

Are Global Statistics Discarded Statistics? An Investigation of What Types of Co-Occurrence Statistics Could Support the Acquisition of Semantic Knowledge

Molly Niehaus

Carnegie Mellon University, Pittsburgh, Pennsylvania, United States

Tijana Slepcev

Carnegie Mellon University, Pittsburgh, Pennsylvania, United States

Catarina Vales

Carnegie Mellon University, Pittsburgh, Pennsylvania, United States

Abstract

Much research examining the development of semantic knowledge has focused on patterns of co-occurrence in language. An often-made distinction concerns 'local' co-occurrences (word pairs that appear close together) and 'global' co-occurrences (word pairs that appear in similar linguistic contexts), with some recent accounts arguing that only local co-occurrences are exploitable by young children. We tested the alternative view that this dichotomy may be misaligned with natural language input. First, we performed a descriptive analysis of the co-occurrence structure present in the input. This analysis suggests that global co-occurrences are frequent in the input to children. Second, we performed a descriptive analysis of the semantic information encoded by local and global statistics in the CHILDES corpus; this analysis suggests that both local and global co-occurrences encode patterns of similarity that could support the acquisition of structured semantic representations.