Quantitatively assessing the development of adjective ordering preferences using child-directed and child-produced speech corpora

Galía Bar-Sever, Rachael Lee, Gregory Scontras, and Lisa Pearl
University of California, Irvine

Adjective ordering preferences

We find this preference in many different languages, whether adjectives are pre- or post-nominal

 Lexical class hypothesis:
words are grouped into hierarchically-arranged lexical semantic classes

 Subjectivity hypothesis:
less subjective adjectives are preferred closer to the modified noun

Recent work by Scontras et al. (2017) and Hahn et al. (2017) suggests that the subjectivity hypothesis best accounts for adult knowledge

But how does this knowledge develop? And how can we tell which representation kids are using?

A process for analyzing the likelihood of child output given their input

To decide which representation hypothesis is active in children at a given age, we compare the predictions of each hypothesis with respect to the observed child input and behavior

What’s going on?

We need to test which representation hypothesis best accounts for the data kids produce given their input

There are two possible abstract adult representations that could be developing, but kids also could be repeating back the input frequencies in an item-based way

Which of these representations best accounts for kids’ use of multi-adj strings?

What are kids hearing?

Data taken from CHILDES North American & UK corpora, ages 2-4

688,428 child-directed utterances

What are kids saying?

1,069,406 child-produced utterances

Given the input, which hypothesis is best at generating the produced data?

Each row presents the logged probability scores for a given age:
more negative = less probable
Item-based input frequency best predicts the data before age 3
Abstract lexical class overtakes it at 4

In the future: What representations are children using across different languages? What happens to emerging representations in populations with delayed acquisition?