Generalizing Without Evidence: Inferring Syntactic Rules From Sparse Input

Longstanding debates between theories of language acquisition center on whether children possess innate syntactic rules or learn them from experience with linguistic input. Computational models address this question via simulation of language learning processes, but what specific input features are necessary for grammatical generalization of specific constructions remains unclear. This study addresses that gap by focusing on English determiners, which provide an ideal test case as early-acquired function words whose syntactic regularity offers a clear window into category formation. Using a Transformer-based model (particularly BERT [1]), we investigate whether abstract grammatical rules for the determiner class can be acquired from input where direct evidence for the rule has been removed.

To isolate the role of input variability, a set of BERT models were trained incrementally from scratch on child-directed speech, following [2]. The training data was extracted from the LDP corpus [3], which documents the real-time linguistic environments of multiple English-speaking children. The experiment featured five conditions, ranging from natural, unrestricted data to a completely restricted condition where each noun was paired with only 'a' or only 'the', but never both. The models' ability to generalize was evaluated on their predictions of masked determiners in child-produced sentences containing nouns they had seen with only a single determiner during training.

The results show that all models, including the one trained on the most restricted input, successfully generalized, using determiners in novel combinations at a substantial rate (over 30% of determiner predictions were generalizations). Interestingly, this behavior was present even in models trained on children's earliest data. However, while all models learned to identify the syntactic category for a determiner at similar rates, their accuracy for predicting the specific determiner that was masked ('a' vs. 'the') decreased as input restrictions increased. Accuracies can be seen dropping from 72% in the unrestricted model to 55% in the fully restricted model.

These findings suggest that direct exposure to determiner-noun variability is not necessary for acquiring a grammatical rule. Instead, the models appear to use broader contextual and distributional cues [4] to build syntactic categories. This result is in line with findings that show that infants can extract grammatical patterns long before they use them productively [5], suggesting a potential dissociation between obtaining knowledge of a category and the ability to use that knowledge. In sum, this work supports the idea that statistical learning mechanisms can enable children to form grammatical representations from relatively limited input, underscoring the value of computational modeling for researching the exact mechanisms of language acquisition.

References

[1] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 4171–4186. Association for Computational Linguistics.

[2] Alhama, R. G., Foushee, R., Byrne, D., Ettinger, A., Goldin-Meadow, S., & Alishahi, A. (2023). Linguistic Productivity: the Case of Determiners in English. Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics, 330–343. https://doi.org/10.18653/v1/2023.ijcnlp-main.21


[3] Goldin-Meadow, S., Levine, S. C., Hedges, L. V., Huttenlocher, J., Raudenbush, S. W., & Small, S. L. (2014). New Evidence About Language and Cognitive Development Based on a Longitudinal Study: Hypotheses for Intervention. The American Psychologist, 69(6), 588–599. https://doi.org/10.1037/a0036886

[4] Redington, M., Chater, N., & Finch, S. (1998). Distributional Information: A Powerful Cue for Acquiring Syntactic Categories. Cognitive Science, 22(4), 425–469. https://doi.org/10.1207/s15516709cog2204_2

[5] Lany, J., & Saffran, J. R. (2010). From Statistics to Meaning. Psychological Science, 21(2), 284–291. https://doi.org/10.1177/0956797609358570