Learning the generative principles of a linguistic system from limited examples

Lei Yuan, Violet Xiang, David Crandall, & Linda Smith

One of the motivating questions of this year's special session is: How are children exposed to a small set of linguistic input but are able to master a complete linguistic system that allows for infinite generalization? Sharing the organizers' vision, we view this question as an important step to understand mature language processing and to arrive at important theoretical unifications about human cognition in general. Proposals to the above question typically either: a) emphasize rapid generalization based on prior principles that are language-specific and part of human core knowledge system, or b) highlight the generative principles discovered through more domain-general mechanisms, including associative learning mechanism. But even in its most advanced form of deep learning models, associative mechanism has been criticized as data-hungry and limited in generalization, thus falling short of accounting for human language learning. In these debates, little attention has been paid to the structure of input data—on which proposed learning mechanisms must operate. Thus, the goal of the current research is to better understand how associative mechanism interfaces with the statistical structure of input data to produce ***far generalization*** of a ***linguistic system*** based on ***limited input data***.

We address the above question by focusing on the language system that underlies the Arabic multi-digit number symbols and their spoken names. As a relatively recent human-invention, this system is less likely to have innately-involved structures, with much evidence suggesting that formal education is necessary to acquire this system and that many school-aged children struggle to do so. But it is also a system with many overlapping features that exhibit a small-world like structure as shown in Fig. 1 (Left), in which multiple redundant, degenerate, imperfect but inter-predictive features offer multiple pathways to the to-be-learned generalizable principles. We ask: How are place value terms (e.g., “hundred") combined in conjunction with single-digit number names (e.g., “three”) to create an in-theory infinite set of possible expressions such as "three hundred gazillions and five"? We hypothesize that a suite of these inter-correlated imperfect predictive components can allow an associative learning mechanism to generalizations that accord with generative principles and can do so despite limited training data.

In Study 1 and 2, 148 preschool children were randomly assigned to either a training condition in which they were given experience with alignable pairs of written multi-digit numbers and their corresponding spoken names in casual learning activities such as storybook reading (e.g., “Johnny wants to save money to buy a new toy car. How much does it cost? Look! It costs forty-two dollars [the experimenter pointed to the written “4” and “2” in sequence].”), or a control condition in which they saw the same material but were shown letters (e.g., “CAR”) rather than numbers. Across three days of 15-minutes daily exposure to a small set of 36 unique numbers, only children in the training condition (but not the control condition) showed significant accuracy improvement from pre- to post-test in recognizing a novel, never-before-seen multi-digit number (e.g., Which is one hundred twenty-five? 125 vs. 251), as shown in Fig. 1 (Right). Study 3 used a deep learning model—as shown in Fig. 2—to provide evidence that the co-predictive properties between number names and their written forms, albeit imperfect and local predictors, are sufficient for an associative learner to make far and systematic generalizations without explicitly representing any rules or principles. As shown in Fig. 3, after training using the same small set of input data as children in Study 1 & 2, the models not only demonstrated significant learning (based on four complementary measures), but also showed the same error pattern as children in responding to items that vary in difficulty levels.

This result—that associative mechanism can lead to far generalization when operated on limited input data with statistical structure that is conducive to learning and in fact pervasive in many real-life domains—is important for understanding how limited data *with a particular statistical structure* gives rise to far generalizations. Implications for learning natural language will be considered.

