Small lexical learners:
The development of adjective ordering preferences

Bar-Sever, Lee, Scontras, & Pearl
presented by: Galia Bar-Sever
BUCLD 42, 2017
“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
not only in English,
but in many different languages
where adjectives occur either
pre- or post-nominally
kitten grey small

Hungarian
Mokilese
Dutch
Selepet
Telugu
Mandarin Chinese
how do adults represent ordering preferences?

simple hypothesis: repeat back what you hear

small grey
how do adults represent ordering preferences?

simple hypothesis: repeat back what you hear

small grey
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

small grey!

small grey

small grey
how do adults represent ordering preferences?

adjectives group into lexical semantic classes that are ordered
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
how do adults represent ordering preferences?

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

dimension
  value
  age
  physical
  shape
  color
  material

Dixon 1982, Cinque 2014
how do adults represent ordering preferences?

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

dimension
  - value
    - age
    - physical
      - shape
        - color
          - material
  - small
    - grey

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?

dimension
  value
    age
      physical
        shape
          color
            material
              small nice young soft round grey plastic
how do adults represent ordering preferences?

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

the faultless disagreement task
operationalizing subjectivity

the faultless disagreement task

Scontras et al. 2017
operationalizing subjectivity

the faultless disagreement task
That kitten is small!

operationalizing subjectivity

the faultless disagreement task

Scontras et al. 2017
You're wrong!
That kitten is not small!!

That kitten is small!

operationalizing subjectivity
the faultless disagreement task

Scontras et al. 2017
Can they both be right?

That kitten is not small!!
You’re wrong! That kitten is not small!!

nope  Can they both be right?  yep

Scontras et al. 2017
you might be more able to faultlessly disagree on whether something is “small” than you would on whether it is “grey”

“small grey kitten”
operationalizing subjectivity

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

How subjective is “small”

not subjective  very subjective

Scontras et al. 2017
how do adults represent ordering preferences?

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

“small grey”

subjectivity
when do children develop abstract knowledge of ordering preferences?

a likely starting point: 
**input frequency** determines output

“small grey”
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
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

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

“small grey”

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

“small”

(2-away)

“grey”

(1-away)

“kitten”
hypothesis comparison: input frequency

\[ H_{InputFreq} : \text{small} \]

\[ p_{2exp}(\text{small}) = \frac{f_{2input}(\text{small})}{N_{input}(\text{small})} \]

depends on how often it was in your input in each position.
hypothesis comparison: input frequency

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

expectation that small occurs 2-away again

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

small grey

nice grey

small white

big grey

nice small

small grey kitten
Hypothesis comparison: input frequency

\[ H_{InputFreq}: \text{small} \]

\[ p_{2exp}(\text{small}) = \frac{f_{2input}(\text{small})}{N_{input}(\text{small})} \]

- small grey
- nice grey
- small white
- big grey
- nice small

# of times small appears 2-away in input
hypothesis comparison: input frequency

\[ H_{InputFreq}: \text{small} \]

\[ p_{2exp}(\text{small}) = \frac{f_{2input}(\text{small})}{N_{input}(\text{small})} \]

# of multi-adjective strings containing small in input

\(\text{small grey}
\)
\(\text{nice grey}
\)
\(\text{small white}
\)
\(\text{big grey}
\)
\(\text{nice small}
\)

\(\text{small grey kitten}\)
hypothesis comparison: lexical class

\[ H_{SemCl}: \text{small} \]

\[
p_{2\text{exp}}(\text{small}) = \frac{f_{\text{input}}(<\text{small}) + 0.5 \times f_{\text{input}}(=\text{small})}{N_{\text{input}}(adj)}
\]

what is the probability that small will appear 2-away with another adjective?
hypothesis comparison: lexical class

expectation that small occurs 2-away again

\[ p_{2exp}(\text{small}) = \frac{f_{input}(\text{<small}) + 0.5 \times f_{input}(\text{=small})}{N_{input}(\text{adj})} \]

\( H_{SemCl}: \text{small} \)
hypothesis comparison: lexical class

# adjective tokens in a closer lexical class than small

\[ p_{2exp}(\text{small}) = \frac{f_{\text{input}}(<\text{small}) + 0.5 \times f_{\text{input}}(=\text{small})}{N_{\text{input}}(\text{adj})} \]

\[ H_{\text{SemCl}}: \text{small} \]
hypothesis comparison: lexical class

\[ H_{SemCl}: \text{small} \]

\[ p_{2exp}(\text{small}) = \frac{f_{input}(<\text{small}) + 0.5 \times f_{input}(=\text{small})}{N_{input}(adj)} \]

