Language Models Can Learn Exceptions to Syntactic Rules Cara Su-Yi Leong and Tal Linzen

Restrictions on Passives

- •Artificial neural networks can generalize productively to novel contexts (e.g. Hupkes et al., 2020; Kim and Linzen, 2020; Lake and Baroni, 2018).
- \rightarrow Can they also *learn exceptions* to productive rules?
- •English passives are productive (Pinker et al., 1987; Brooks and Tomasello, 1999) but contain lexical restrictions for which there is often *no negative evidence* in the linguistic input.
- (1) a. The dog ate four bones. regular b. Four bones were eaten by the dog.
- (2) a. The dog murfed four balloons. nonce b. Four balloons were murfed by the dog.
- (3) a. The muffin cost four dollars. exceptional b. * Four dollars was/were cost by the muffin.

Does exposure to (human-scale) linguistic data provide sufficient indirect evidence to learn exceptions to passivization?

Method and Materials

- •We collected acceptability judgments on active and passive sentences containing verbs in five test verb classes and two control verb classes.
- \rightarrow If model judgments match human judgments in gradience and exceptionality, then the model learned human-like exceptions.
- •Human judgments: 84 Prolific participants (10 excluded) rated sentence acceptability from 0-100.
- •Model scores: sentence scores for each sentence were obtained from 5 GPT-2 models trained on 100M words of OpenWebText by summing the log probabilities of each token in the sentence.

Martin D S Braine and Patricia J Brooks. 1995. Verb Argument Structure and the Problem of Avoiding an Overgeneral Grammar. In Michael Tomasello. 1999. Young children learn to produce passives with nonce verbs. Developmental Psychology, 35(1):29. * Dieuwke Hupkes, Verna Dankers, Mathijs Mul, and Elia Bruni. 2020. COGS: A compositional is passives with nonce verbs. Developmental Psychology, 35(1):29. * Dieuwke Hupkes, Verna Dankers, Mathijs Mul, and Elia Bruni. 2020. COGS: A compositional is passives with nonce verbs. Developmental Psychology, 35(1):29. * Dieuwke Hupkes, Verna Dankers, Mathijs Mul, and Elia Bruni. 2020. COGS: A compositional is passives with nonce verbs. Developmental Psychology, 35(1):29. * Dieuwke Hupkes, Verna Dankers, Mathijs Mul, and Elia Bruni. 2020. COGS: A compositional is passives with nonce verbs. Developmental Psychology, 35(1):29. * Dieuwke Hupkes, Verna Dankers, Mathijs Mul, and Elia Bruni. 2020. COGS: A compositional is passive with nonce verbs. Developmental Psychology, 35(1):29. * Dieuwke Hupkes, Verna Dankers, Mathijs Mul, and Elia Bruni. 2020. COGS: A compositional is passive with nonce verbs. Developmental Psychology, 35(1):29. * Dieuwke Hupkes, Verna Dankers, Mathijs Mul, and Elia Bruni. 2020. COGS: A compositional is passive with nonce verbs. Developmental Psychology, 35(1):29. * Dieuwke Hupkes, Verna Dankers, Mathijs Mul, and Elia Bruni. 2020. COGS: A compositional is passive with nonce verbs. Developmental Psychology, 35(1):29. * Dieuwke Hupkes, Verna Dankers, Mathijs Mul, and Elia Bruni. 2020. COGS: A compositional is passive with nonce verbs. Developmental Psychology, 35(1):29. * Dieuwke Hupkes, Verna Dankers, Mathijs Mul, and Elia Bruni. 2020. COGS: A compositional is passive with nonce verbs. Developmental Psychology, 35(1):29. * Dieuwke Hupkes, Verna Dankers, Mathijs Mul, and Elia Bruni. 2020. COGS: A compositional is passive with nonce verbs. Developmental Psychology, 35(1):29. * Dieuwke Hupkes, Verna Dankers, Mathi generalization challenge based on semantic interpretation. In Proceedings of the 2020 EMNLP. & Brenden M. Lake and Marco Baroni. 2018. Generalization without Systematicity: On the Compositional Skills of Sequence-to-Sequence Recurrent Networks. In Proceedings of the 35th ICML 2018. & Tal Linzen, Emmanuel Dupoux, and Yoav Goldberg. 2016. Assessing the Ability of LSTMs to Learn Syntax-Sensitive Dependencies. TACL, 4:521–535. & Steven Pinker, David S. Lebeaux, and Loren Ann Frost. 1987. Productivity and constraints in the acquisition of the passive. Cognition, 26(3):195–267.

Results





A Potential Explanation?