Take-home message for today

1) intake \neq input

because sensitivities of learner may change over development

2) Complex interaction between UG/Hypothesis space and statistical inference mechanisms
Inside the Language Acquisition Device

Examine
1) perceptual encoding
2) role of prior knowledge in statistical inference
3) role of UG in defining acquisitional intake
Noun classes

Noun divided into classes based on external consequences

- verb agreement
- adjective agreement
- pronominalization
- complementizers...
Noun Classes

Subclasses of nouns triggering characteristic agreement patterns

Tsez (Nakh-Dagestanian; 6000 speakers; no prior psycholinguistics):

- class 1: øigu uži (good boy)
- class 2: jigu kid (good girl)
- class 3: bigu k’etu (good cat)
- class 4: rigu čorpa (good soup)
Noun classes

Building an agreement paradigm (Pinker, 1984)

Yet, children struggle with noun classes into the school years (MacWhinney 1978, Karmiloff-Smith 1979, Mills 1986)
acquiring noun classes

<table>
<thead>
<tr>
<th>No Systematicity (Unlearnable)</th>
<th>Partial Systematicity (Learnable)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>Class 2</td>
</tr>
<tr>
<td>mul-ja, mul-du</td>
<td>sif-no, sif-bi</td>
</tr>
<tr>
<td>don-ja, don-du</td>
<td>jav-no, jav-bi</td>
</tr>
<tr>
<td>kap-ja, kap-du</td>
<td>bip-no, bip-bi</td>
</tr>
<tr>
<td>gav-ja, gav-du</td>
<td>dit-no, dit-bi</td>
</tr>
<tr>
<td>mul-a-ja, mul-a-du</td>
<td>sifo-no, sifo-bi</td>
</tr>
<tr>
<td>dona-ja, dona-du</td>
<td>javo-no, javo-bi</td>
</tr>
<tr>
<td>kap-ja, kap-du</td>
<td>bip-no, bip-bi</td>
</tr>
<tr>
<td>gav-ja, gav-du</td>
<td>dit-no, dit-bi</td>
</tr>
</tbody>
</table>

Frigo & McDonald 1998; Gerken et al 2002, 2005
Unlearnable

data[,1]
data[,2]
Learnable
acquiring noun classes

Correlated Cue Hypothesis:

children need partial systematicity among members of a class in order to acquire noun classes

Predictions:

natural languages will contain sufficient quantity of correlated cues

children are sensitive to class commonalities throughout noun class acquisition

Asymmetries between children and adults may reveal how statistical information is used
Noun Classes

Subclasses of nouns triggering characteristic agreement patterns

Tsez (Nakh-Dagestanian; 6000 speakers; no prior psycholinguistics):

- class 1: øigu uži (good boy)
- class 2: jigu kid (good girl)
- class 3: bigu k’etu (good cat)
- class 4: rigu čorpa (good soup)
Noun Classes

Two kinds of distributional features

External: morphosyntactic agreement on V/Adj

• Vowel initial V/A stems only

Internal: shared properties of some members of class

• semantic/phonological features of head N
## Tsez

### Internal distributional characteristics of noun classes

<table>
<thead>
<tr>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3</th>
<th>Class 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>all male humans</td>
<td>all female humans</td>
<td>all other animates</td>
<td>many other things</td>
</tr>
<tr>
<td>only male humans</td>
<td>many other things</td>
<td>many other things</td>
<td></td>
</tr>
<tr>
<td>~13% of nouns</td>
<td>~12% of nouns</td>
<td>~41% of nouns</td>
<td>~34% of nouns</td>
</tr>
</tbody>
</table>
correlated cues in Tsez

what are the correlated cues in tsez?

