Announcements

Homework 3 will be returned on Tuesday

Homework 4 will be assigned today, and due next Thursday (5/8/08)

Quiz 4 will be on Tuesday (5/6/08)

Grammatical Categorization

Computational Problem: Identify grammatical categories
These will tell you how words are used in the language.

“This is a DAX.”  "He is sibbing.”

DAX = noun  Sib = verb
Categorization: How?
How might children initially learn what categories words belong to?

Deriving Categories from Semantic Information
Semantic Bootstrapping Hypothesis (Pinker 1984)

Children can initially determine a word’s category by observing what kind of entity in the world it refers to.

- objects, substance = noun
- action = verb

Word’s semantic category (meaning) is then linked to innate grammatical category knowledge (noun, verb)

Slight problem: hard to identify the referent in the world for words sometimes (like verbs)

"Look! He’s frepping!"

frep = climb, perch, glower, grab, yell, …?

Another problem: mapping rules are not perfect
Ex: not all action-like words are verbs

"active", "action" action-like meaning, but they’re not verbs
Categorization: How?
How might children initially learn what categories words belong to?

Distributional Learning

Children can initially determine a word’s category by observing the linguistic environments in which words appear:

- relative location of words in an utterance: "He likes to SIB."
- phonological regularities within classes of words: the, a, an = short (monosyllabic) words, simple syllables
- co-occurrence relations between grammatical categories:
  - Determiner Noun (the goblin) = Determiners (a, the, an, …) precede Nouns (goblin)

Categorization: How?
How might children initially learn what categories words belong to?

Distributional Learning (Evidence)

Children are sensitive to the distributional properties of their native language when they’re born (Shi, Werker, & Morgan 1999).

- 7 month olds can recognize and track specific functor words (a, the, to, will…) in fluent speech (Höhle & Weissenborn 2003)
- 15-16 month German infants can determine novel words are nouns, based on the distributional information around the novel words (Höhle et al. 2004)
- 18 month English infants can track distributional information like "is…-ing" to signal that a word is a verb (Santelmann & Jusczyk 1998)

Categorization: How?
How might children initially learn what categories words belong to?

Idea (Gómez & Lakusta 2004)

(1) Sound properties of certain words can be tracked distributionally (monosyllabic, simple syllables = noticeable to infants).
(2) Infants can group words together into categories based on these properties.
About Categorization

Data Observed

<table>
<thead>
<tr>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>X5</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1 = the</td>
<td>king</td>
<td>girl</td>
<td>baby</td>
<td>goblin</td>
</tr>
<tr>
<td>A2 = a</td>
<td>king</td>
<td>girl</td>
<td>baby</td>
<td>goblin</td>
</tr>
<tr>
<td>Y1</td>
<td>Y2</td>
<td>Y3</td>
<td>Y4</td>
<td>Y5</td>
</tr>
<tr>
<td>B1 = will</td>
<td>sing</td>
<td>laugh</td>
<td>steal</td>
<td>run</td>
</tr>
<tr>
<td>B2 = can</td>
<td>sing</td>
<td>laugh</td>
<td>steal</td>
<td>run</td>
</tr>
</tbody>
</table>

*the* goes with these words

*the* behavior = precedes "king", "girl", "baby", etc.
About Categorization

Data Observed

<table>
<thead>
<tr>
<th>A1 = the</th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>X5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>king</td>
<td>girl</td>
<td>baby</td>
<td>goblin</td>
<td>dwarf</td>
</tr>
<tr>
<td>A2 = a</td>
<td>king</td>
<td>girl</td>
<td>baby</td>
<td>goblin</td>
<td>dwarf</td>
</tr>
</tbody>
</table>

B1 = will

<table>
<thead>
<tr>
<th>Y1</th>
<th>Y2</th>
<th>Y3</th>
<th>Y4</th>
<th>Y5</th>
</tr>
</thead>
<tbody>
<tr>
<td>sing</td>
<td>laugh</td>
<td>steal</td>
<td>run</td>
<td>sneeze</td>
</tr>
</tbody>
</table>

B2 = can

<table>
<thead>
<tr>
<th>Y1</th>
<th>Y2</th>
<th>Y3</th>
<th>Y4</th>
<th>Y5</th>
</tr>
</thead>
<tbody>
<tr>
<td>sing</td>
<td>laugh</td>
<td>steal</td>
<td>run</td>
<td>sneeze</td>
</tr>
</tbody>
</table>

Inference: "a" goes with almost all the same words

"a" has almost the same distribution as "the", so "a" is the same category as "the"

Prediction:

"will" goes with these words

"will" behavior = precedes "sing", "laugh", "steal", etc.

