Psych 156A/ Ling 150: Acquisition of Language II

Lecture 17
Structure II

Announcements

Work on structure review questions

Final review this Thursday 6/2/16

Final exam next Tuesday 6/7/16 between 4:00 and 6:00 pm (taken online through EEE).

Consider taking more language science classes in the future!

Language variation: Recap from before

While languages may differ on many levels, they have many similarities at the level of language structure (syntax). Even languages with no shared history seem to share similar structural patterns.

Japanese/Navajo

NP
S
VP
Verb

Edo/English

S
NP
VP
Verb
NP
Object
PP
P
preposition

NP
Object
PP
P
postposition

One way for children to learn the complex structures of their language is to have them already be aware of the ways in which human languages can vary. Linguistic nativists believe this is knowledge contained in Universal Grammar. Then, children listen to their native language data to decide which patterns their native language follows.
Language variation: Recap from before

Languages can be thought to vary structurally on a number of linguistic parameters. One purpose of parameters is to explain how children learn some hard-to-notice structural properties.

- **Head directionality = head first**

Interacting Parameters

Parameter 1: Head-directionality

Edo/English: Head first

Basic word order: Subject Verb Object [SVO]

Prepositions: Preposition Noun Phrase

Possessed before Possessor

Possession Possessor

An issue: Learning parameter values

The observable data are often the result of a combination of interacting parameters. That is, the observable data are the result of some unobservable process, and the child has to reverse engineer the observable data to figure out what parameter values might have produced the observable data - even if the child already knows what the parameters are!

Interacting Parameters

Parameter 1: Head-directionality

Japanese/Navajo: Head-final

Basic word order: Subject Object Verb [SOV]

Postpositions:

Noun Phrase Postposition

Possessor before Possessed

Possessor Possession
Parameter 2: Verb Second
Verb moves to second phrasal position, some other phrase moves to the first position (German)
Sarah das Buch liest
Sarah the book reads

Interacting Parameters

Parameter 2: Verb Second
Verb moves to second phrasal position, some other phrase moves to the first position (German)
Sarah liest Sarah das Buch liest
Sarah reads the book “Sarah reads the book.”

Interacting Parameters

Parameter 2: Verb Second
Verb moves to second phrasal position, some other phrase moves to the first position (German)
Sarah liest Sarah das Buch liest
Sarah reads the book “Sarah reads the book.”

Interacting Parameters

Parameter 2: Verb Second
Verb moves to second phrasal position, some other phrase moves to the first position (German)
Sarah liest Sarah das Buch liest
Sarah reads the book “Sarah reads the book.”

Interacting Parameters

Parameter 2: Verb Second
Verb moves to second phrasal position, some other phrase moves to the first position (German)
Sarah liest Sarah das Buch liest
Sarah reads the book “Sarah reads the book.”
Verb moves to second phrasal position, some other phrase moves to the first position (German)

Sarah liest das Buch

"Sarah reads the book."

Verb does not move (English)

Sarah reads the book.

Observable (spoken) form of the sentence

Which grammars can analyze this data point?

G1
+head-first
+V2

G2
+head-first
+V2

G3
-head-first
+V2

G4
-head-first
+V2
What do the grammars that can analyze this data point have in common?
We don’t know whether it’s +head-first or -head-first since there’s a grammar of each kind (though there are more +head-first).

This data point isn’t unambiguous for any of the parameters we’re interested in because the parameters interact…even though we feel like it might be somewhat informative for +head-first and +V2 because these occur in more grammars that are compatible.

Another example of unambiguous issues

Spanish: +subject-drop

Patterns allowed:

- Vamos
  - go-1st-pl-pres
  - “We go”

- Nosotros vamos
  - 1st-pl-go-1st-pl-pres
  - “We go”
Another example of unambiguous issues

Parameter 1: subject-drop

English: -subject-drop
Patterns allowed:
- go-1st-pl-pres
  “go” - “we go”
- 1st-pl go-pres
  “We go”

Another example of unambiguous issues

Parameter 2: Head-directionality

Edo/English: Head first
Basic word order:
Subject Verb Object [SVO]
Prepositions:
Preposition Noun Phrase
Possessed before Possessor
Possession Possessor

Japanese/Navajo: Head-final
Basic word order:
Subject Object Verb [SOV]
Postpositions:
Noun Phrase Postposition
Possessor before Possessed
Possession Possessor

Data point:
Subject Object Verb

Grammars available:
- +subj-drop
- +head-first
- -subj-drop
- -head-first
Another example of unambiguous issues

Data point: Subject Object Verb

Which grammars can analyze this data point?