   Internal cues: predictive information on the noun (semantic/phonological)

   External cues: overt agreement with the nouns with this predictive information

how often do the two appear together?
correlated cues in Tsez

Predictive information on nouns

Plaster et al (2009) found predictive features to classify ~70% of tsez nouns

this is (some) of the information that is predictive of class

naturalistic input would be better input than a dictionary
correlated cues in Tsez

Supervised learning algorithms build decision trees

Corpus of Tsez child directed speech, tagged with relevant features

```
male ?
  yes: Class 1  no: female?
    yes: Class 2  no: animate?
      yes: Class 3  ???
```
correlated cues in Tsez

most predictive features in Tsez

<table>
<thead>
<tr>
<th>Feature type</th>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3</th>
<th>Class 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>privileged semantic</td>
<td>male human</td>
<td>female human</td>
<td>animate</td>
<td></td>
</tr>
<tr>
<td>idiosyncratic</td>
<td>paper clothing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>phonological</td>
<td>γ- initial</td>
<td>b- initial</td>
<td>r- initial</td>
<td>i- final</td>
</tr>
</tbody>
</table>
correlated cues in Tsez

<table>
<thead>
<tr>
<th>Feature</th>
<th>Feature Value</th>
<th>Class Predicted</th>
<th>Class predicts the feature (%)</th>
<th>Feature predicts the class (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>phonological (first segment)</td>
<td>b</td>
<td>Class 3</td>
<td>10</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>r</td>
<td>Class 4</td>
<td>9</td>
<td>61</td>
</tr>
<tr>
<td>phonological (last segment)</td>
<td>i</td>
<td>Class 4</td>
<td>34</td>
<td>54</td>
</tr>
<tr>
<td>biological semantic</td>
<td>animate</td>
<td>Class 3</td>
<td>13</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>female</td>
<td>Class 2</td>
<td>22</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>Class 1</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>other semantic</td>
<td>paper, clothing</td>
<td>Class 2</td>
<td>4</td>
<td>52</td>
</tr>
</tbody>
</table>
correlated cues in Tsez

are nouns with predictive features seen with agreement?

corpus analysis:

84% verb tokens and 77% adjective tokens show overt agreement

% of nouns with predictive features triggering overt agreement:

- class 1: 100%
- class 2: 52%
- class 3: 51%
- class 4: 45%
acquiring noun classes

Correlated Cue Hypothesis:

- children need partial systematicity among members of a class in order to acquire noun classes

Predictions:

- ✔ natural languages will contain sufficient quantity of correlated cues
  - children are sensitive to class commonalities throughout noun class acquisition

Asymmetries between children and adults may reveal how statistical information is used
experiments

to elicit classification of real and nonce words with different combinations of predictive features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Values</th>
<th>Class Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>phonological (first segment)</td>
<td>b</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>r</td>
<td>4</td>
</tr>
<tr>
<td>phonological (last segment)</td>
<td>i</td>
<td>4</td>
</tr>
<tr>
<td>biological semantic</td>
<td>animate</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>female</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>1</td>
</tr>
<tr>
<td>other semantic</td>
<td>paper, clothing</td>
<td>2</td>
</tr>
<tr>
<td>biological semantic &amp; first segment</td>
<td>animate &amp; r-initial</td>
<td>3 &amp; 4</td>
</tr>
<tr>
<td></td>
<td>female &amp; r-initial</td>
<td>2 &amp; 4</td>
</tr>
<tr>
<td></td>
<td>male &amp; b-initial</td>
<td>1 &amp; 3</td>
</tr>
<tr>
<td>biological semantic &amp; last segment</td>
<td>animate &amp; i-final</td>
<td>3 &amp; 4</td>
</tr>
<tr>
<td>actual class &amp; first segment</td>
<td>class 4 &amp; b-initial</td>
<td>4 &amp; 3</td>
</tr>
</tbody>
</table>
Eat/Don’t Eat Task: -iš (eat, intransitive), and -ac’o (eat, transitive) show overt agreement with the subject and object respectively.

Assistant introduces each character and item, participant tells the character to eat, then what to eat/not eat. Class assignment is evident in the agreement.