Conclusion:

"a dwarf" is in language
### About Categorization

#### Data Observed

<table>
<thead>
<tr>
<th></th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>X5</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>the</td>
<td>king</td>
<td>girl</td>
<td>baby</td>
<td>goblin</td>
</tr>
<tr>
<td>A2</td>
<td>a</td>
<td>king</td>
<td>girl</td>
<td>baby</td>
<td>goblin</td>
</tr>
<tr>
<td>Y1</td>
<td>will</td>
<td>sing</td>
<td>laugh</td>
<td>steal</td>
<td>run</td>
</tr>
<tr>
<td>B1</td>
<td>can</td>
<td>sing</td>
<td>laugh</td>
<td>steal</td>
<td>run</td>
</tr>
</tbody>
</table>

Inference: “can” goes with almost all the same words as “will”.

#### Prediction

“can” acts like “will” so “can” goes with “sneeze”.

#### Conclusion

“can sneeze” is in language.

### Gómez & Lakusta 2004: Categorization Experiment

Testing 12 month olds, using artificial language paradigm (so children couldn’t have any experience with the categories beforehand).

**General procedure:**
- Infants exposed to one of two training languages (L1 or L2).
- Used same set of vocabulary (all novel words).
- L1 generalization: a goes with X, b goes with Y (aX, bY language)
- L2 generalization: a goes with Y, b goes with X (aY, bX language)
### Gómez & Lakusta 2004: Categorization Experiment

<table>
<thead>
<tr>
<th>A1 = alt</th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>X5</th>
<th>X6</th>
<th>Disyllabic words</th>
</tr>
</thead>
<tbody>
<tr>
<td>coomo</td>
<td>fengle</td>
<td>kicey</td>
<td>loga</td>
<td>paylig</td>
<td>wazil</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A2 = ush</th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>X5</th>
<th>X6</th>
<th>Disyllabic words</th>
</tr>
</thead>
<tbody>
<tr>
<td>coomo</td>
<td>fengle</td>
<td>kicey</td>
<td>loga</td>
<td>paylig</td>
<td>wazil</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Y1</th>
<th>Y2</th>
<th>Y3</th>
<th>Y4</th>
<th>Y5</th>
<th>Y6</th>
<th>Monosyllabic words</th>
</tr>
</thead>
<tbody>
<tr>
<td>deech</td>
<td>ghope</td>
<td>jic</td>
<td>skige</td>
<td>vabe</td>
<td>tam</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B1 = ong</th>
<th>Y1</th>
<th>Y2</th>
<th>Y3</th>
<th>Y4</th>
<th>Y5</th>
<th>Y6</th>
<th>Monosyllabic words</th>
</tr>
</thead>
<tbody>
<tr>
<td>deech</td>
<td>ghope</td>
<td>jic</td>
<td>skige</td>
<td>vabe</td>
<td>tam</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B2 = erd</th>
<th>Y1</th>
<th>Y2</th>
<th>Y3</th>
<th>Y4</th>
<th>Y5</th>
<th>Y6</th>
<th>Monosyllabic words</th>
</tr>
</thead>
<tbody>
<tr>
<td>deech</td>
<td>ghope</td>
<td>jic</td>
<td>skige</td>
<td>vabe</td>
<td>tam</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Association:** alt/ush (a1,a2) go with these words (X1-X6)

