Psych 156A/ Ling 150: Psychology of Language Learning

Lecture 9
Structure

Announcements
Reminder: Office hours for Lisa this week on Thursday from 3:30 - 5:30pm, not today.
Pick up midterms
Homework 2 due today
Review questions posted for phrases
Homework 3 posted (due 2/24/09)

About Language Structure
Sentences are not just strings of words.
The girl danced with the goblin king.

About Language Structure
Sentences are not just strings of words. Words cluster into larger units called phrases, based on their grammatical category.
Noun (N) = girl, goblin, dream, laughter, …
Determiner (Det) = a, the, an, these, …
Adjective (Adj) = lovely, stinky, purple, …
Verb (V) = laugh, dance, see, defeat, …
Adverb (Adv) = lazily, well, rather, …
Preposition (P) = with, on, around, towards, …
Sentences are not just strings of words. Words cluster into larger units called phrases, based on their grammatical category.

Noun Phrases (NP)

Can be replaced with pronouns like "he", "she", "it", ...
About Language Structure

Sentences are not just strings of words. Words cluster into larger units called phrases, based on their grammatical category.

Preposition Phrases (PP)

Det N V P Det Adj N
The girl danced with the Goblin King.

Can be replaced with words like “here” and “there”
Sentences are not just strings of words. Words cluster into larger units called phrases, based on their grammatical category. The girl danced with the Goblin King.

Verb Phrases (VP)

Can be replaced with words like “do so” and “did so”
Computational Problem

How do children figure out which words belong together (as phrases) and which words don’t?

The girl danced with the goblin king.

Learning Phrases

One way we’ve seen that children can learn things is by tracking the statistical information available.

Saffran, Aslin, & Newport (1996): Transitional Probability is something 8-month-olds can track.

\[ \text{Prob(“stlebe”) < Prob(“castle”) \quad \text{Prob(“stlebe”) < Prob(“beyond”)}} \]

to the castle beyond the goblin city

Posit a word boundary at the minimum of the transitional probabilities between syllables.

Learning Phrases

One way we’ve seen that children can learn things is by tracking the statistical information available.

Thompson & Newport (2007): Transitional Probability to divide words into phrases?

<table>
<thead>
<tr>
<th>A B C D E F</th>
</tr>
</thead>
<tbody>
<tr>
<td>The goblin easily steals the child.</td>
</tr>
</tbody>
</table>

A look at real language properties in action with transitional probabilities

Example: Optional phrases

the girl and the dwarf
Example: Optional phrases

\[
\begin{array}{cccccc}
A & B & C & D & E & F \\
\end{array}
\]

The goblin easily steals the child.

If the child only ever sees this order of categories, there’s no way to know how the words break up into phrases using transitional probabilities.

Why?

\[
\begin{align*}
\text{TrProb}(AB) &= \text{TrProb}(BC) = \text{TrProb}(CD) = \\
&= \text{TrProb}(DE) = \text{TrProb}(EF) = 1
\end{align*}
\]

But suppose C is an optional word/phrase.

\[
\begin{array}{cccccc}
A & B & C & D & E & F \\
\end{array}
\]

The goblin steals the child.

Data without C sometimes will appear.

With the optional phrase left out, TrProb(BC) is less than 1 since sometimes B is followed by D instead of always being followed by C. A transitional probability learner later encountering ABCDEF might posit a phrase boundary between B and C because Tr(AB) and TrProb(CD) are still 1.

Conclusion: AB is a unit, CDEF is a unit.

\[
\begin{array}{cccccc}
A & B & C & D & E & F \\
\end{array}
\]

The goblin easily steals the child.

\[
\begin{array}{cccccc}
A & B & C & D & E & F \\
\end{array}
\]

The goblin steals the child.

The goblin (= NP) easily steals the child (= VP)
A look at real language properties in action with transitional probabilities

Example: Optional phrases

A B C D E F
The goblin easily steals the child.

For ABDEF, Tr(AB) and Tr(DE) ≠ 1, but TrProb(BD) < 1. So, a transitional probability learner will posit a boundary between B and D.

Conclusion: AB is a unit, DE is a unit, the goblin (= NP) steals the child (= VP)

The goblin steals the child.

Artificial Language Experiments

Adults (not children) listened to data from an artificial language for 20 minutes on multiple days

Assumption: Adults who are learning an artificial language will behave like children who are learning their first language since the adults have no prior experience with the artificial just as children have no prior experience with their first language

Is this a good assumption to make?

Adults in Artificial Language Experiments = Children in First Language?

Maybe yes, if children’s brains are able to track the same information as adults’ brains. Then, the fact that adults can learn phrases from transitional probabilities means children should also be able to learn phrases from transitional probabilities.