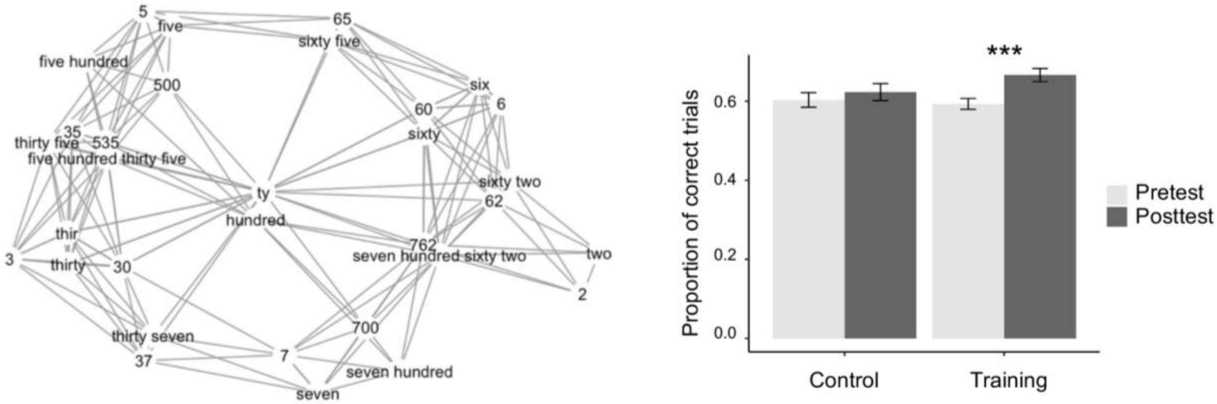

Figure 1. Left: Illustration of the massive overlapping and redundant partial mappings within and between written number symbols and names for four randomly chosen numbers 37, 65, 535 and 762. Nodes depict individual components; edges depict co-occurrences and partial mappings among the nodes. Right: Pre- and post-test performance for children in the training and control condition in Study 1 & 2, showing significant improvement for the training (but not the control) condition after merely three daily 15 mins exposure with 36 unique numbers.

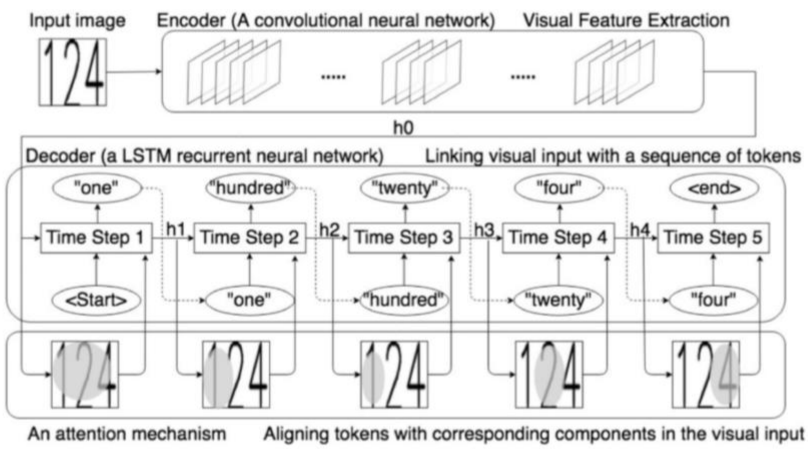

Figure 2. Illustration of the model architecture: 1) A convolutional neural network (CNN) as an encoder for extracting visual features from an input image, 2) A Long short-term memory (LSTM) recurrent neural network (RNN) as a decoder for linking visual input with a sequence of tokens (i.e., components in the number name), and 3) An attention mechanism that learns to align tokens with corresponding parts in the visual input (the shaded region represents the parts of the image that are most relevant to the token in the current time step).

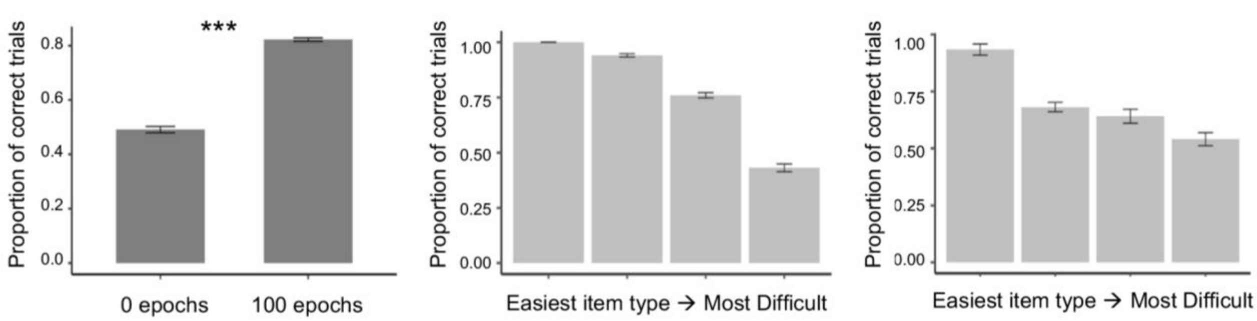

Figure 3. Left: model performance before and after training. Middle: model's after training performance for four different types of testing items that vary in difficulty. Right: children's after training performance for the same four different types of testing items that vary in difficulty. DNN models not only learned from the training but also exhibited the same error patterns as human children.