English-learning children hear and use multiple meanings for words in early speech

Many theories hold that children initially expect that a word will have only one meaning; this expectation is thought to facilitate vocabulary development, based on the assumption that most words that children encounter will be unambiguous.[1,2] Yet this assumption has rarely been tested. Here, we show that children not only hear words directed to them with multiple meanings from early in life, but also themselves use multiple word meanings in their own productions.

To collect the first large-scale estimate of the word meanings directed to and used by children, a combination of expert and crowd annotators tagged CHILDES[3] transcripts using the well-established WordNet sense inventory.[4] We chose two longitudinal home-based corpora, Manchester[5] and Providence,[6] yielding data for 18 children, 12 - 48 m.o. For tractability, we selected words in the North American English CDI (WS instrument),[7] and annotated up to 100 tokens of each word for each speaker in each 3-month time period. For each token, annotators selected one or more senses using a web app which displayed each target word in transcript context using childes-db.[8] The analyses here reflect ~60,000 meaning-tagged words spoken by children and ~120,000 by adults.

As a descriptive statistic, we report the number of unique senses per unique word types per time period for caregivers and the child from each family (Fig 1.) We fit two Bayesian logistic mixed effects regressions using the BRMS[9] package, predicting non-dominant vs. dominant sense usage for adults and children over time (Fig. 2). The dominant sense was defined as the most frequent sense within each time period for each speaker; non-dominant included all other senses. We treat child age as a fixed effect and include by-family random intercepts and age-by-family random slopes. Adults show an intercept of β_0 = -2.10, corresponding to a probability of non-dominant meaning usage of .12, which suggests children are exposed to significant lexical ambiguity in early development. Adults do not show a statistically significant increase over time (β_{age} = 0.01, 95% Bayesian credible interval = -0.004 - 0.025). Children also showed evidence of lexical ambiguity at the earliest age bin (β_0 = -2.51, CI = -2.79 - -2.23) corresponding to a probability of non-dominant meaning usage of .075. The model revealed a significant positive effect of child age (β_{age} = .03, CI = .02 - .04), suggesting that children have a growing command of multiple meanings.

We next investigated which forms of polysemy are most frequent in children's speech. We counted the number of word types that each child used in each combination of WordNet sense categories, normalized by child, and averaged across children to produce a *relative prevalence* score. The most prevalent form of nominal polysemy was artifact-quantity (ex. a *cup* / a *cup* of milk) and food-plant (*pumpkin* pie / *pumpkin* patch), and the most prevalent verbal polysemy was contact-motion (*get* that / *get* [over] here) and motion-change (Fig. 3). These colexifications highlight a set of conceptual relations that may be frequent and/or salient to children in early development. (500)

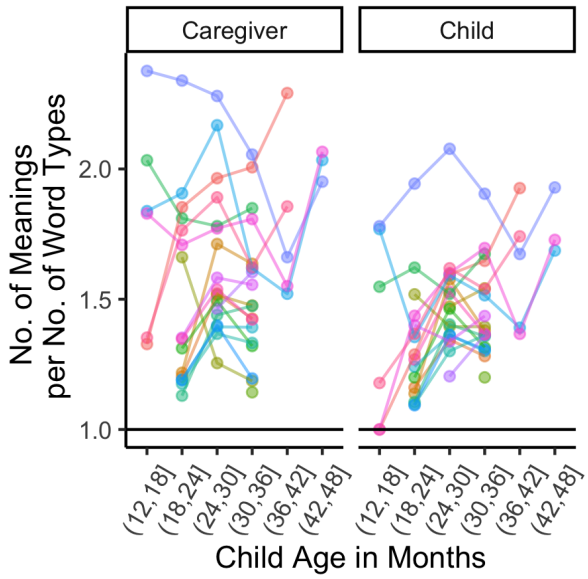

Fig. 1 Number of meanings per number of word types for 18 families in the Providence and Manchester Corpora