**Abstraction:** alt/ush (a1,a2) go with disyllabic words

**Categorization:** alt/ush are a category whose behavior is to go with disyllabic words
**Gómez & Lakusta 2004:**

**Categorization Experiment**

<table>
<thead>
<tr>
<th>L1</th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>X5</th>
<th>X6</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1 = alt</td>
<td>alt</td>
<td>como</td>
<td>fengle</td>
<td>kicey</td>
<td>loga</td>
<td>paylig</td>
</tr>
<tr>
<td>A2 = ush</td>
<td>ush</td>
<td>como</td>
<td>fengle</td>
<td>kicey</td>
<td>loga</td>
<td>paylig</td>
</tr>
</tbody>
</table>

**Diarylable words**

<table>
<thead>
<tr>
<th>Y1</th>
<th>Y2</th>
<th>Y3</th>
<th>Y4</th>
<th>Y5</th>
<th>Y6</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1 = ong</td>
<td>deech</td>
<td>ghope</td>
<td>jic</td>
<td>skige</td>
<td>vabe</td>
</tr>
<tr>
<td>B2 = erd</td>
<td>deech</td>
<td>ghope</td>
<td>jic</td>
<td>skige</td>
<td>vabe</td>
</tr>
</tbody>
</table>

**Monosyllabic words**

Association: ong/erd (b1,b2) go with these words (Y1-Y6)

Abstraction: ong/erd (b1,b2) go with monosyllabic words

Categorization: ong/erd are a category whose behavior is to go with monosyllabic words

<table>
<thead>
<tr>
<th>L2</th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>X5</th>
<th>X6</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1 = alt</td>
<td>alt</td>
<td>deech</td>
<td>ghope</td>
<td>jic</td>
<td>skige</td>
<td>vabe</td>
</tr>
<tr>
<td>A2 = ush</td>
<td>ush</td>
<td>deech</td>
<td>ghope</td>
<td>jic</td>
<td>skige</td>
<td>vabe</td>
</tr>
</tbody>
</table>

**Monosyllabic words**

<table>
<thead>
<tr>
<th>Y1</th>
<th>Y2</th>
<th>Y3</th>
<th>Y4</th>
<th>Y5</th>
<th>Y6</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1 = ong</td>
<td>como</td>
<td>fengle</td>
<td>kicey</td>
<td>loga</td>
<td>paylig</td>
</tr>
<tr>
<td>B2 = erd</td>
<td>como</td>
<td>fengle</td>
<td>kicey</td>
<td>loga</td>
<td>paylig</td>
</tr>
</tbody>
</table>

General procedure:

Infants exposed to one of two training languages (L1 or L2).

Used same set of vocabulary (all novel words).

L1 generalization: a goes with X, b goes with Y (aX, bY language)

L2 generalization: a goes with Y, b goes with X (aY, bX language)

Test phase:

Infants exposed to new phrases from their training language

L1 children: new aX, bY examples

L2 children: new aY, bX examples
Gómez & Lakusta 2004: Categorization Experiment

<table>
<thead>
<tr>
<th>L1 test</th>
<th>A1 = alt beevit meeper gackle roosa nawlup binnow</th>
<th>A2 = ush beevit meeper gackle roosa nawlup binnow</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>X2</td>
<td>X3</td>
</tr>
<tr>
<td>Y1</td>
<td>Y2</td>
<td>Y3</td>
</tr>
<tr>
<td>Y4</td>
<td>Y5</td>
<td>Y6</td>
</tr>
</tbody>
</table>

The point: Children needed to complete association, abstraction, and categorization in order to realize that these new instances of aX and bY were part of the artificial language L1.

---

Gómez & Lakusta 2004: Categorization Experiment

<table>
<thead>
<tr>
<th>L1 process</th>
<th>A1 = alt coomo fengle .... wazil</th>
<th>A2 = ush coomo fengle .... wazil</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>X2 ... X6</td>
<td></td>
</tr>
<tr>
<td>Y1</td>
<td>Y2 ... Y6</td>
<td></td>
</tr>
</tbody>
</table>

Association

---

Gómez & Lakusta 2004: Categorization Experiment

<table>
<thead>
<tr>
<th>L1 process</th>
<th>A1 = alt coomo fengle .... wazil</th>
<th>A2 = ush coomo fengle .... wazil</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>X2 ... X6</td>
<td></td>
</tr>
<tr>
<td>Y1</td>
<td>Y2 ... Y6</td>
<td></td>
</tr>
<tr>
<td>B1 = ong</td>
<td>vot pel tood rud biff foge</td>
<td></td>
</tr>
<tr>
<td>B2 = erd</td>
<td>vot pel tood rud biff foge</td>
<td></td>
</tr>
</tbody>
</table>

