Computation of Language Laboratory
UC Irvine

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

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

Adjective ordering preferences

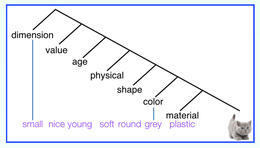
✓ "small grey kitten"

✗ "grey small kitten"

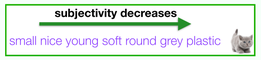
We find this preference in many different languages, whether adjectives are pre- or post-nominal

How do adults represent these preferences?

Lexical class hypothesis:
words are grouped into hierarchically-arranged lexical semantic classes
(Dixon, 1982; Cinque, 1994)



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?

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?

Child-directed speech

Data taken from **CHILDES** North American & UK corpora, **ages 2-4** (MacWhinney, 2000)

688,428 child-directed utterances

age	#strings	#adj tokens	#adj types
2	1440	2880	131
3	881	1762	128
4	745	1490	124

What are kids saying?

Child-produced speech

1,069,406 child-produced utterances

age	#strings	#adj tokens	#adj types
2	466	932	79
3	274	584	72
4	235	470	81

Of all the multi-adj strings, **3.46%** were **direct repetitions** and only **0.50%** of the strings were of a **child directly repeating an adult**

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

Child-produced speech

log(p(D|H))

age	input frequency	lexical class	subjectivity
2	**-202.6**	-334.9	-274.6
3	**-125.1**	-164.0	-163.0
4	-182.9	**-165.2**	-193.5

difference scores

lexical vs. best	subjectivity vs. best
-132.3	-72
-38.9	-37.9
0	-28.3

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

lexical class and subjectivity perform better as children age, demonstrating the **emergence of abstract knowledge**

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

"small grey kitten"
2-away **1-away**

Child-directed speech

$$p_2exp(adj_x)$$

expected probability of target adjective appearing the **2-away** position in the output

"small grey"

$$\frac{f_{2input}(adj_x)}{N_{input}(adj_x)}$$

number of times target adjective appeared in the **2-away** position / number of times target adjective appeared **at all**

OR

$$\frac{f_{input}(< adj_x) + 0.5 * f_{input}(= adj_x)}{N_{input}(adj)}$$

(# of adjs in a **closer** lexical class than the target + 0.5 × # of adjs in the **same** lexical class as the target) / **total number** of adjectives in the input

OR

subjectivity

$$\frac{f_{input}(< adj_x) + 0.5 * f_{input}(= adj_x)}{N_{input}(adj)}$$

(# of adjs with **lower** subjectivity than the target + 0.5 × # of adjs with **equal** subjectivity as the target) / **total number** of adjectives in the input

Child-produced speech

$$p(D(adj_x)|H) = \binom{N}{f}(p_2exp(adj_x))^f(1-p_2exp(adj_x))^{N-f}$$

how probable the actual distribution of the adjective in the output **data** D(adj) is **given** the representation **hypothesis** H

for each adj

product calculated over each adjective in the child's output

$$p(D|H) = \prod_{adj_x \in A} p(D(adj_x)|H)$$

Discussion

Using **corpus analysis** and **quantitative approaches**, we can see **when** more abstract underlying representations emerge for adjective ordering preferences: around **age 4**

It remains unclear **when subjectivity overtakes lexical class**; this likely depends on children's development of the **conceptual underpinnings** of subjectivity, which occurs late (Foushee & Srinivasan, 2017)

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

Cinque, G. (1994). On the evidence for partial N-movement in the Romance DP. In R. S. Kayne, G. Cinque, J. Koster, J.-Y. Pollock, L. Rizzi, & R. Zanuttini (Eds.), *Paths Towards Universal Grammar. Studies in Honor of Richard S. Kayne* (pp. 85–110). Washington DC: Georgetown University Press. **Dixon, R. M. (1982).** *Where have all the adjectives gone? and other essays in semantics and syntax* (Vol. 107). Walter de Gruyter. **Foushee, R., & Srinivasan, M. (2017).** Could both be right? Children's and adults' sensitivity to subjectivity in language. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (p. 379-3384). London, UK: Cognitive Science Society. **Hahn, M., Futrell, R., & Degen, J. (2017).** Exploring adjective ordering preferences via artificial language learning. Talk presented at the first California Meeting on Psycholinguistics, December 2-3, 2017, UCLA. **MacWhinney, B. (2000).** *The CHILDES project: The database (Vol. 2).* Psychology Press. **Scontras, G., Degen, J., & Goodman, N. D. (2017).** Subjectivity predicts adjective ordering preferences. *Open Mind: Discoveries in Cognitive Science*, 1, 53-65.

Contact: gbarseve@uci.edu