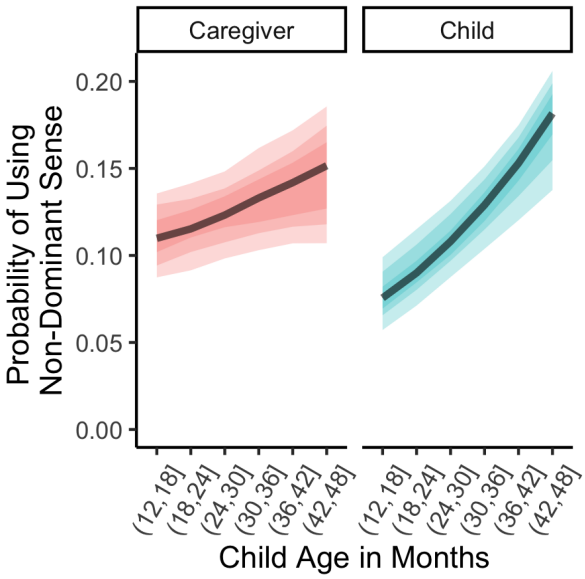

Fig. 2 Probability of non-dominant sense usage across developmental age under a Bayesian mixed effects logistic regression. Regions reflect 50, 80, and 95% CIs.

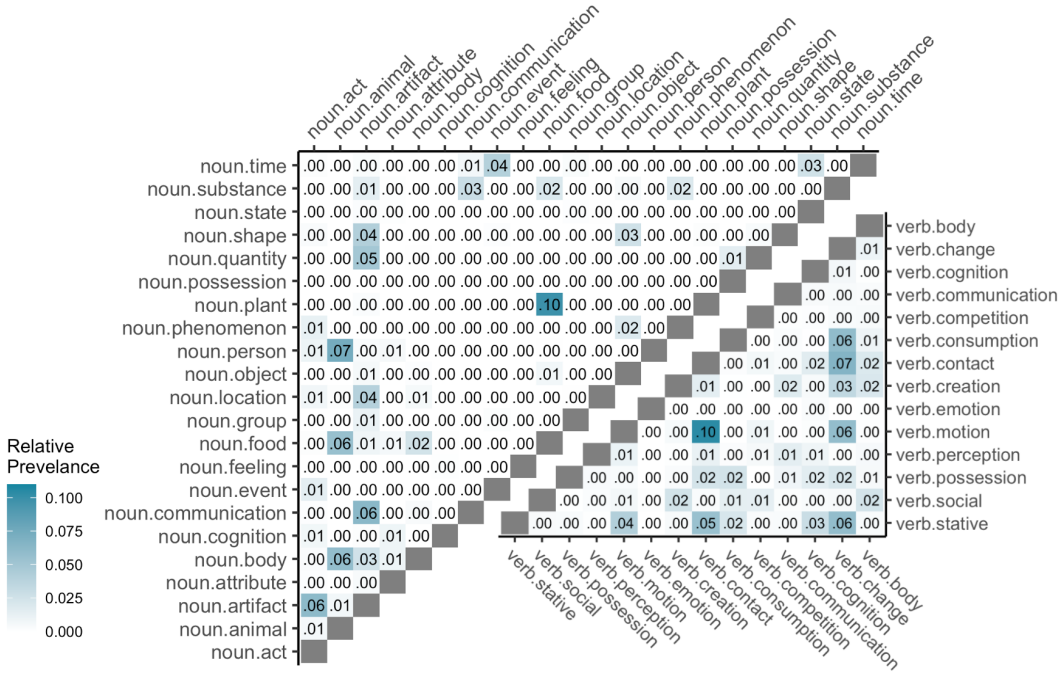

Fig 3. The relative prevalence of polysemous relationships: the proportion of each child's lexicon (in word types) that they use in a pair of categories, averaged across children.

[1]Markman, 1989 [2]Trueswell et al., 2013 [3]MacWhinney, 2000, [4]Miller et al., 1990 [5]Theakston et al., 2001 [6]Demuth et al., 2006, [7]Fenson et al., 2007 [8]Sanchez et al. 2019 [9]Bürkner et al., 2017