Semantic Training Signals Promote Hierarchical Syntactic Generalization in Neural Networks

Aditya Yedetore (yedetore@bu.edu)
Najoung Kim (najoung@bu.edu)
Boston University

Background:

- Humans generalize hierarchically, but neural networks generalize linearly. Why?

(H0) Nature: Innate biases [1]
(H1) Nurture: External evidence [2]

In prior work: Neural networks trained on form alone generalize linearly [3,4]

- However, children learn language from both form and meaning

Our hypothesis: Networks trained on both form and meaning will generalize hierarchically

Methods:

- We compared the generalization of networks trained on form alone with those additionally trained to translate forms to meaning (as in [5])

Results

- LSTMs and Transformers trained on form alone: 98% linear, 2% hierarchical (replicating [3])
- LSTMs trained on form & meaning: 0% linear, 100% hierarchical
- Transformers trained additionally on meaning: 38% linear, 62% hierarchical

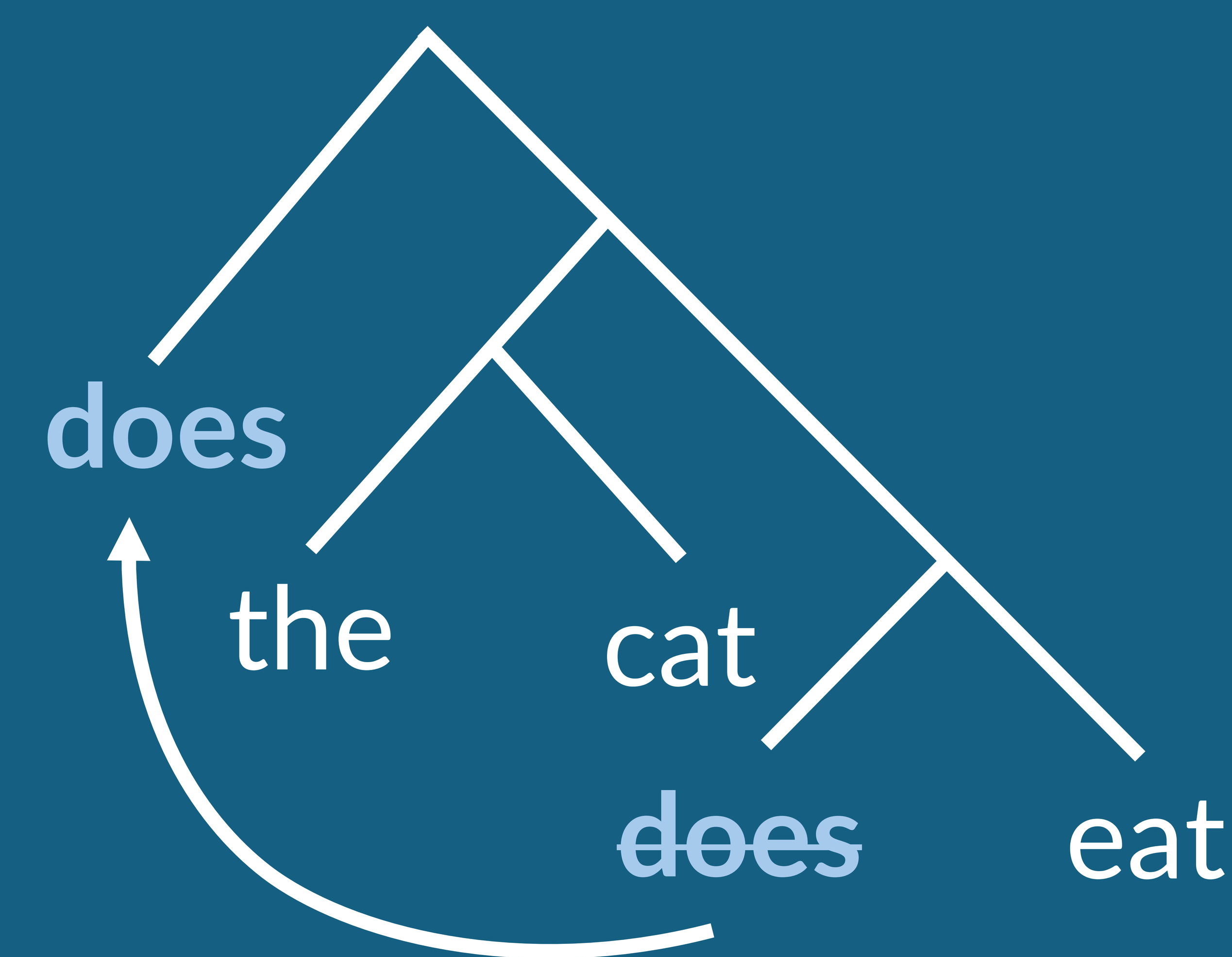
References

[1] Chomsky, Noam. *Reflections on Language*. United Kingdom, Pantheon Books, 1975.
[2] John Lewis and Jeffrey Elman. 2001. Learnability and the statistical structure of language: Poverty of stimulus arguments revisited. *Proceedings of the 26th Annual Boston University Conference on Language Development*, 1.