Maybe no, if there are other factors that could interfere, such as adults having more cognitive resources to process information or using their native language experience to help them learn something about the artificial language. Then, just because adults succeed doesn’t mean children will also succeed.

Artificial Language Similar To Real Language?

Properties of the artificial language that are similar to real language properties

optional phrases (the goblin chased a chicken in the castle) PP is optional in the sentence

repeated phrases (NP Verb NP PP) More than one NP is used in the sentence

moved phrases (In the castle the goblin chased a chicken) PP is moved to the front of the sentence
Artificial Language Experiments

Baseline pattern: ABCDEF

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>kof</td>
<td>hox</td>
<td>jes</td>
<td>sot</td>
<td>fal</td>
<td>ker</td>
</tr>
<tr>
<td>Daz</td>
<td>hox</td>
<td>REF</td>
<td>(ref)</td>
<td>2DB</td>
<td>(exe)</td>
</tr>
<tr>
<td>MBR</td>
<td>(br)</td>
<td>LEX</td>
<td>(lex)</td>
<td>TCD</td>
<td>(tcd)</td>
</tr>
</tbody>
</table>

The goblin easily steals the child.

How do we tell if learning happened?

Baseline assessment: Can subjects actually realize all these nonsense words belong to 6 distinct categories?

Can they categorize?

<table>
<thead>
<tr>
<th>kof</th>
<th>hox</th>
<th>jes</th>
<th>sot</th>
<th>fal</th>
<th>ker</th>
</tr>
</thead>
<tbody>
<tr>
<td>daz</td>
<td>neb</td>
<td>tid</td>
<td>zor</td>
<td>rud</td>
<td>sib</td>
</tr>
</tbody>
</table>

How do we tell if learning happened?

Phrase learning assessment: If they can categorize, do they learn what the phrases are (AB, CD, EF)?

Example: test between AB and non-phrase BC

Sample test item - which one do they think belongs together?

| kof | hox | vs. | hox | jes |

Nonse Words Assigned to Each Form Class

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>kof</td>
<td>hox</td>
<td>jes</td>
<td>sot</td>
<td>fal</td>
<td>ker</td>
</tr>
<tr>
<td>Daz</td>
<td>(hox)</td>
<td>REF</td>
<td>(ref)</td>
<td>2DB</td>
<td>(exe)</td>
</tr>
<tr>
<td>MBR</td>
<td>(br)</td>
<td>LEX</td>
<td>(lex)</td>
<td>TCD</td>
<td>(tcd)</td>
</tr>
</tbody>
</table>

Artificial Language Phrases

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td>CD</td>
<td>EF</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Learning a language with optional phrases

Baseline pattern: ABCDEF

Other patterns heard (phrases AB CD EF missing):
CDEF, ABEF, ABCD

kof hox jes sot fal ker
rel zor taf nav
mer neb rud sib
daz lev tid lum

Control subjects:
Control language (remove one adjacent pair at a time)
Additional control patterns heard:
BCDE, ABCF, ADEF

Learning a language with optional phrases

Transitional Probabilities in the Optional Phrase language and the Control language are different. The Optional Phrase language has lower probability across phrase boundaries than within phrases. The control language has the same probability no matter what.

<table>
<thead>
<tr>
<th></th>
<th>A→B</th>
<th>B→C</th>
<th>C→D</th>
<th>D→E</th>
<th>E→F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optional phrases</td>
<td>.90</td>
<td>.90</td>
<td>1.00</td>
<td>.80</td>
<td>1.00</td>
</tr>
<tr>
<td>Optional control</td>
<td>.90</td>
<td>.90</td>
<td>1.00</td>
<td>.90</td>
<td>.90</td>
</tr>
</tbody>
</table>

Learning a language with repeated phrases

Baseline pattern: ABCDEF

Other patterns heard (phrases AB CD EF repeated):
ABCDEFAB, ABCDEFCD, ABCDEFF

kof hox jes sot fal ker
kof hox rel zor taf nav daz neb
mer neb jes zor rud sib tid sot
daz lev tid lum fal nav taf ker

Control subjects:
Control language (repeat one adjacent pair at a time)
Additional control patterns heard:
ABCDEFBC, ABCDEFDE, ABCDEFFA

Learning a language with repeated phrases

Transitional Probabilities in the Repeated Phrase language and the Control language are different. The Repeated Phrase language has lower probability across phrase boundaries than within phrases. The control language has almost the same probability no matter what.