# adjective tokens in the same semantic class as small \times 0.5

wee, evil, big, teeny, woolen, grey, fluffy, small, round, nice
hypothesis comparison: lexical class

\[ H_{SemCl}: \text{small} \]

\[ p_{2exp}(\text{small}) = \frac{f_{\text{input}}(\leq \text{small}) + 0.5 \times f_{\text{input}}(= \text{small})}{N_{\text{input}}(adj)} \]

- number of total adjective tokens in input
hypothesis comparison: subjectivity

\[ H_{Subj}: \text{small} \]

\[ p_{2exp}(\text{small}) = \frac{f_{input}(\text{<small}) + 0.5 \times f_{input}(\text{=small})}{N_{input}(adj)} \]

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

\[ p_{2exp}(\text{small}) = \frac{f_{\text{input}}(<\text{small}) + 0.5 \times f_{\text{input}}(=\text{small})}{N_{\text{input}}(\text{adj})} \]

*subjectivity scores come from adult MTurk judgments
Hypothesis comparison: subjectivity

\[ p_{2exp}(\text{small}) = \frac{f_{input}(\text{small}) + 0.5 \times f_{input}(=\text{small})}{N_{input}(adj)} \]

\# adjective tokens less subjective than small

- wee: 0.56
- grey: 0.28
- fluffy: 0.23
- evil: 0.55
- woolen: 0.11
- teeny: 0.65
- small: 0.56
- nice: 0.67
- round: 0.33
hypothesis comparison: subjectivity

\[ H_{Subj}: \text{small} \]

\[ p_{2exp}(\text{small}) = \frac{f_{input}(<\text{small}) + 0.5 \times f_{input}(=\text{small})}{N_{input}(adj)} \]

<table>
<thead>
<tr>
<th>adjective</th>
<th>probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>wee</td>
<td>0.56</td>
</tr>
<tr>
<td>grey</td>
<td>0.28</td>
</tr>
<tr>
<td>fluffy</td>
<td>0.23</td>
</tr>
<tr>
<td>evil</td>
<td>0.55</td>
</tr>
<tr>
<td>teeny</td>
<td>0.65</td>
</tr>
<tr>
<td>nice</td>
<td>0.67</td>
</tr>
<tr>
<td>woolen</td>
<td>0.11</td>
</tr>
<tr>
<td>big</td>
<td>0.9</td>
</tr>
<tr>
<td>small</td>
<td>0.56</td>
</tr>
<tr>
<td>round</td>
<td>0.33</td>
</tr>
</tbody>
</table>

# adjective tokens equally as subjective as small \times 0.5
hypothesis comparison: subjectivity

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

$p_{2exp}(\text{small}) = \frac{f_{input}(\text{<small}) + 0.5 \times f_{input}(\text{=small})}{N_{input}(\text{adj})}$

# of total adjective tokens in input
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
for each hypothesis, we calculate the likelihood of the data given the hypothesis for each adjective in the child’s output

\[ p(D(adj_x)|H) = \binom{N}{f} (p_2 \exp(adj_x))^f (1 - p_2 \exp(adj_x))^{N-f} \]
hypothesis comparison

\[ p(D(\text{adj}_x) | H) = \binom{N}{f} (p_2 \exp(\text{adj}_x))^f (1 - p_2 \exp(\text{adj}_x))^{N-f} \]

total # of multi-adjective strings

- small grey
- small fluffy
- nice small

???
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

\[ p(D_{adj_x}|H) = \binom{N}{f} (p_2 \exp(adj_x))^f (1 - p_2 \exp(adj_x))^{N-f} \]

small grey
small fluffy
nice small
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

<table>
<thead>
<tr>
<th>age</th>
<th>input frequency</th>
<th>lexical class</th>
<th>subjectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2yrs</td>
<td>-202.6</td>
<td>-334.9</td>
<td>-322.4</td>
</tr>
<tr>
<td>3yrs</td>
<td>-125.1</td>
<td>-164.0</td>
<td>-187.4</td>
</tr>
<tr>
<td>4yrs</td>
<td>-182.9</td>
<td>-165.2</td>
<td>-211.0</td>
</tr>
</tbody>
</table>

2yrs 3yrs 4yrs
results

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

<table>
<thead>
<tr>
<th>age</th>
<th>input frequency</th>
<th>lexical class</th>
<th>subjectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-202.6</td>
<td>-334.9</td>
<td>-322.4</td>
</tr>
</tbody>
</table>
## results