[3] R. Thomas McCoy, Robert Frank, and Tal Linzen. 2020. Does Syntax Need to Grow on Trees? Sources of Hierarchical Inductive Bias in Sequence-to-Sequence Networks. *Transactions of the Association for Computational Linguistics*, 8:125–140.
[4] Aditya Yedetore, Tal Linzen, Robert Frank, and R. Thomas McCoy. 2023. How poor is the stimulus? Evaluating hierarchical generalization in neural networks trained on child-directed speech. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 9370–9393, Toronto, Canada. Association for Computational Linguistics.
[5] Kim N, Linzen T. 2020 COGS: a compositional generalization challenge based on semantic interpretation. In *Proc. Of the 2020 Conf. on Empirical Methods in Natural Language Processing (EMNLP*, pp. 9087-9105. Online: Association for Computational Linguistics

Neural networks generalize more like humans when trained on both form & meaning than when trained on form alone.

Form Alone
→ **Linear Rule:**

Form & Meaning
→ **Hierarchical Rule:**



Form:

- The cat does eat.
- Does the cat eat?

Meaning:

- Eat(ιx.Cat(x))

My prior work on hierarchical generalization in neural networks [4].

Training Data

Example *Form* Training Instances

(1) a. The cat does eat. DECLARATIVE
The cat does eat.
b. The cat does eat. QUESTION
Does the cat eat?

