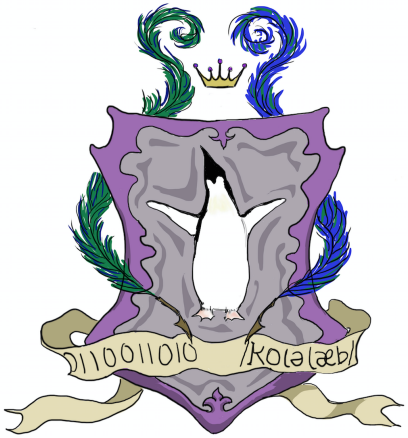
Small lexical learners:
The development of adjective ordering preferences

Bar-Sever, Lee, Scontras, & Pearl
presented by: Galia Bar-Sever
BUCLD 42, 2017



Computation of Language Laboratory

UC Irvine

“The small grey kitten”

“The grey small kitten”

“The small grey kitten”

“The grey small kitten”

robust adjective ordering preferences

not only in English,
but in **many different languages**
where adjectives occur either
pre- or **post-nominally**

small grey kitten

robust adjective ordering preferences

Mokilese

Hungarian

not only in English,
but in **many different languages**
where adjectives occur either
pre- or **post-nominally**

Selepet

Dutch

Telugu

Mandarin Chinese

kitten grey small

how do adults represent ordering preferences?

simple hypothesis: **repeat** back what you hear

how do adults represent ordering preferences?

simple hypothesis: **repeat** back what you hear

how do adults represent ordering preferences?

simple hypothesis: **repeat** back what you hear

how do adults represent ordering preferences?

however, in adults it seems like something **more abstract** is going on

how do adults represent ordering preferences?

adjectives group into
lexical semantic classes that are ordered

Dixon 1982, Cinque 2014

how do adults represent ordering preferences?

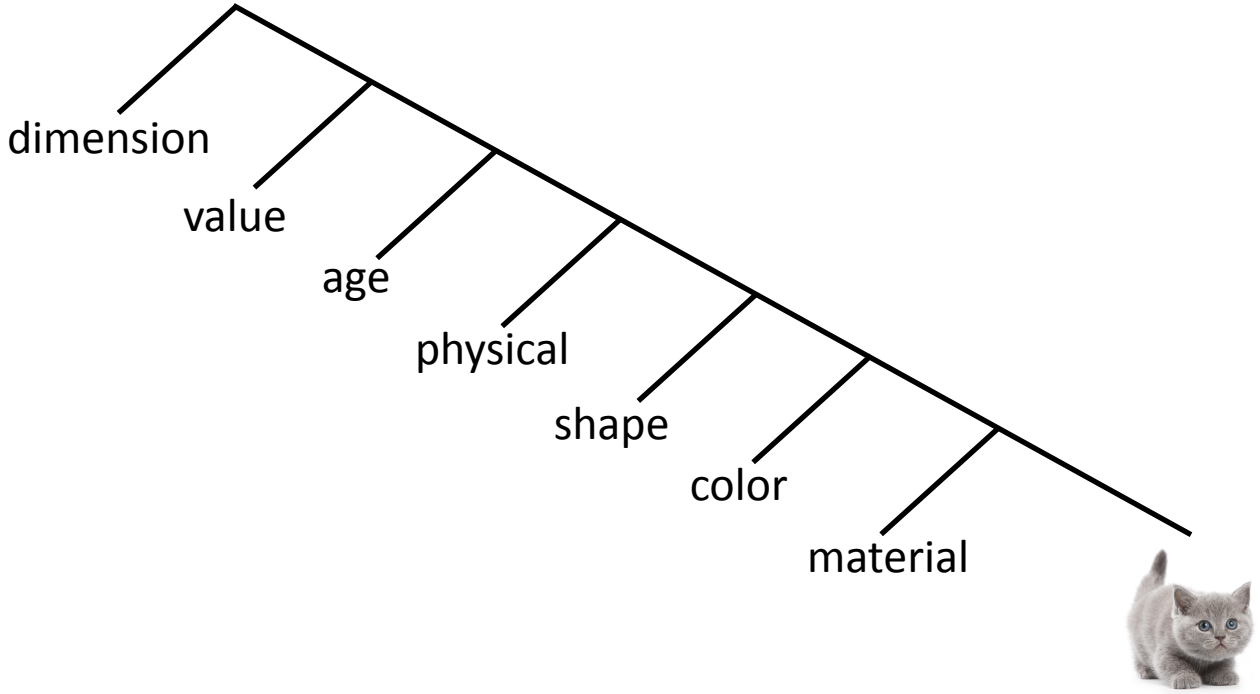
adjectives group into
lexical semantic classes that are ordered

lexical class ordering could be determined by
hierarchical abstract syntax

Dixon 1982, Cinque 2014

how do adults represent ordering preferences?

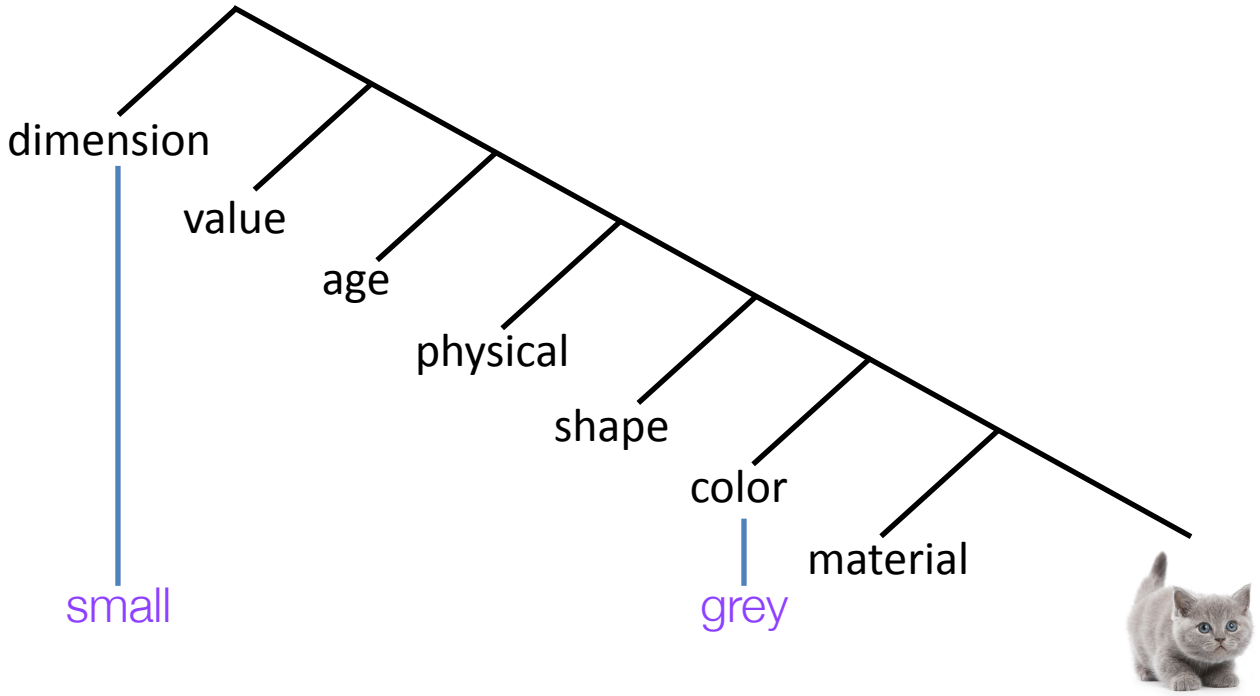
lexical class ordering could be determined by **hierarchical abstract syntax**



Dixon 1982, Cinque 2014

how do adults represent ordering preferences?