<table>
<thead>
<tr>
<th></th>
<th>A→B</th>
<th>B→C</th>
<th>C→D</th>
<th>D→E</th>
<th>E→F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repeated phrases</td>
<td>.90</td>
<td>.94</td>
<td>1.00</td>
<td>.99</td>
<td>1.00</td>
</tr>
<tr>
<td>Repeated control</td>
<td>.92</td>
<td>.94</td>
<td>1.00</td>
<td>.94</td>
<td>.93</td>
</tr>
</tbody>
</table>
Learning a language with moved phrases

Baseline pattern: ABCDEF

Other patterns heard (phrases AB CD EF moved):
ABCDDEF, ABEFCD, CDAEF, CDEFAB, EFABCD, EFCDAB

Example strings heard:
kof hox jes sot fal ker
daz neb taf nav rel zor
...

Control subjects:
Control language (move one adjacent pair at a time)
Additional control patterns heard:
BCAFDE, AFDEBC, DEAFBC, DEBCAF

Learning a language with moved phrases

Transitional Probabilities in the Moved Phrase language and the Control language are different. The Moved Phrase language has lower probability across phrase boundaries than within phrases. The control language has the same probability no matter what.

<table>
<thead>
<tr>
<th></th>
<th>A→B</th>
<th>B→C</th>
<th>C→D</th>
<th>D→E</th>
<th>E→F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moved</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Moved control</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
</tr>
</tbody>
</table>

Artificial Language Learning:
Categorization, Day 1

Generally above chance performance (50%), control group performing about the same or a little worse than test groups.

Artificial Language Learning:
Categorization, Day 5

General improvement, though test groups still a little better than control groups. Still, subjects generally capable of categorization.
In each case, even after only 20 minutes of exposure (day 1), test subjects are better than control subjects for each of the languages with optional, repeated, or moved phrases.

Artificial Language Learning: Phrases, Day 1

After 5 days of exposure (100 minutes), the difference between control subjects and test subjects becomes apparent.

Artificial Language Learning: Phrases, Day 5

Some properties seem easier to pick up on than others (repeated and moved language subjects are much better than control subjects).

Interestingly, control subjects in the optional phrase condition actually did really well - this is unexpected since the transitional probabilities were uninformative.
Learning a language with optional phrases, repeated phrases, and moved phrases

Baseline pattern: ABCDEF

Other patterns heard (phrases AB CD EF moved):
CDEF, ABEF, ABCD, ABCDEFAB, ABCDEFCD, ABCDEFEE, ABCDEF, ABFCD, CDAEBF, CDEFAB, EFABCD, EFCDAB

<table>
<thead>
<tr>
<th>Pattern</th>
<th>A→B</th>
<th>B→C</th>
<th>C→D</th>
<th>D→E</th>
<th>E→F</th>
</tr>
</thead>
<tbody>
<tr>
<td>All-combined</td>
<td>1.00</td>
<td>0.73</td>
<td>1.00</td>
<td>0.25</td>
<td>1.00</td>
</tr>
<tr>
<td>All-combined</td>
<td>0.67</td>
<td>0.71</td>
<td>0.58</td>
<td>0.59</td>
<td>0.47</td>
</tr>
</tbody>
</table>

Transitional Probabilities in the “All-combined” language and the Control language are different. The “All-combined” language has lower probability across phrase boundaries than within phrases. The control language probabilities are more uniform, though they do vary.

Predictions for all-combined condition?

One idea: Harder

Why? There are many more patterns that are acceptable for the artificial language. Even if transitional probability is informative, it’s a lot of information to track because there are so many patterns that are acceptable and even more potential patterns that are unacceptable.

Prediction: Test subjects don’t do much better than control subjects.

Second idea: The same, or easier.

Why? There are many more patterns that subjects’ minds can catch. If even one of the variations (optional, repeated, moved phrases) is helpful, three of these will be even more helpful. This is reflected in the transitional probabilities, which are much lower across phrases than within phrases.

Prediction: Test subjects do much better than control subjects.

Artificial Language: Categorization

Test subjects do about as well as control subjects for being able to categorize. This is good, since it means subjects can abstract across the artificial words and realize some belong to the same category.

Artificial Language: Phrases

Test subjects much better than control subjects. Second prediction is supported: finding phrases is easier when more variations are available, even though there are more patterns to learn.
Statistically Learning Phrases

Thompson & Newport (2007): Adults can learn phrases in artificial languages if there are “sentences” that show the kinds of variation real sentences can have.

Interesting Point: When there are more variation types (optional, repeated, and moving phrases), adults are even better at unconsciously identifying phrases.

Open Questions:
1. How well will this work for real language data? (Remember Gambell & Yang (2006) found that transitional probabilities don’t work so well for word segmentation when the data is realistic child-directed speech samples.)

2. Will children be able use transitional probabilities to find phrases?