Association

---
### Gómez & Lakusta 2004: Categorization Experiment

#### L1 process

<table>
<thead>
<tr>
<th></th>
<th>X1</th>
<th>X2</th>
<th>...</th>
<th>X6</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>alt</td>
<td>coomo</td>
<td>fengle</td>
<td>...</td>
</tr>
<tr>
<td>A2</td>
<td>ush</td>
<td>coomo</td>
<td>fengle</td>
<td>...</td>
</tr>
<tr>
<td>B1</td>
<td>ong</td>
<td>deech</td>
<td>ghope</td>
<td>...</td>
</tr>
<tr>
<td>B2</td>
<td>erd</td>
<td>deech</td>
<td>ghope</td>
<td>...</td>
</tr>
</tbody>
</table>

**Abstraction:**
- Disyllabic words:
  - A1 = alt coomo fengle ... wazil
  - A2 = ush coomo fengle ... wazil
  - B1 = ong deech ghope ... tam
  - B2 = erd deech ghope ... tam

### L1 process

<table>
<thead>
<tr>
<th></th>
<th>X1</th>
<th>X2</th>
<th>...</th>
<th>X6</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>alt</td>
<td>coomo</td>
<td>fengle</td>
<td>...</td>
</tr>
<tr>
<td>A2</td>
<td>ush</td>
<td>coomo</td>
<td>fengle</td>
<td>...</td>
</tr>
<tr>
<td>B1</td>
<td>ong</td>
<td>deech</td>
<td>ghope</td>
<td>...</td>
</tr>
<tr>
<td>B2</td>
<td>erd</td>
<td>deech</td>
<td>ghope</td>
<td>...</td>
</tr>
</tbody>
</table>

**Categorization based on similar distribution:**
- Disyllabic words:
  - A1 = alt coomo fengle ... wazil
  - A2 = ush coomo fengle ... wazil
  - B1 = ong deech ghope ... tam
  - B2 = erd deech ghope ... tam

### L1 process

<table>
<thead>
<tr>
<th></th>
<th>X1</th>
<th>X2</th>
<th>...</th>
<th>X6</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>alt</td>
<td>coomo</td>
<td>fengle</td>
<td>...</td>
</tr>
<tr>
<td>A2</td>
<td>ush</td>
<td>coomo</td>
<td>fengle</td>
<td>...</td>
</tr>
<tr>
<td>B1</td>
<td>ong</td>
<td>deech</td>
<td>ghope</td>
<td>...</td>
</tr>
<tr>
<td>B2</td>
<td>erd</td>
<td>deech</td>
<td>ghope</td>
<td>...</td>
</tr>
</tbody>
</table>

**Extension to new examples:**
- alt beevit
- ong pel
Gómez & Lakusta 2004: Categorization Experiment

Results:
12 month olds listened longer to the test items that obeyed the categorizations of the language they were trained on, even though the words in the test items were ones they had never heard before.

This suggests that 12 month olds were able to complete association, abstraction, and categorization for this artificial language - based only on the distributional information available.

Specifically, the distributional information was the occurrence of one item next to another one in the training phase (L1: aX, bY).

Mintz 2003: Digital Children & Categorization

Idea: Children may be attending to other kinds of distributional information available in the linguistic environment

There is evidence that children can track information that is non-adjacent in the speech stream (Santelmann & Jusczyk 1998, Gómez 2002)

he is running

Mintz 2003: Digital Children & Categorization

Idea: What categorization information is available if children track frequent frames?

Frequent frame: X___Y
where X and Y are words that frame another word and appear frequently in the child’s linguistic environment

Examples: the___is can___him
the king is... can trick him...
the goblin is... can help him...
the girl is... can hug him...
Data representing child's linguistic environment:
6 corpora of child-directed speech from the CHILDES database

Definition of "frequent" for frequent frames:
Frames appearing a certain number of times in a given corpus (e.g., 45 times).