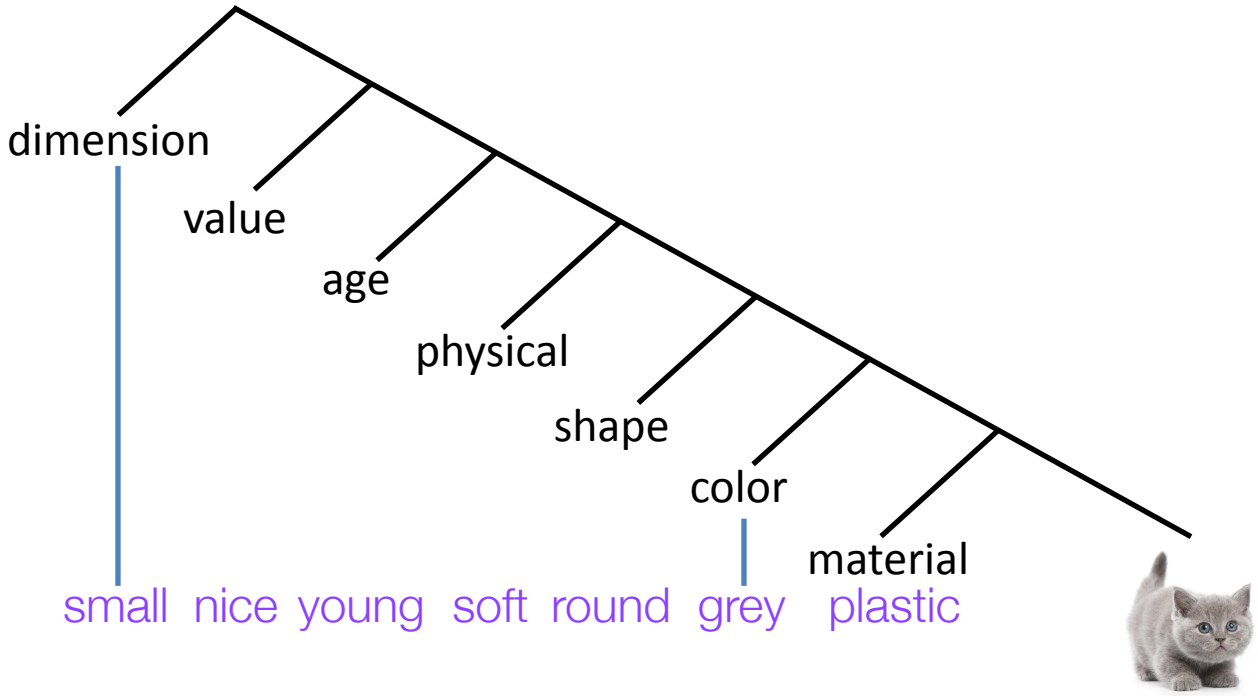
lexical class ordering could be determined by **hierarchical abstract syntax**



Dixon 1982, Cinque 2014

how do adults represent ordering preferences?

lexical class ordering could be determined by **hierarchical abstract syntax**



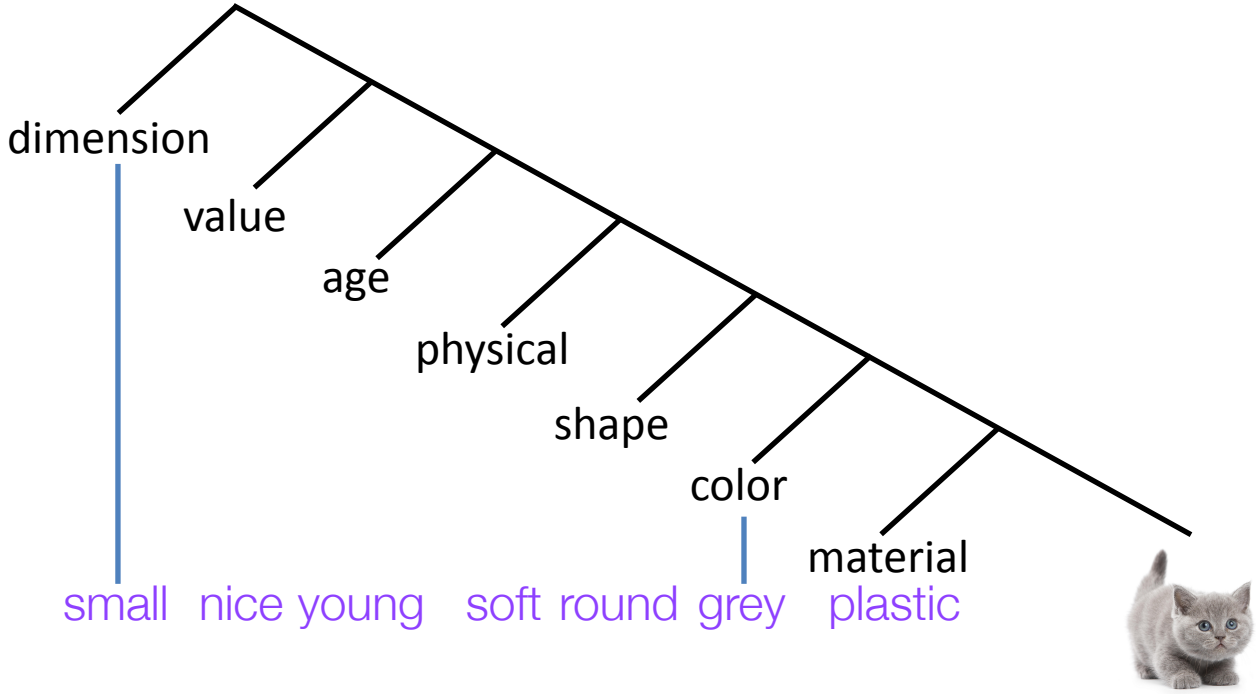
Dixon 1982, Cinque 2014

how do adults represent ordering preferences?

internal representation explicitly encodes hierarchical syntactic ordering of lexical semantic classes