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

<table>
<thead>
<tr>
<th>age</th>
<th>input frequency</th>
<th>lexical class</th>
<th>subjectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-202.6</td>
<td>-334.9</td>
<td>-322.4</td>
</tr>
<tr>
<td>3</td>
<td>-125.1</td>
<td>-164.0</td>
<td>-187.4</td>
</tr>
</tbody>
</table>
results

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

<table>
<thead>
<tr>
<th>age</th>
<th>input frequency</th>
<th>lexical class</th>
<th>subjectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-202.6</td>
<td>-334.9</td>
<td>-322.4</td>
</tr>
<tr>
<td>3</td>
<td>-125.1</td>
<td>-164.0</td>
<td>-187.4</td>
</tr>
<tr>
<td>4</td>
<td>-182.9</td>
<td>-165.2</td>
<td>-211.0</td>
</tr>
</tbody>
</table>
results

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

<table>
<thead>
<tr>
<th>age</th>
<th>input frequency</th>
<th>lexical class</th>
<th>subjectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-202.6</td>
<td>-334.9</td>
<td>-322.4</td>
</tr>
<tr>
<td>3</td>
<td>-125.1</td>
<td>-164.0</td>
<td>-187.4</td>
</tr>
<tr>
<td>4</td>
<td>-182.9</td>
<td>-165.2</td>
<td>-211.0</td>
</tr>
</tbody>
</table>
at 4, a lexical class representation is the best fit

<table>
<thead>
<tr>
<th>age</th>
<th>input frequency</th>
<th>lexical class</th>
<th>subjectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-202.6</td>
<td>-334.9</td>
<td>-322.4</td>
</tr>
<tr>
<td>3</td>
<td>-125.1</td>
<td>-164.0</td>
<td>-187.4</td>
</tr>
<tr>
<td>4</td>
<td>-182.9</td>
<td>-165.2</td>
<td>-211.0</td>
</tr>
</tbody>
</table>
results

we can see the emergence of more abstract knowledge

<table>
<thead>
<tr>
<th>age</th>
<th>input frequency</th>
<th>lexical class</th>
<th>subjectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-202.6</td>
<td>-334.9</td>
<td>-322.4</td>
</tr>
<tr>
<td>3</td>
<td>-125.1</td>
<td>-164.0</td>
<td>-187.4</td>
</tr>
<tr>
<td>4</td>
<td>-182.9</td>
<td>-165.2</td>
<td>-211.0</td>
</tr>
</tbody>
</table>
Let’s look at how close the lexical class hypothesis is to the input frequency hypothesis in terms of data coverage.

<table>
<thead>
<tr>
<th>age</th>
<th>input frequency</th>
<th>lexical class</th>
<th>subjectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-202.6</td>
<td>-334.9</td>
<td>-322.4</td>
</tr>
<tr>
<td>3</td>
<td>-125.1</td>
<td>-164.0</td>
<td>-187.4</td>
</tr>
<tr>
<td>4</td>
<td>-182.9</td>
<td>-165.2</td>
<td>-211.0</td>
</tr>
</tbody>
</table>
results

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

<table>
<thead>
<tr>
<th>age</th>
<th>input frequency</th>
<th>lexical class</th>
<th>subjectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td><strong>-202.6</strong></td>
<td><strong>-334.9</strong></td>
<td><strong>-322.4</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.65</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td><strong>-125.1</strong></td>
<td><strong>-164.0</strong></td>
<td><strong>-187.4</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.31</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td><strong>-182.9</strong></td>
<td><strong>-165.2</strong></td>
<td><strong>-211.0</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.90</td>
<td></td>
</tr>
</tbody>
</table>
## Results

**Interpretation:**
Emergence of lexical class knowledge

<table>
<thead>
<tr>
<th>Age</th>
<th>Input Frequency</th>
<th>Lexical Class</th>
<th>Subjectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-202.6</td>
<td>-334.9</td>
<td>-322.4</td>
</tr>
<tr>
<td>3</td>
<td>-125.1</td>
<td>-164.0</td>
<td>-187.4</td>
</tr>
<tr>
<td>4</td>
<td>-182.9</td>
<td>-165.2</td>
<td>-211.0</td>
</tr>
</tbody>
</table>
results

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

<table>
<thead>
<tr>
<th>age</th>
<th>input frequency</th>
<th>lexical class</th>
<th>subjectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-202.6</td>
<td>-334.9</td>
<td>-322.4</td>
</tr>
<tr>
<td>3</td>
<td>-125.1</td>
<td>-164.0</td>
<td>-187.4</td>
</tr>
<tr>
<td>4</td>
<td>-182.9</td>
<td>-165.2</td>
<td>-211.0</td>
</tr>
</tbody>
</table>
When do children develop abstract knowledge of ordering preferences?

A starting point: **input frequency** determines output

“small grey”
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!