Meant to represent the idea that the child will encounter these frames often enough to recognize them and use them for categorization.

Trying out frequent frames on a corpus of child-directed speech.

Frame: the ___ is
"the radio is in the way...but the doll is...and the teddy is..."
radio, doll, teddy = Category 1 (similar to Noun)

Frame: you ___ it
"you draw it so that he can see it...you dropped it on purpose!...so he hit you with it..."
draw, dropped, with = Category 2 (similar to Verb)

Determining success with frequent frames:

Precision = # of words identified correctly as Category within frame
# of words identified as Category within frame

Recall = # of words identified correctly as Category within frame
# of words that should have been identified as Category
Mintz 2003: Digital Children & Categorization

Determining success with frequent frames:

Precision = \# of words identified correctly as Category within frame / \# of words identified as Category within frame

Recall = \# of words identified correctly as Category within frame / \# of words that should have been identified as Category

Frame: you ___ it
draw, dropped, with = Category 2 (similar-ish to Verb)

\# of words correctly identified as Verb = 2
\# of words identified as Verb = 3
Precision = 2/3

Mintz 2003: Digital Children & Categorization

Determining success with frequent frames:

Precision = \# of words identified correctly as Category within frame / \# of words identified as Category within frame

Recall = \# of words identified correctly as Category within frame / \# of words that should have been identified as Category

Frame: you ___ it
draw, dropped, with = Category 2 (similar-ish to Verb)

\# of words correctly identified as Verb = 2
\# of words should be identified as Verb = many (all verbs in corpus)
Recall = 2/many = small number

Mintz 2003: Digital Children & Categorization

Some actual results of frequent frames

Frame: you ___ it
put, want, do, see, take, turn, taking, said, sure, lost, like, leave,
got, find, throw, threw, think, sing, reach, picked, get, dropped,
seen, lose, know, knocked, hold, help, had, gave, found, fit, enjoy,
eat, chose, catch, with, wind, wear, use, took, told, throwing, stick,
share, sang, roll, ride, recognize, reading, ran, pulled, pull, press,
pouring, pick on, need, move, manage, make, load, liked, lift,
licking, let, left, hit, hear, give, flapped, fix, finished, drop, driving,
done, did, cut, crashed, change, calling, bring, break, because, banged
Mintz 2003:  
Digital Children & Categorization

Some actual results of frequent frames
Frame: the ___ is
moon, sun, truck, smoke, kitty, fish, dog, baby, tray, radio,
powder, paper, man, lock, lipstick, lamb, kangaroo, juice, ice,
flower, elbow, egg, door, donkey, doggie, crumb, cord, clip,
chicken, bug, brush, book, blanket, Mommy

Mintz 2003:  
Digital Children & Categorization

Precision & Recall of frequent frames across corpora

Precision: Above 90% for all corpora (high)
Interpretation: When a frequent frame clustered words together
into category, they often did belong together. (Nouns together,
verbs together, etc.)
Recall: Around 10% for all corpora (very low)
Interpretation: A frequent frame made lots of little clusters, rather
than being able to cluster all the verbs together and all the nouns
together.

Mintz 2003:  
Digital Children & Categorization

Getting better recall (forming one category of Verb, Noun, etc.)

Many frames overlap in the words they identify.
the_ is the_ was a_ is that_ is ...
dog dog dog cat cat
cat goblin goblin king king
girl teddy girl teddy

What about putting clusters together that have a certain number
of words in common?
Mintz 2003:
Digital Children & Categorization

Getting better recall (forming one category of Verb, Noun, etc.)
Many frames overlap in the words they identify.
the/a/that_is/was
dog
teddy
cat
goblin
king
girl

Recall goes up to 91% (very high).
Precision stays above 90% (very high)

Mintz 2003:
Digital Children & Categorization

Summary
Frequent frames are non-adjacent co-occurring words with one word in between them.
They are likely to be information young children are able to track, based on experimental evidence.
When tested on realistic child-directed speech, frequent frames do very well at grouping words into clusters which are very similar to actual grammatical categories like Noun and Verb.
Frequent frames could be a very good strategy for children to use.

Questions?