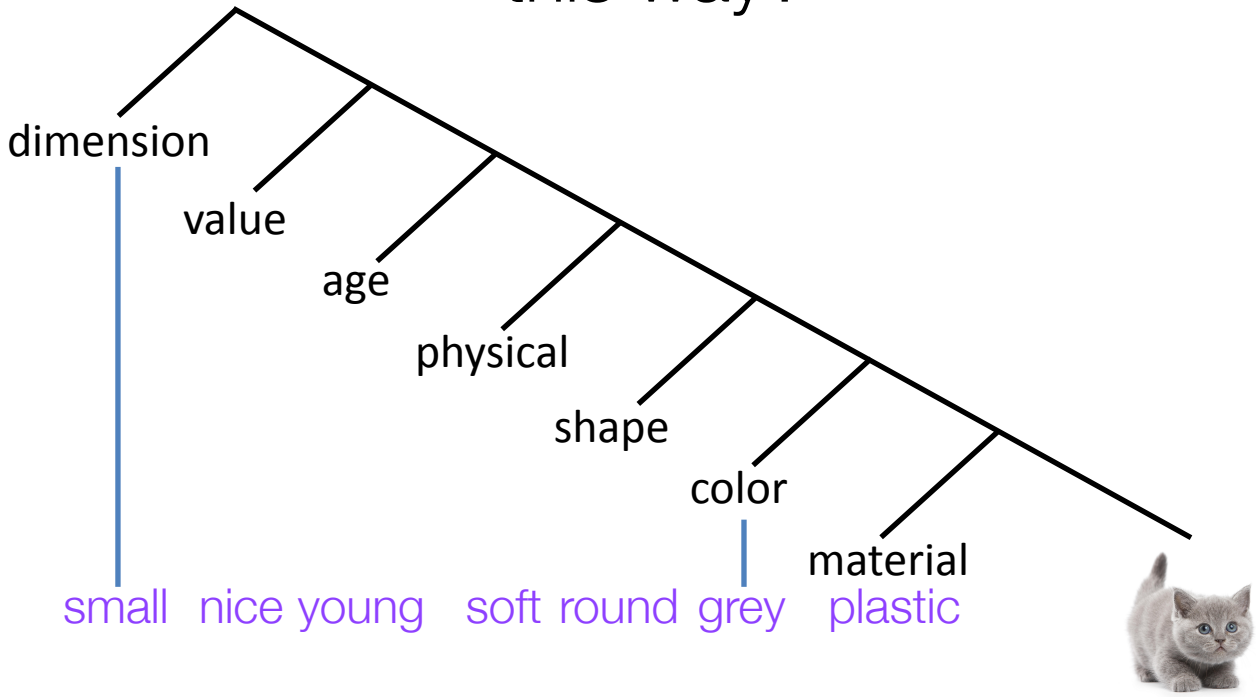
how do adults represent ordering preferences?

but **why this ordering**
of lexical semantic classes?



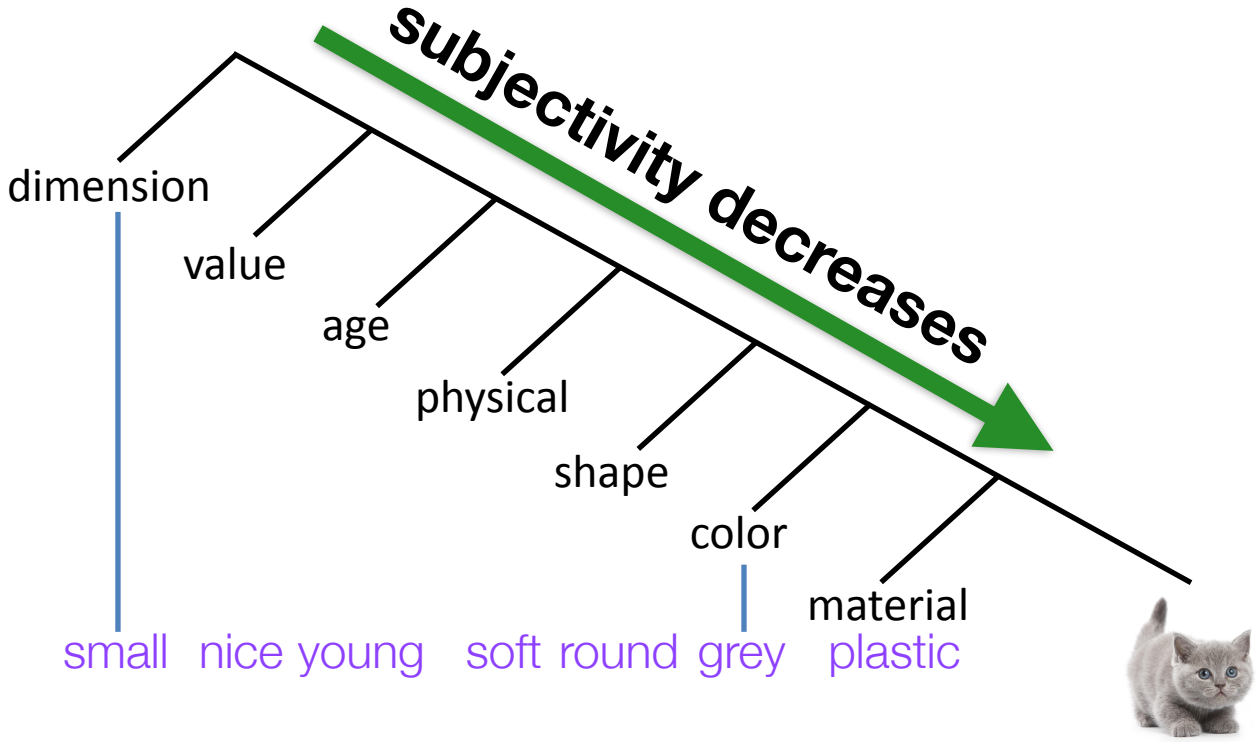
how do adults represent ordering preferences?

is there some deeper reason why the classes should be ordered in this way?



how do adults represent ordering preferences?

adults are sensitive to the relative **subjectivity** of the adjectives they are ordering



Scontras et al. 2017

how do adults represent ordering preferences?

adults are sensitive to the relative **subjectivity** of the adjectives they are ordering

the observed lexical ordering could derive from this **subjectivity** ordering

subjectivity decreases

small nice young soft round grey plastic

Scontras et al. 2017

operationalizing subjectivity

the faultless disagreement task

Scontras et al. 2017

operationalizing subjectivity

the faultless disagreement task

Scontras et al. 2017

operationalizing subjectivity

the faultless disagreement task

Scontras et al. 2017

operationalizing subjectivity

the faultless disagreement task

Scontras et al. 2017

operationalizing subjectivity

the faultless disagreement task

Scontras et al. 2017

operationalizing subjectivity

the faultless disagreement task

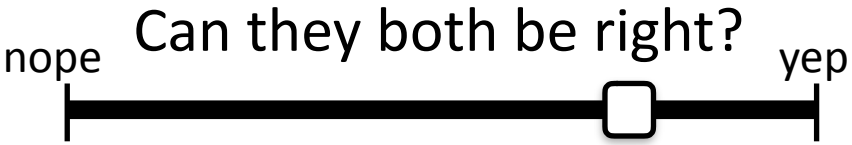
That kitten is small!

