

# **Abstract Syntactic Knowledge or Limited-Scope Formulae: A Computational Study of Childrens Early Utterances**

**Qihui Xu**

Graduate Center, City University of New York, New York, New York, United States

**Martin Chodorow**

Graduate Center, City University of New York, New York, New York, United States

**Virginia Valian**

Graduate Center, City University of New York, New York, New York, United States

**Xiaomeng Ma**

Graduate Center, City University of New York, New York, New York, United States

## **Abstract**

Do childrens early utterances reflect abstract syntactic knowledge or slot-filler formulae developed through word imitation? This study compares development of part-of-speech (POS) sequences with word sequences using language models (LMs) trained on mothers utterances (N=1,272,139) from CHILDES English corpora, in which POS tags are automatically assigned by MOR and POST programs (MacWhinney, 2000). Word-based and POS-based LM probabilities for childrens multi-word utterances in the Providence corpus (Brschinger et al., 2013, 15-36 months, Nchildren=6, Nutterances=50,717) were calculated as a function of age. Word-based LM probability of childrens multi-word utterances first increases with age and then levels off after 23 months. By contrast, POS-based probability remains high and stable across all ages. This suggests children have adult-like syntactic knowledge even at a very early age when their word sequences are still not adult-like. The pattern of results supports the abstract syntax view. Additional studies will use more accurate POS-taggers and larger datasets.