- kid (girl)  
  Class 2, Semantic Cue

- buq (sun)  
  Class 3, Phonological Cue

- k’uraj (onion)  
  Class 4, no Cue

- zamil (nonce)  
  Class 3, Semantic Cue
Eat/Don’t Eat Task: -iš (eat, intransitive), and -ac’o (eat, transitive) show overt agreement with the subject and object respectively

4-6 yr olds (n=12)

7-9 yr olds (n=12)

adults (n=12)
experiment

Eat/Don’t Eat Task: -iš (eat, intransitive), and -ac’o (eat, transitive) show overt agreement with the subject and object respectively
Results

Measure: % words assigned to target class

target class = class most strongly predicted

Factors: age (young, old, adult)

cue type: no cue

- biological semantic
- other semantic
- phonological
- conflicting

Quantifying the results:

![Graph showing biological semantic cue for young children]

Young Children

Target
Non-target
(male humans, female humans, non human animates)
### Results

Real words:

<table>
<thead>
<tr>
<th></th>
<th>Biological Semantic</th>
<th>Other Semantic</th>
<th>Phonological</th>
<th>No Cue</th>
<th>Conflicting</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Young Children</strong></td>
<td>79</td>
<td>71</td>
<td>84</td>
<td>77</td>
<td>42*</td>
</tr>
<tr>
<td><strong>Older Children</strong></td>
<td>86</td>
<td>58</td>
<td>94</td>
<td>78</td>
<td>47*</td>
</tr>
<tr>
<td><strong>Adults</strong></td>
<td>87</td>
<td>75</td>
<td>92</td>
<td>86</td>
<td>71</td>
</tr>
</tbody>
</table>

Classification of **real words** was compared to the words’ actual class
Results

Nonce words:

<table>
<thead>
<tr>
<th></th>
<th>Biological Semantic</th>
<th>Other Semantic</th>
<th>Phonological</th>
<th>Conflicting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Young Children</td>
<td>54</td>
<td>8*</td>
<td>61</td>
<td>38*</td>
</tr>
<tr>
<td>Older Children</td>
<td>65</td>
<td>9*</td>
<td>63</td>
<td>53</td>
</tr>
<tr>
<td>Adults</td>
<td>53</td>
<td>23*</td>
<td>61</td>
<td>55</td>
</tr>
</tbody>
</table>

Classification of **nonce words** with cues was compared to a base distribution of classification of nonce words without cues.
summary of results

children and adults use predictive information probabilistically

children: phonological > semantic info, despite differences in the predictiveness

children do not take advantage of all possible generalizations

sensitivity out of proportion with statistical reliability
interpretation

Bias in learner’s expectations about cues to classes

some semantic cues ignored by learners despite high predictiveness

Detectability of cue shapes classification more than statistical predictiveness

phonological cues more powerful than semantic cues, despite being less predictive

because phonological information is detectable earlier in development and more reliably throughout development
Perceptual intake ≠ input
- perception delivers imperfect information
- imperfect intake impacts learning
Subset-Superset Issue

• G1 is a subset of G2
Size Principle: Logic via Learner’s Expectation of Data Points

• If only subset data points are encountered, a restriction to the subset becomes more and more likely.

• The more subset data points encountered, the more the learner is biased towards G1.

• Learning rate determined by rate of examples encountered
Size Principle: Logic via simple arithmetic

- If only subset data points are encountered, a restriction to the subset becomes more and more likely.

\[ |G_1| = a \quad |G_2| = a + b \]
\[ d = \frac{1}{a} \quad d = \frac{1}{a+b} \]

\[ \frac{1}{a} > \frac{1}{a+b} \]
Therefore, \( d \) is more likely to have been drawn from \( G_1 \) than \( G_2 \).
Figure 3. Twelve training sets of labeled objects used in Experiment 1, drawn from all three domains (animals, vegetables, and vehicles) and all four test conditions (one example, three subordinate examples, three basic-level examples, and three superordinate examples). The circled number underneath each object is used to index that object’s location in the hierarchical clustering shown in Figure 7.
subset principle: predictions

Bar graph showing the probability of picking hypotheses for 1 Sub. Example and 3 Sub. Examples. The graph plots the probability of picking hypotheses against the number of examples. Data for hypotheses includes Subordinate, Basic, and Superordinate.
finding the prior
size principle: predictions

- Sub. Example
- 3 Sub. Examples

Data

Probability of picking hypothesis

- Subordinate
- Basic
- Superordinate
experimental results
### Descriptions of 30-month old’s vocabulary (mechanical turk)

| Category | Answer  | Av. Count | p(dim|C) |
|----------|---------|-----------|--------|
| noun     | kind    | 308       | 0.91   |
| noun     | property| 3         | 0.012  |
| noun     | both    | 25        | 0.077  |
| adjective| kind    | 2         | 0.045  |
| adjective| property| 59        | 0.911  |
| adjective| both    | 2         | 0.044  |
rated similarity of experimental items
model results
comparing results
The point