You're wrong! That kitten is **not** small!!

Scontras et al. 2017

operationalizing subjectivity

the faultless disagreement task

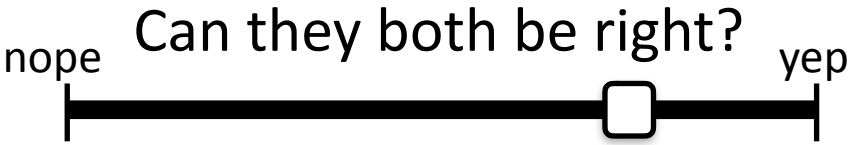


That kitten is small!

You're wrong! That kitten is **not** small!!

Scontras et al. 2017

operationalizing subjectivity



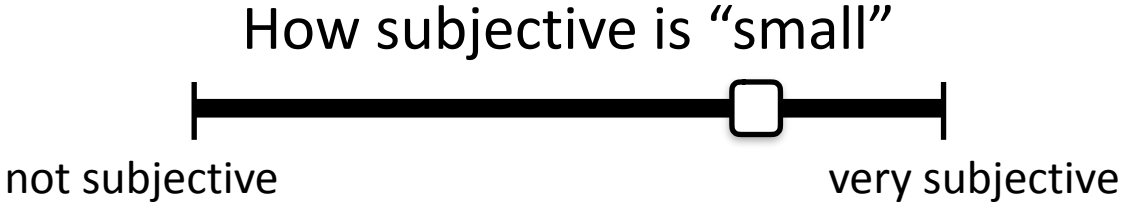
you might be more able to faultlessly disagree on whether something is “small” than you would on whether it is “grey”

“small grey kitten”

Scontras et al. 2017

operationalizing subjectivity

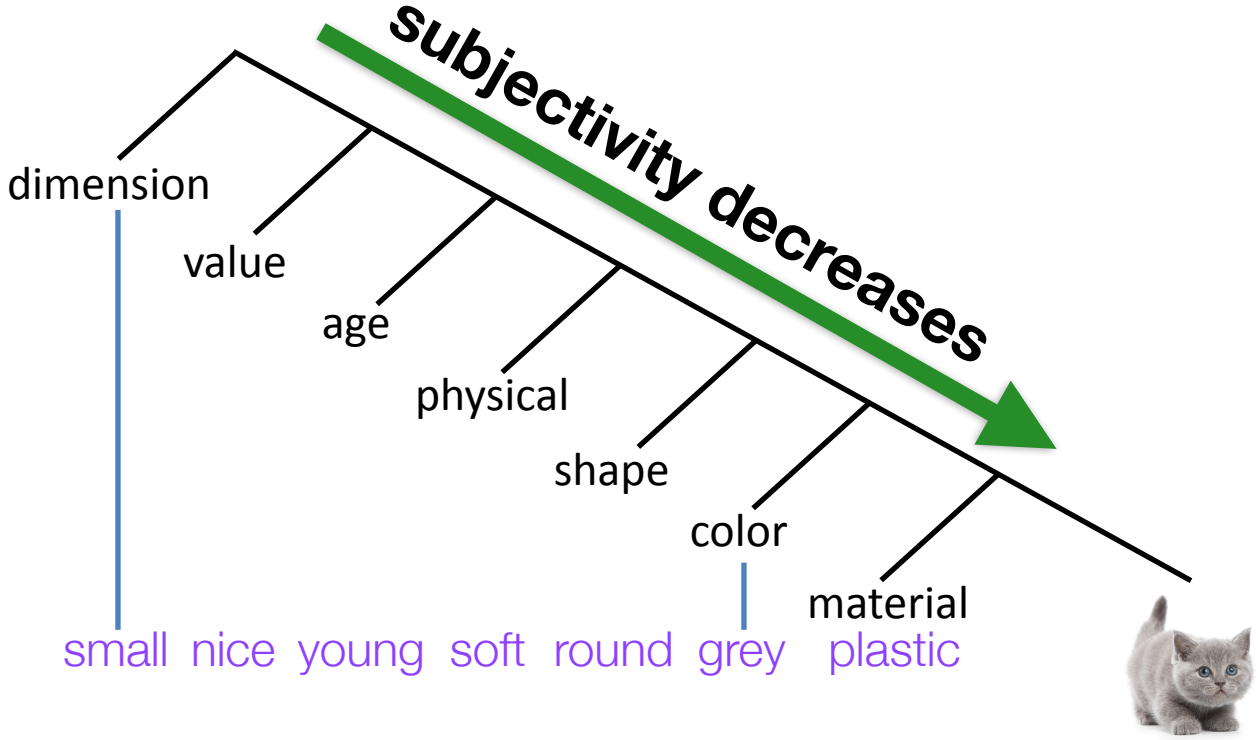
we can also just ask people how **“subjective”** an adjective is:



Scontras et al. 2017

how do adults represent ordering preferences?

lexical class ordering might derive from the perceived **subjectivity** of adjectives

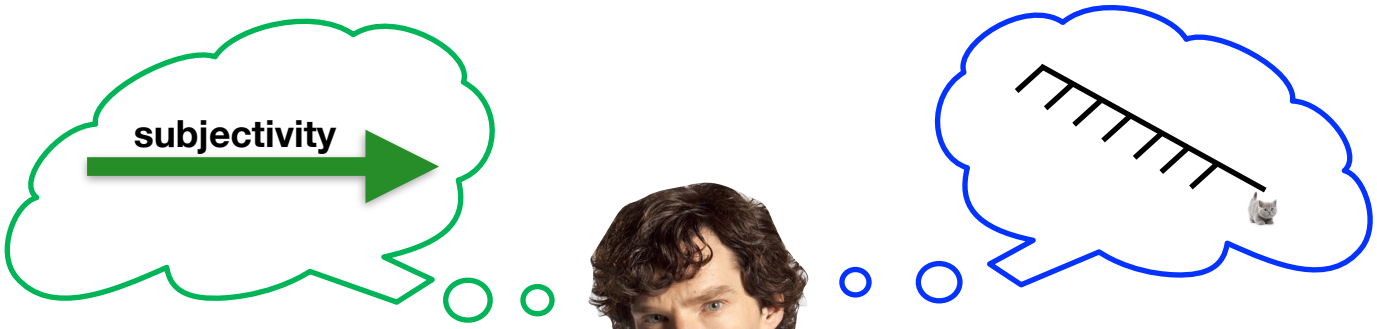


Scontras et al. 2017

how do adults represent ordering preferences?

lexical class ordering might derive from the perceived **subjectivity** of adjectives

two options for adult representations:



ordering with respect to **subjectivity**

ordering with respect to lexical **semantic classes**

what about kids?

when do children develop abstract knowledge of ordering preferences?

when do children develop abstract knowledge of ordering preferences?

we think this knowledge does develop, because the preferences aren't there to begin with, and children become more adult-like as they get older

