Colorless Green Ideas Sleep Furiously Revisited: A Statistical Perspective

Florencia Reali (fr34@cornell.edu)
Rick Dale (rad28@cornell.edu)
Morten H. Christiansen (mhc27@cornell.edu)

Department of Psychology; Cornell University; Ithaca, NY 14853 USA

Abstract

In the present study we provide empirical evidence that human learners succeed in an artificial-grammar learning task that involves recognizing grammatical sequences whose bigram frequencies from the training corpus are zero. This result begs explanation: Whatever strategy is being used to perform the task, it cannot rely on the simple co-occurrence of elements in the training corpus. While rule-based mechanisms may offer an account, we propose that a statistical learning mechanism is able to capture these behavioral results. A simple recurrent network is shown to learn sequences that contain null-probability bigram information by simply relying on distributional information in a training corpus. The present results offer a simple but stark challenge to previous objections to statistical learning approaches to language acquisition that are based on sparseness of the primary linguistic data.