Cynthia Fisher, PhD

Psychology Department

University of Illinois

**Title:** The origins of syntactic bootstrapping: A computational model

**Abstract:**  To learn the words and the grammar of their native language, children must analyze sentences into their constituent parts, and work out how the composition and arrangement of these constituents conveys abstract relational meanings, telling us *who does what to whom*. Traditional accounts assume children solve this problem largely because they can often infer the meaning of input sentences. The syntactic bootstrapping account proposes that even young children use partial knowledge of syntax to guide sentence interpretation and word learning.  But how could syntactic structure guide sentence interpretation before children have learned much about the syntax of the native language?  We present a computational model of the origins of syntactic bootstrapping, based on systems for automatic semantic-role labeling (SRL).  SRL models learn to identify sentence constituents that fill semantic roles, and to determine their roles. The 'BabySRL' instantiates the structure-mapping account of syntactic bootstrapping (Fisher et al., 2010). On this account, sentence comprehension is grounded in noun learning and skeletal expectations about linguistic predicate-argument structure.  Once some nouns are known, children can construct a partial sentence representation. Children treat each noun as a candidate argument, and thus find the number of nouns in this representation inherently meaningful. We used the computational model to investigate the consequences of these assumptions for learning from natural corpora of child-directed speech. The BabySRL simulations yield strong evidence for the structure-mapping account: partial sentence representations consisting of an ordered set of nouns (a) are useful as a foundation for further syntactic learning, and (b) these representations can be built via unsupervised distributional analysis and a proposed algorithm for identifying word clusters as nouns or verbs. This model makes explicit one way in which distributional learning, via simple innate expectations about linguistic predicate-argument structure, could provide an independent source of information about sentence meaning.