Bever 1970, Martin and Molfese 1972, Hare and Otto 1978

when do children develop abstract knowledge of ordering preferences?

what underlying representation do children have at different ages and **how can we tell**?

when do children develop abstract knowledge of ordering preferences?

a likely starting point:
repeat what they hear in their input

when do children develop abstract knowledge of ordering preferences?

a likely starting point:
input frequency determines output

when do children develop abstract knowledge of ordering preferences?

later, children may begin to organize their knowledge according to **lexical classes**

when do children develop abstract knowledge of ordering preferences?

eventually, children may recognize **subjectivity** as a stable predictor of preferences

a developmental puzzle

how we can tell what the underlying representation could be?

first, we need a really good sample of **what children are saying** at different ages and **what they are hearing**


small grey
nice grey
small white
big grey
nice small


small grey
small fluffy
nice small

corpus analysis

data:

English data on the CHILDES database, North American and United Kingdom corpora

utterances:

1,069,406 child-produced utterances

688,428 child-directed utterances

ages:

2 to 4 years of age

method

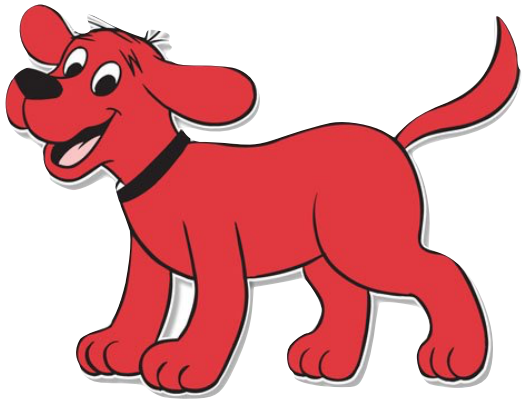
1. extract [adjective adjective noun] phrases from corpora

2. calculate mean distance of each adjective from the noun

3. assign adjectives to a lexical class and associate them with subjectivity scores

child-directed utterances

*MOT: my dog is a big red dog
%mor: … (1)adj|big (1)adj|red (1)n|dog



child-produced utterances

*CHI: nice fresh air
%mor: (1)adj|nice (1)adj|fresh (1)n|air

adjective instances

age; produced/ directed	#multi-adjective strings	#adj tokens	#adj types
2;			
p:	466	932	79
d:	1440	2880	131
3;			
p:	274	584	72
d:	881	1762	128
4;			
p:	235	470	81
d:	745	1490	124

repetitions

were children just parroting adults?

repetitions

were children just parroting adults?

2 years old:
3.79% repetitions
0.57% child repeating adult

3 years old:
2.8% repetitions
0.33% child repeating adult

4 years old:
1.92% repetitions
0.50% child repeating adult

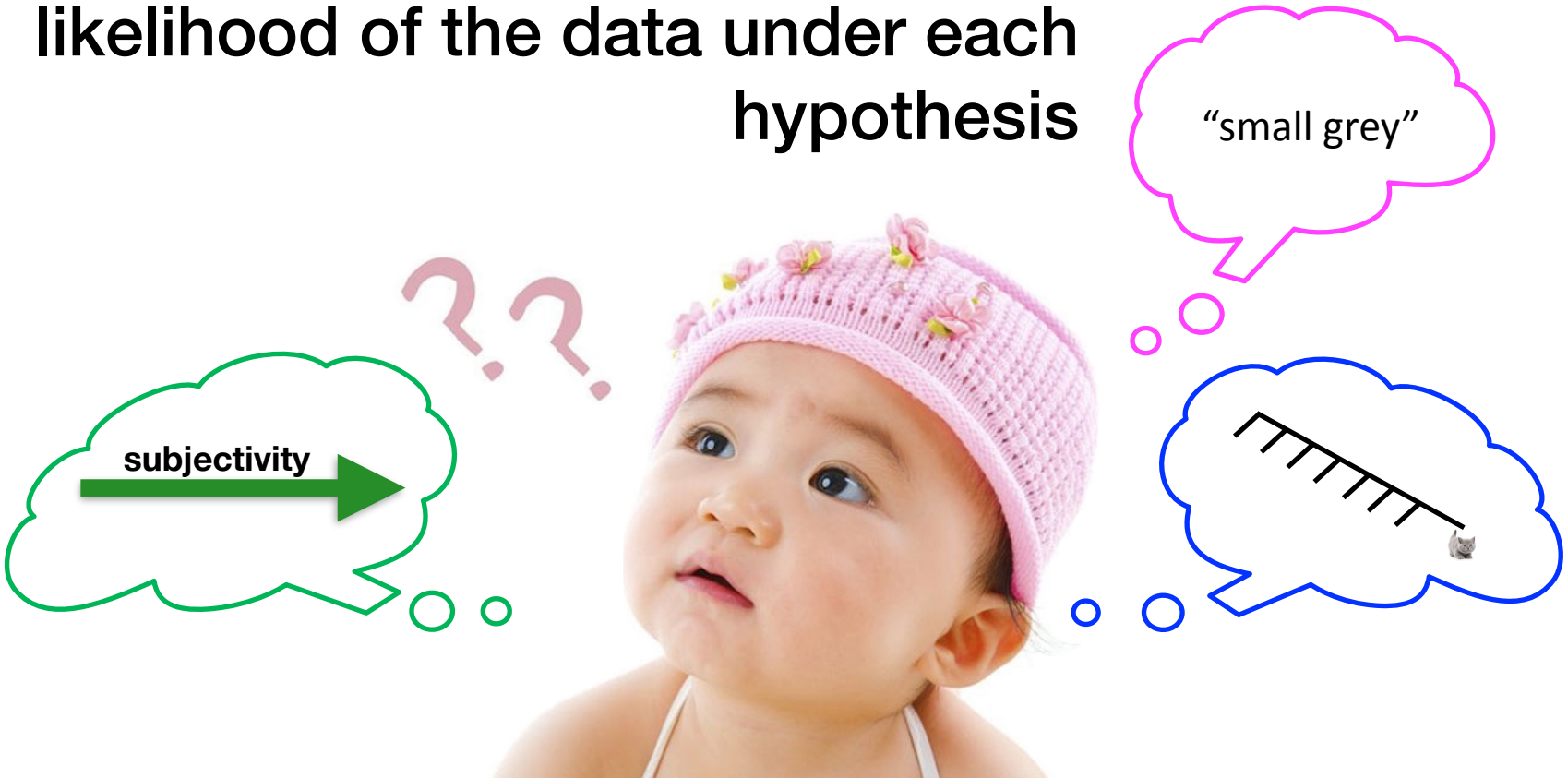
repetitions

were children just parroting adults?