+subj-drop +head-first -subj-drop +head-first
G1 G3 G2 G4

Another example of unambiguous issues

Data point: Subject Object Verb

G1? +subj-drop allows Subject to be spoken
+head-first predicts SVO

G2? +subj-drop allows Subject to be spoken
+head-first predicts SVO

Another example of unambiguous issues

Data point: Subject Object Verb

G3? -subj-drop requires Subject to be spoken
+head-first predicts SOV

G4? -subj-drop requires Subject to be spoken
+head-first predicts SOV
Another example of unambiguous issues

Data point: Subject Object Verb

G4? ✔️ -subj-drop requires Subject to be spoken ✔️ -head-first predicts SOV

X G1 +subj-drop ✔️ G2 +subj-drop
+head-first ✔️ -head-first

G3 -subj-drop ✔️ +head-first -head-first

Another example of unambiguous issues

Data point: Subject Object Verb

There’s more than one grammar compatible with this data point... even though we feel like it *should definitely* be informative for -head-first (since that’s the only value in the compatible grammars).

X G1 +subj-drop ✔️ G2 +subj-drop
+head-first ✔️ -head-first

G3 -subj-drop ✔️ +head-first -head-first

So what can we do?

Learning structure with statistical learning: Linguistic parameters and probability
Linguistic knowledge for learning structure

Parameters = constraints on language variation. Only certain rules/patterns are possible. This is linguistic knowledge.

A language’s grammar
= combination of language rules
= combination of parameter values

Idea: use statistical learning to learn which value (for each parameter) that the native language uses for its grammar. This is a combination of using linguistic knowledge & statistical learning.

Yang 2004: Variational learning

Idea taken from evolutionary biology:
In a population, individuals compete against each other. The fittest individuals survive while the others die out.

How do we translate this to learning language structure?

Individual = grammar (combination of parameter values that represents the structural properties of a language)

Fitness = how well a grammar can analyze the data the child encounters

Yang 2004: Variational learning

Idea taken from evolutionary biology:
A child’s mind consists of a population of grammars that are competing to analyze the data in the child’s native language.

Population of grammars
Yang 2004: Variational learning

Intuition: The most successful (fittest) grammar will be the native language grammar because it can analyze all the data the child encounters. This grammar will “win”, once the child encounters enough native language data because none of the other competing grammars can analyze all the data.

If this is the native language grammar, this grammar can analyze all the input while the other two can’t.

Variational learning details

Before the child has encountered any native language data, all grammars are equally likely. So, initially all grammars have the same probability, which is 1 divided the number of grammars available.

<table>
<thead>
<tr>
<th>Grammar</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammar 1</td>
<td>1/3</td>
</tr>
<tr>
<td>Grammar 2</td>
<td>1/3</td>
</tr>
<tr>
<td>Grammar 3</td>
<td>1/3</td>
</tr>
</tbody>
</table>

If there are 3 grammars, the initial probability for any given grammar = 1/3

Variational learning details

At any point in time, a grammar in the population will have a probability associated with it. This represents the child’s belief that this grammar is the correct grammar for the native language.

Variational learning details

As the child encounters data from the native language, some of the grammars will be more fit because they are better able to account for the structural properties in the data.

Other grammars will be less fit because they cannot account for some of the data encountered.

<table>
<thead>
<tr>
<th>Grammar</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammar 1</td>
<td>1/3</td>
</tr>
<tr>
<td>Grammar 2</td>
<td>1/20</td>
</tr>
<tr>
<td>Grammar 3</td>
<td>3/20</td>
</tr>
</tbody>
</table>

Grammars that are more compatible with the native language data will have their probabilities increased while grammars that are less compatible will have their probabilities decreased over time.
After the child has encountered enough data from the native language, the native language grammar should have a probability near 1.0 while the other grammars have a probability near 0.0.

Unambiguous data from the native language can only be analyzed by grammars that use the native language’s parameter value. This makes unambiguous data very influential data for the child to encounter, since these data are only compatible with the parameter value that is correct for the native language.

Problem: Do unambiguous data exist for entire grammars? This requires data that are incompatible with every other possible parameter value of every other possible grammar....