Example *Meaning* Training Instance

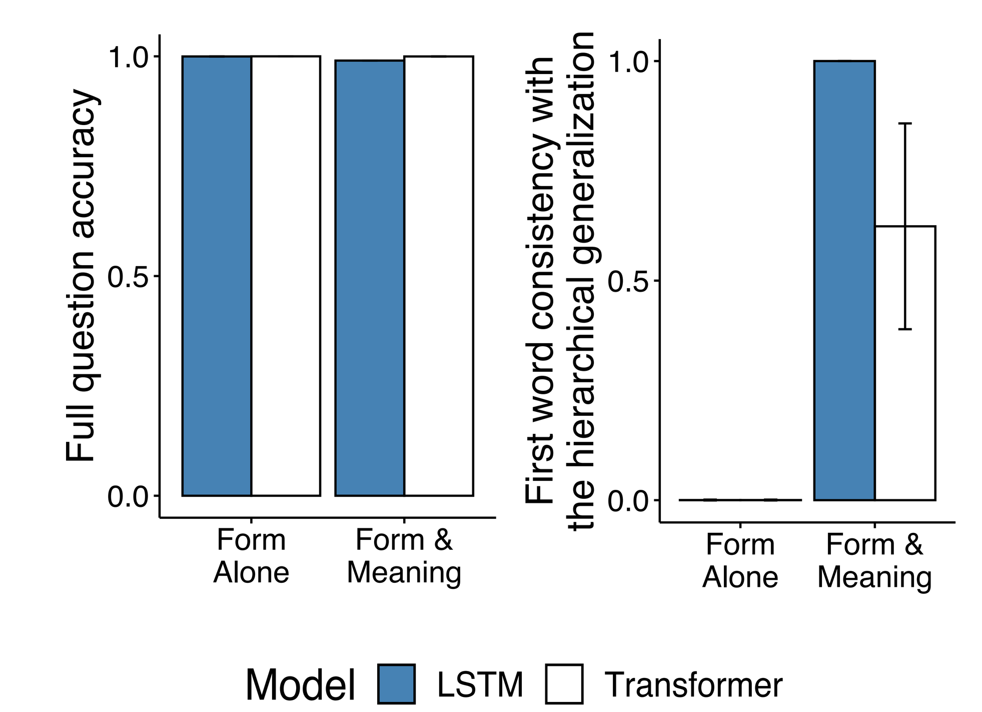
(2) The cat does eat. TRANSLATE
Eat(ιx.Cat(x))

Hierarchical Evaluation Data

Example Generalization Instance

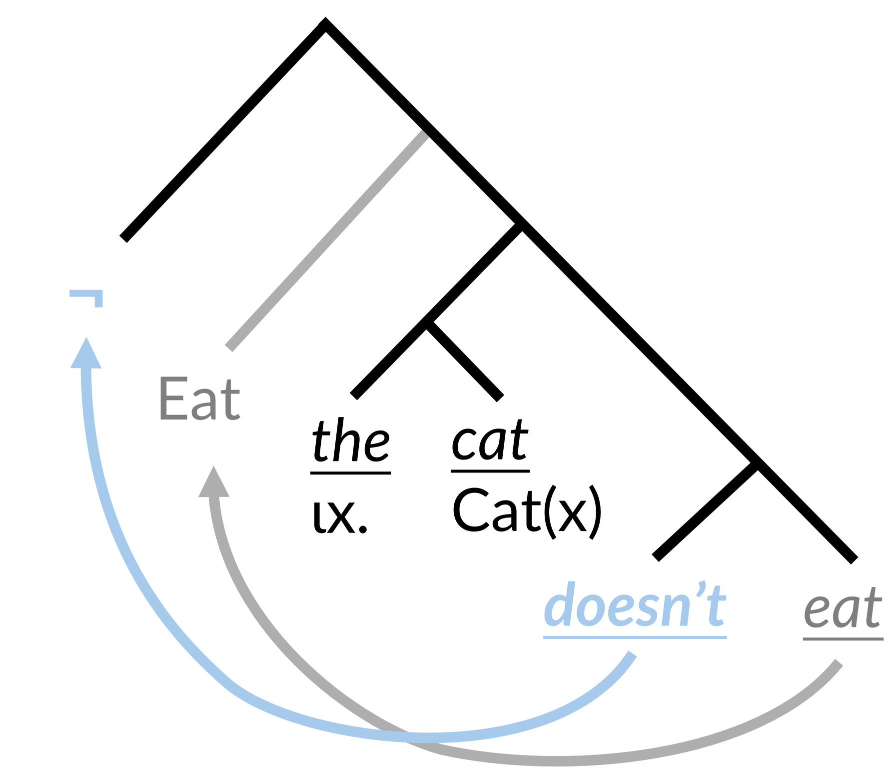
(3) The cat that does eat is happy QUESTION
a. Is the cat that does eat happy?
b. *Does the cat that eat is happy?

Test Set Results **Gen Set Results**



A Shared Representation for Form and Meaning

(4) The cat doesn't eat. TRANSLATE
¬Eat(ιx.Cat(x))



BOSTON UNIVERSITY