2-4 years old
3.46% repetitions
0.50% child repeating adult

hypothesis comparison

we can evaluate **how well a hypothesis predicts our data** by calculating and comparing **the likelihood of the data under each hypothesis**



hypothesis comparison

calculate the probability that a given adjective in the input will appear "**2-away**" in a new multi-adjective string under each hypothesis

"small grey kitten"

(2-away) (1-away)

hypothesis comparison: input frequency

$H_{InputFreq}$: small

$$p_2 exp(\text{small}) = \frac{f_{2input}(\text{small})}{N_{input}(\text{small})}$$

depends on how often it was in your input in each position

hypothesis comparison: input frequency

expectation that small occurs 2-away again

$H_{InputFreq}$: small

$$p_2exp(\text{small}) = \frac{f_{2input}(\text{small})}{N_{input}(\text{small})}$$

small grey

nice grey

small white

big grey

nice small

hypothesis comparison: input frequency

$$H_{InputFreq}: \text{small}$$

# of times small appears 2-away in input

$$p_2exp(\text{small}) = \frac{f_{2input}(\text{small})}{N_{input}(\text{small})}$$

small grey

nice grey

small white

big grey

nice small

small grey kitten

hypothesis comparison: input frequency

$H_{InputFreq}$: small

# of multi-adjective strings containing small in input

$$p_2exp(\text{small}) = \frac{f_{2input}(\text{small})}{N_{input}(\text{small})}$$

small grey

nice grey

small white

big grey

nice small

hypothesis comparison: lexical class

$$H_{SemCl}: \text{small}$$

$$p_2exp(\text{small}) = \frac{f_{input}(<\text{small}) + 0.5 * f_{input}(=\text{small})}{N_{input}(adj)}$$

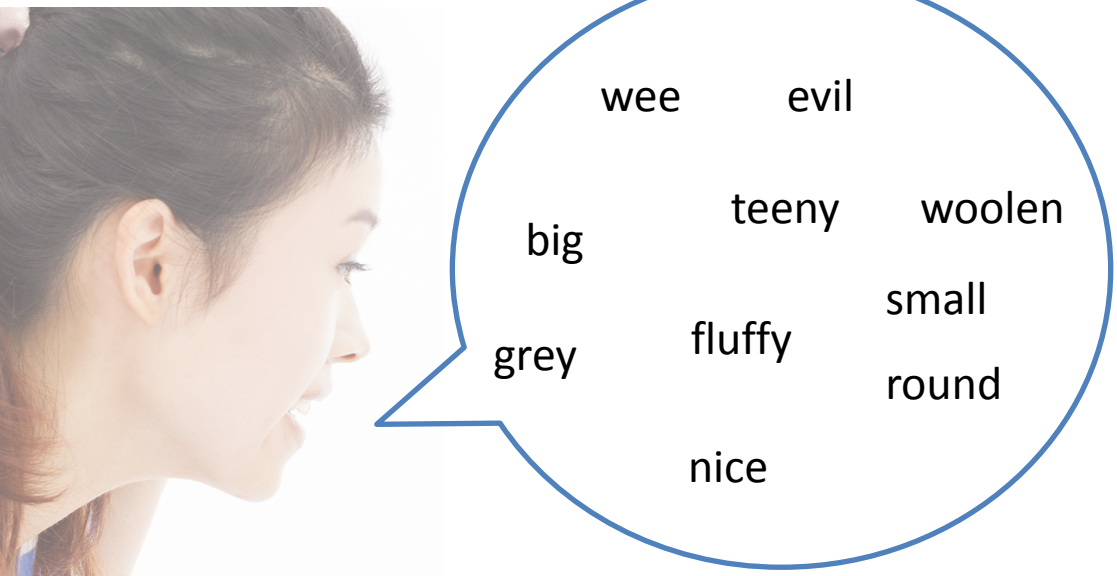
what is the probability that small will appear 2-away with another adjective?

hypothesis comparison: lexical class

expectation that small occurs 2-away again

$$H_{SemCl}: \text{small}$$

$$p_2exp(\text{small}) = \frac{f_{input}(<\text{small}) + 0.5 * f_{input}(=\text{small})}{N_{input}(adj)}$$

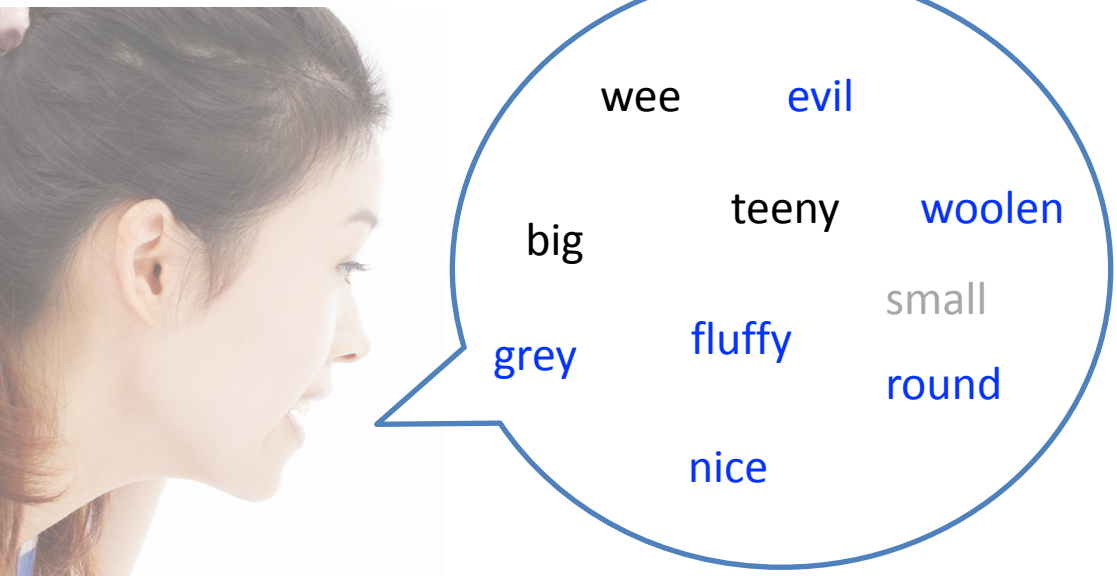


hypothesis comparison: lexical class

adjective tokens in a closer lexical class than small

$$H_{SemCl}: \text{small}$$

$$p_2 exp(\text{small}) = \frac{f_{input}(<\text{small}) + 0.5 * f_{input}(=\text{small})}{N_{input}(adj)}$$



hypothesis comparison: lexical class

$$H_{SemCl:\,\text{small}}$$

adjective tokens in the same semantic class as small × 0.5

$$p_2exp(\text{small}) = \frac{f_{input}(<\text{small}) + 0.5 * f_{input}(=\text{small})}{N_{input}(adj)}$$

hypothesis comparison: lexical class

$$H_{SemCl}: \text{small}$$

$$p_2exp(\text{small}) = \frac{f_{input}(<\text{small}) + 0.5 * f_{input}(=\text{small})}{N_{input}(adj)}$$

# of total adjective tokens in input

hypothesis comparison: subjectivity

$$H_{Subj}: \text{small}$$

$$p_2exp(\text{small}) = \frac{f_{input}(<\text{small}) + 0.5 * f_{input}(=\text{small})}{N_{input}(adj)}$$

what is the probability that small will appear 2-away with another adjective?