Size principle should not be taken to be a replacement for knowledge

Size principle is weaker than prior based on knowledge of grammatical categories
Inside the Language Acquisition Device

statistical inference engine depends on prior knowledge
One-substitution

learnability problem

*I like this yellow bottle and you like that one*

situations compatible with

\[ \text{one} = N^o \]

situations compatible with

\[ \text{one} = N\text{-bar} \]
Poverty of Stimulus Arg.

1) PLD is compatible with $H[\text{one}=N']$, $H[\text{one}=N^o]$

2) It is possible to define data that would distinguish $H[\text{one}=N']$ from $H[\text{one}=N^o]$

3) Learners do not have access to such data

4) But they all acquire $H[\text{one}=N']$

5) Therefore only $H[\text{one}=N']$ is in the learner’s HS
Poverty of Stimulus Arg.

1) PLD is compatible with H[one=N’], H[one=Nº]

2) It is possible to define data that would distinguish H[one=N’] from H[one=Nº]

3) Learners do not have access to such data

4) But they all acquire H[one=N’]

5) Therefore only H[one=N’] is in the learner’s HS
One-substitution

1) is there evidence in the speech to children that leads them to reject the hypothesis that \textit{one} = \textit{Nº}? [step 3]

2) if not, do children at the earliest stages of syntactic development know that \textit{one} = \textit{N’} [step 4]

if 1=No and 2=Yes, then we have an argument from the poverty of the stimulus
Corpus data

Experiment 1: corpus analysis

parental speech from

Adam (2;3 - 4;10) Brown Corpus.

Nina (1;11 - 3;3) in Suppes Corpus.

(approx. 56,000 adult utterances)
Results: CHILDES

1,129 parental uses of *one*

792 pronominal uses

<table>
<thead>
<tr>
<th>antecedent</th>
<th>Det N</th>
<th>Det Adj N</th>
<th>Det N PP</th>
</tr>
</thead>
<tbody>
<tr>
<td>number</td>
<td>750</td>
<td>32</td>
<td>5</td>
</tr>
<tr>
<td>percent</td>
<td>95%</td>
<td>3.5%</td>
<td>0.5%</td>
</tr>
</tbody>
</table>

of 37 cases w/phrasal antecedent: 2 unambig. (0.2%)

# of ungrammatical uses of *one* = 4 (0.5%)

One-substitution: CHILDES

Conclusions:

**data sparseness:** no reliable evidence in the input to tell the child that *one* is anaphoric to N’ and not N°

only learner’s expectation that nothing is anaphoric to X° can lead them to adult grammar
Poverty of Stimulus Arg.

1) PLD is compatible with $H[one=N']$, $H[one=N^o]$

2) It is possible to define data that would distinguish $H[one=N']$ from $H[one=N^o]$

3) Learners do not have access to such data

4) But they all acquire $H[one=N']$

5) Therefore only $H[one=N']$ is in the learner’s HS
One-substitution in infants

Questions:

what do infants know about one-substitution?

is one anaphoric to phrasal categories for infants?

do infants build hierarchical phrase structure in natural language?
One-substitution in infants

Experiment 2: preferential looking

infants’ eye-movements to images on large screen TV monitored

children’s attention is directed by language
Preferential Looking
Preferential Looking

Children’s eye movements coded offline (30 frames/sec)
Preferential Looking

Where's the train?
Preferential Looking

Where's the train?
One-substitution in infants

Experiment 2: preferential looking

18-month-olds familiarized to labeled object

look. here’s a yellow bottle

then tested in 2 conditions:

control vs. anaphoric

what do you see now? do you see another one?
anaphoric condition
Results: *Do you see another one?*

![Graph showing Looking Time for Control and Anaphoric conditions with familiar and novel stimuli, with asterisks indicating statistical significance at p < 0.01.](image)

*Lidz, Waxman and Freedman, Cognition, 2003*