



SCHOOLING CHOMSKY'S IMAGINATION: UNIFYING MORPHOLOGY, SEMANTICS, AND POS TAGGING AS VECTOR COMPUTATION

Dr. Chris Westury

Professor, Department of Psychology
Cognition Seminar

www.psych.ualberta.ca/~westburylab/

Friday, October 27

3:00 PM - 4:30 PM

BS P-319N



In his (in)famous (1956) paper 'Three Models for the Description of Language', Chomsky made up the sentence 'colorless green ideas sleep furiously' to motivate the idea that "Whatever the other interest of statistical approximation [...] may be, it is clear that it can shed no light on the problems of grammar." (p. 116) Because the first-order co-occurrence probabilities of the words in his sentence are approximately zero, Chomsky concluded that language learners cannot possibly rely on statistical properties of language to make syntactic judgments. He later extended this claim to argue that children only received positive evidence which was insufficient for them to learn language, his 'poverty of the stimulus argument' (Chomsky, 1980), sometimes called the 'logical problem of language acquisition'. Recent work unifying predictive co-occurrence models (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013) with animal models of discriminant learning (Rescorla & Wagner, 1972) suggest that Chomsky's clear intuitions were incorrect. We demonstrate that it is possible to extract accurate information about syntactic category and morphological family membership directly from patterns of word co-occurrence. Since discriminant learning depends on confirming or disconfirming predictions, it supplies the negative evidence that Chomsky failed to notice.