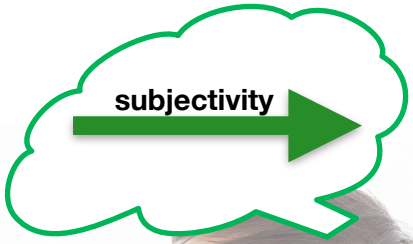
hypothesis comparison: subjectivity

expectation that small occurs 2-away again

H_{Subj}: small

$$p_2exp(\text{small}) = \frac{f_{input}(<\text{small}) + 0.5 * f_{input}(=\text{small})}{N_{input}(adj)}$$

wee 0.56
evil 0.55
woolen 0.11
big 0.9
teeny 0.65
small 0.56
grey 0.28
fluffy 0.23
round 0.33
nice 0.67



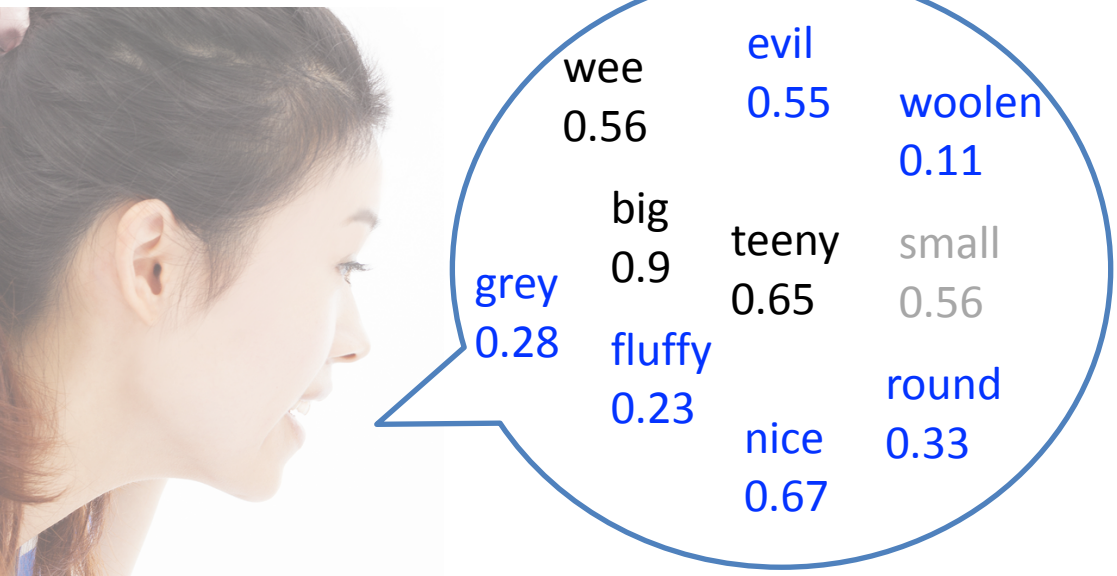
***subjectivity scores come from adult MTurk judgments**

hypothesis comparison: subjectivity

adjective tokens less subjective than small

H_{Subj}: small

$$p_2exp(\text{small}) = \frac{f_{input}(<\text{small}) + 0.5 * f_{input}(=\text{small})}{N_{input}(adj)}$$



hypothesis comparison: subjectivity

H_{Subj}: small

adjective tokens equally as subjective as small × 0.5

$$p_2exp(\text{small}) = \frac{f_{input}(<\text{small}) + 0.5 * f_{input}(=\text{small})}{N_{input}(adj)}$$

wee 0.56
evil 0.55
woolen 0.11
big 0.9
teeny 0.65
small 0.56
grey 0.28
fluffy 0.23
round 0.33
nice 0.67

hypothesis comparison: subjectivity

H_{Subj}: small

# of total adjective tokens in input

$$p_2exp(\text{small}) = \frac{f_{input}(<\text{small}) + 0.5 * f_{input}(=\text{small})}{N_{input}(adj)}$$

hypothesis comparison

how do we get from the representation to output?

use the expected probability of an adjective appearing in a 2-away position (vs. a 1-away position) to calculate **how probable the actual distribution of that adjective is** in the child-produced multi-adjective strings

hypothesis comparison

for each hypothesis, we calculate the **likelihood of the data given the hypothesis for each adjective** in the child's output

$$p(D(adj_x)|H) = \binom{N}{f}(p_2 exp(adj_x))^f(1 - p_2 exp(adj_x))^{N-f}$$

"small"

small grey

small fluffy

nice small

???

hypothesis comparison

total # of multi-adjective strings

$$p(D(adj_x)|H) = \binom{N}{f}(p_2 exp(adj_x))^f(1 - p_2 exp(adj_x))^{N-f}$$

hypothesis comparison

probability of being 2-away

$$p(D(adj_x)|H) = \binom{N}{f} (p_2 exp(adj_x))^f (1 - p_2 exp(adj_x))^{N-f}$$

# of times 2-away

small grey

small fluffy

nice small

???

hypothesis comparison

probability in 1-away position

# of times 1-away

$$p(D(adj_x)|H) = \binom{N}{f}(p_2 exp(adj_x))^f (1 - p_2 exp(adj_x))^{N-f}$$

hypothesis comparison

for all adjectives in the child's production, the likelihood of that hypothesis is:

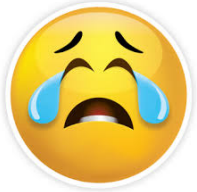
$$p(D|H) = \prod_{adj_x \in A} p(D(adj_x)|H)$$

results

log probabilities

because the probabilities are so small, results are given in logged probabilities

scores range from
0 (best, highly probable) to **-infinity** (worst, not probable)



results

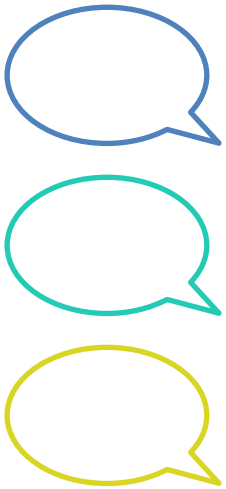
remember: trying to capture different data for each age

age	input frequency	lexical class	subjectivity

results

log probability scores for each hypothesis at 2, 3, and 4 years old

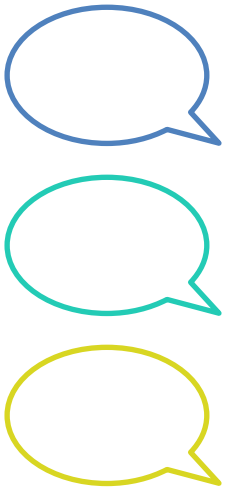
age	input frequency	lexical class	subjectivity
2	**-202.6**	-334.9	-322.4



results

log probability scores for each hypothesis at 2, 3, and 4 years old

age	input frequency	lexical class	subjectivity
2	**-202.6**	-334.9	-322.4
3	**-125.1**	-164.0	-187.4