This seems unlikely for real language data because parameters connect with different types of patterns, which may have nothing to do with each other, as we saw from the previous examples of interacting parameters.

Yang (2004)’s algorithm can take advantage of the fact that grammars are really sets of parameter values. Parameter values can be probabilistically accessed, depending on the level of belief (probability) the learner currently has in each one.
For each data point encountered in the input...

(1) Choose a grammar to test out on a particular data point. Select a grammar by choosing a set of parameter values, based on the probabilities associated with each parameter value.

---

Denison, Bonawitz, Gopnik, & Griffiths 2013: Experimental evidence from 4 and 5-year-olds suggests that children are sensitive to the probabilities of complex representations (which parameters are), and so this kind of sampling is not unrealistic.

---

(2) Try to analyze the data point with this grammar.

If this grammar can analyze the data point, increase the probability of all participating parameters values slightly (reward each value).

---

Actual update equation for reward:

\[ p_v_{\text{updated}} = p_v + \gamma(1- p_v) \]

\[ p_o_{\text{updated}} = (1-\gamma)p_o \]

\[ \gamma = \text{learning rate} \ (\text{ex: } \gamma = .125) \]

---
The learning algorithm

(2) Try to analyze the data point with this grammar.

If this grammar can analyze the data point, increase the probability of all participating parameters values slightly (reward each value).

Actual update equation for reward:
If $p_v = .2$ and $p_o = .8$...

$p_v_{\text{updated}} = .2 + .125(1-.2) = .3$
$p_o_{\text{updated}} = (1-.125).8 = .7$

$\gamma = \text{learning rate (ex: } \gamma = .125)$

The learning algorithm

(2) Try to analyze the data point with this grammar.

If this grammar cannot analyze the data point, decrease the probability of all participating parameters values slightly (punish each value).

Actual update equation for reward:
$p_v = \text{previous value of successful parameter value } .2$
$p_o = \text{previous value of opposing parameter value } .8$

$\gamma = \text{learning rate (ex: } .125)$
The learning algorithm

(2) Try to analyze the data point with this grammar.

If this grammar cannot analyze the data point, decrease the probability of all participating parameters values slightly (punish each value).

Actual update equation for reward:

\[ p_{v_{\text{updated}}} = (1-\gamma)p_v \]
\[ p_{o_{\text{updated}}} = \gamma + (1-\gamma)p_o \]

\( \gamma = \text{learning rate (ex: } \gamma = .125) \)

Unambiguous data

Problem ameliorated!
Unambiguous data are much more likely to exist for individual parameter values instead of entire grammars.
Unambiguous issues – not as big a problem!

Data point: Subject Object Verb

In this case, if either G2 or G4 were selected, -head-first would be rewarded (in addition to whichever subj-drop value was used).

Because this data point is unambiguous for -head-first, grammars using that value would be rewarded and its probability as a parameter value would become 1.0 over time.

Meanwhile, grammars using +head-first would be punished every time, and its probability as a parameter value would approach 0.0 over time.
Unambiguous data

Idea from Yang 2004: The more unambiguous data there are, the faster the native language's parameter value will “win” (reach a probability near 1.0). This means that the child will learn the associated structural pattern faster.

Example: the more unambiguous +subject-drop data the child encounters, the faster a child should learn that the native language allows subjects to be dropped.

Question: Is it true that the amount of unambiguous data the child encounters for a particular parameter determines when the child learns that structural property of the language?

Yang 2004, 2011: Unambiguous data learning examples

Wh-fronting for questions:

Wh-word moves to the front (like English)

Sarah will see who?

Underlying form of the question

Who will Sarah see who?

Observable (spoken) form of the question
Wh-fronting for questions

Wh-word moves to the front (like English)

Who will Sarah see who?

Wh-word stays “in place” (like Chinese)

Sarah will see who?

Observable (spoken) form of the question

Parameter: +/- wh-fronting

Native language value (English): +wh-fronting

Unambiguous data: any (normal) wh-question, with wh-word in front (ex: “Who will Sarah see?”)

