Revisiting the debate: Is cross-situational word learning too computationally complex for infants in the real world?

Priscilla Fung & Elizabeth K. Johnson

University of Toronto

In highly simplified lab settings, infants can resolve referential ambiguity by tracking co-occurrence probabilities across situations (cross-situational word learning, henceforth CSWL). But is CSWL possible in the real world? Work using the Human Simulation Paradigm (HSP) – where participants watch muted videos with a beep occurring each time a 'mystery word' is uttered – suggests multiple labelling utterances do not improve word learning; instead, infants acquire words from a single encounter. But the original HSP lacks ecological validity. First, overt guesses after each labelling instance may introduce memory and/or attentional strategies into the learning process. Second, muted videos remove the visual prosody between labels and events in the visual scene – information that has been shown to support early referent-label mapping. It is possible that infants could use this information to limit the range of hypotheses to be considered in CSWL. In Experiment 1, we address this first point. Adults watched 20-s vignettes of caregivers teaching target words to their 11- or 18-month-olds. A visual cue indicated the mystery word's occurrence. Participants (n=72) watched either 1 or 6 vignettes before making a guess. Participants' performance was much stronger after watching 6 vignettes, suggesting cross-situational information helps resolve referential ambiguity. In Experiment 2 (pre-registered), we asked whether visual prosody supports cross-situational word learning, and whether multilingualism might strengthen CSWL skills. We included muted (as in Experiment 1) as well as low-pass filtered trials (obscuring word identity but retaining visual prosody). Preliminary results (n=73; target n=144) support both of our predictions. Participants' accuracy is higher with low-pass filtered speech than muted vignettes and bilinguals outperform monolinguals. Though not yet statistically significant, the effect of low-pass filtered speech is stronger in interactions with 11- than 18-month-olds, in line with research suggesting that visual prosody cues to labelling intent are stronger in speech to younger infants.