results

log probability scores for each hypothesis at 2, 3, and 4 years old

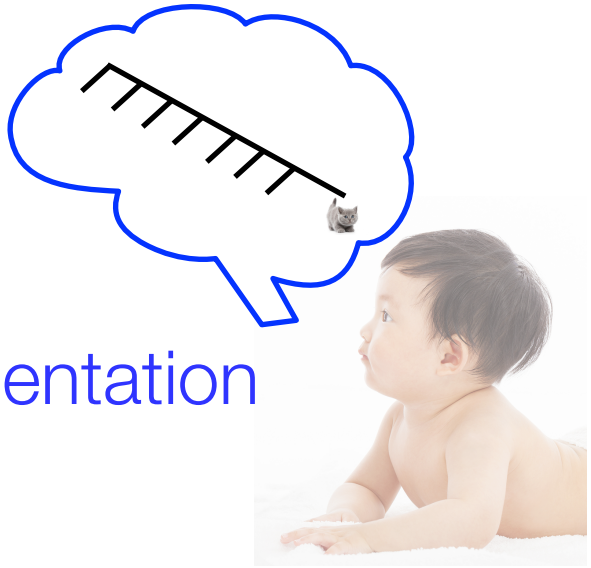
age	input frequency	lexical class	subjectivity
2	**-202.6**	-334.9	-322.4
3	**-125.1**	-164.0	-187.4
4	-182.9	**-165.2**	-211.0

results

simply using the input frequency positions is the best fit for ages 2 and 3

age	input frequency	lexical class	subjectivity
2	**-202.6**	-334.9	-322.4
3	**-125.1**	-164.0	-187.4
4	-182.9	**-165.2**	-211.0

results

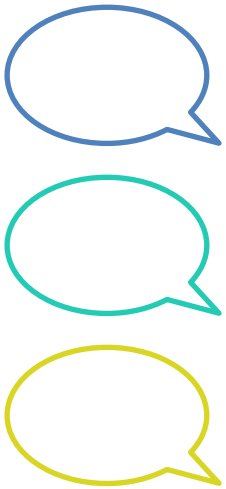


at 4, a lexical class representation is the best fit

age	input frequency	lexical class	subjectivity
2	**-202.6**	-334.9	-322.4
3	**-125.1**	-164.0	-187.4
4	-182.9	**-165.2**	-211.0

results

we can see the emergence of more abstract knowledge



age	input frequency	lexical class	subjectivity
2	**-202.6**	-334.9	-322.4
3	**-125.1**	-164.0	-187.4
4	-182.9	**-165.2**	-211.0

results

let's look at how close the lexical class hypothesis is to the input frequency hypothesis in terms of data coverage

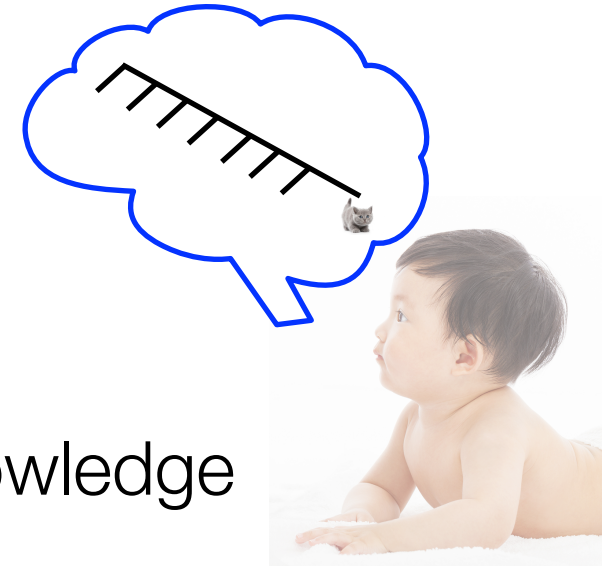
age	input frequency	lexical class	subjectivity
2	**-202.6**	-334.9	-322.4
3	**-125.1**	-164.0	-187.4
4	-182.9	**-165.2**	-211.0

results

take the ratio between log probabilities:
the gap narrows as children get older

age	input frequency	lexical class	subjectivity
2	**-202.6**	-334.9	-322.4
		1.65	
3	**-125.1**	-164.0	-187.4
		1.31	
4	-182.9	**-165.2**	-211.0
		0.90	

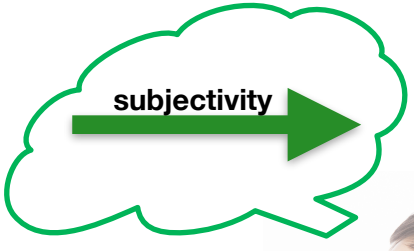
results



interpretation:
emergence of lexical class knowledge

age	input frequency	lexical class	subjectivity
2	**-202.6**	-334.9	-322.4
		1.65	
3	**-125.1**	-164.0	-187.4
		1.31	
4	-182.9	**-165.2**	-211.0
		0.90	

results



the same is true for subjectivity:
the gap narrows over time

age	input frequency	lexical class	subjectivity
2	**-202.6**	-334.9	-322.4
			1.59
3	**-125.1**	-164.0	-187.4
			1.49
4	-182.9	**-165.2**	-211.0
			1.34

when do children develop abstract knowledge of ordering preferences?

a starting point:
input frequency determines output

when do children develop abstract knowledge of ordering preferences?

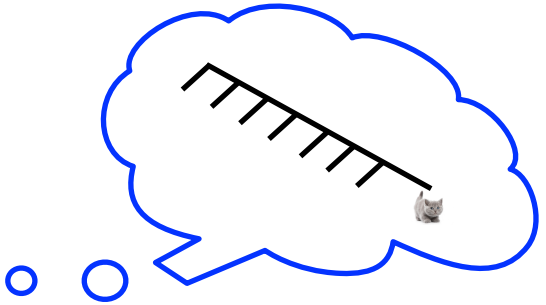
later, around age 4:
children begin to organize their knowledge according to
lexical classes

when do children develop abstract knowledge of ordering preferences?

eventually, children may recognize **subjectivity** as a stable predictor of preferences

take-home points

by using corpus analysis and quantitative approaches, we can see when more abstract underlying representations emerge for adjective ordering preferences (~4)



still unclear when (or whether) subjectivity overtakes lexical class — may depend on children's development of the conceptual underpinnings of subjectivity

thank you!

Computation of Language Laboratory

UC Irvine

age	input frequency	lexical class	subjectivity	binned-sub
2	**-202.6**	-334.9	-322.4	-274.6
3	**-125.1**	-164.0	-187.4	-163.0
4	-182.9	**-165.2**	-211.0	-193.5

Progress:

Consider the following situation:

Greg and Logan see the same carrot.

Greg says: **"That carrot is big."**

Logan responds: **"You're wrong. That carrot is not big."**

Can both Greg and Logan be right?

No, somebody must be wrong. Yes, it's a matter of opinion.

Continue

Progress:

Consider the following adjective:

large

How subjective is the adjective "large"?

completely objective completely subjective

Continue