

Surprisal and developmental sentence processing: exploring the role of language exposure through neural language models

Kuan-Jung Huang

University of Maryland College Park, College Park, Maryland, United States

Roger Levy

Massachusetts Institute of Technology, Cambridge, Massachusetts, United States

Yi Ting Huang

University of Maryland College Park, College Park, Maryland, United States

Abstract

Children’s language processing differs from adults’ in idiosyncratic ways. Within adult data, various incremental processing phenomena have been shown to be predicted by neural language models (LMs) using surprisal as the linking hypothesis, where processing effort is determined by a word’s log inverse conditional probability. Since LMs seem without strong inductive bias for natural-language-specific structures, with their word predictions determined by training on naturalistic data, these results potentially support exposure-based theories. However, it remains unclear how well LMs explain the developmental trajectory of human language processing. Here we evaluate LMs with developmentally-realistic training data—how well they predict six established child language processing phenomena, including cases where child and adult patterns differ. Our LMs correctly predict four of the six, but fail in cases involving thematic-role and pragmatic knowledge. Our results highlight the limitations of language-exposure-based theories and call for further empirical research on human language processing patterns throughout development.