained, the model's results matched the empirical data for complex yes/no questions.

# McCoy et al. (2018)

- McCoy et al. (2018) also used a neural network approach, but assumed that the learner is trying to identify a particular structure-dependent representation from the hypothesis space of possible representations.

- The modeled learner was given the declarative version of the complex yes/no question **instead of** the meaning representation. It also accessed declarative utterances similar to those children might hear.

- The model's output was comparable to empirical data despite unresolvable ambiguity in the input, resulting in poverty of the stimulus.

# Interpreting these investigations

## Useful biases for overcoming poverty of the stimulus:

1. Restricting the relevant input easily leads unresolvable ambiguity. Therefore, a bias to consider any individual linguistic knowledge piece as part of a larger linguistic system is useful.

2. Preferring “rational rules” that prioritize compact representations and representations that rely on reusable building blocks is a useful non-linguistic bias.

3. Linguistic biases pertaining to (i) preferred structural building blocks and (ii) sensitivity to form and meaning connection.

# Consensus

- The linguistic nativist viewpoint is supported by the innate biases in current solutions to this poverty of the stimulus (which seem to be linguistic).

  *(Unless, of course, the linguistic biases turn out to be derived from non-linguistic biases, which would support the non-linguistic nativist view.)*

thank you!