Frequency of unambiguous data to children: 25% of input

Age of +wh-fronting acquisition: very early (before 1 yr, 8 months)

Topic drop

Chinese (+topic-drop): can drop NP (subject or object) if it is the understood topic of the discourse

Understood topic: Jareth

Speakers had been talking about Jareth

Chinese (+topic-drop): can drop NP (subject or object) if it is the understood topic of the discourse

Understood topic: Jareth

Mingtian gui ji hui xiayu.
Tomorrow estimate will rain
‘It is tomorrow that Jareth believes it will rain’

Speaker doesn’t have to say “Jareth”
Yang 2004, 2011: Unambiguous data learning examples

**Topic drop**

Chinese (+topic-drop): can drop NP (subject or object) if it is the understood topic of the discourse

Understood topic: Jareth

Mingtian gui ji hui xiayu. Tomorrow estimate will rain ‘It is tomorrow that (Jareth) believes it will rain’

English (-topic-drop): can’t drop topic NP

*It is tomorrow that believes it will rain. It is tomorrow that Jareth believes it will rain.

---

Yang 2004, 2011: Unambiguous data learning examples

**Subject drop**

Italian (+subject-drop): can drop the subject

Verrá? 3rd-sg-will-come “Will s/he come?”

English (-subject-drop): can’t drop subject NP

*Will come? Will he come?

---

Yang 2004, 2011: Unambiguous data learning examples

**Topic drop**

Parameter: +/- topic-drop

Native language value (Chinese): +topic-drop

Unambiguous data: any utterance where the object NP is dropped because it is the topic

Frequency of unambiguous data to children: 12% of input

Age of +topic-drop acquisition: very early (before 1 yr, 8 months)

---

Yang 2004, 2011: Unambiguous data learning examples

**Subject drop**

Parameter: +/- subject-drop

Native language value (Italian): +subject-drop

Unambiguous data: Dropped subjects in questions

Frequency of unambiguous data to children: 10% of input

Age of +subject-drop acquisition: very early (before 1 yr, 8 months)
Yang 2004, 2011:
Unambiguous data learning examples

Subject drop
Parameter: +/- subject-drop
Native language value (English): -subject-drop
Unambiguous data: Expletive subjects (ex: It seems he’s going to come after all.)
Frequency of unambiguous data to children: 1.2% of input
Age of -subject-drop acquisition: 3 years old

Verb raising
Verb moves “above” (before) the adverb/negative word (French)
Jean voit souvent voit Marie
Jean often sees Marie
Jean voit pas voit Marie
Jean not sees Marie

Observable (spoken) form of the sentence

Underlying form of the sentence

Yang 2004, 2011:
Unambiguous data learning examples
Verb raising
Verb moves “above” (before) the adverb/negative word (French)
Jean voit souvent voit Marie
Jean often sees Marie
Jean voit pas voit Marie
Jean not sees Marie

Observable (spoken) form of the sentence
Yang 2004, 2011:
Unambiguous data learning examples
Verb raising
Parameter: +/- verb-raising
Native language value (French): +verb-raising
Unambiguous data: data points that have both a verb and an adverb/ negative word in them, where the positions of each can be seen (“Jean voit souvent Marie”)
Frequency of unambiguous data to children: 7% of input
Age of +verb-raising acquisition: 1 yr, 8 months

Yang 2004, 2011:
Unambiguous data learning examples
Verb Second
Parameter: +/- verb-second
Native language value (German): +verb-second
Unambiguous data: Object Verb Subject data points in German (“Das Buch liest Sarah”), since they show the Object and the Verb in front of the Subject
Frequency of unambiguous data to children: 1.2% of input
Age of +verb-second acquisition: ~3 yrs

Yang 2004, 2011:
Unambiguous data learning examples
Verb moves to second phrasal position, some other phrase moves to the first position (German)
Sarah liest das Buch liest
Sarah reads the book “Sarah reads the book.”
Das Buch liest Sarah das Buch liest
The book reads Sarah “Sarah reads the book.”
Verb does not move (English)
Sarah reads the book.
Observable (spoken) form of the sentence

Yang 2004, 2011:
Unambiguous data learning examples
Intermediate wh-words in complex questions
(Hindi, some German) Observable (spoken) form of the question
Wer glaubst du wer Recht hat?
Who think-2nd-sg you who right has
“Who do you think has the right?”
**Yang 2004, 2011:**

Unambiguous data learning examples

Intermediate wh-words in complex questions

(Hindi, some German)

Wer glaubst du wer Recht hat?

_Who think-2nd-sg you who right has_

“Who do you think has the right?”

No intermediate wh-words in complex questions (English)

Who do you think has the right?

*Observable (spoken) form of the question*

---

**Yang 2004, 2011:**

Unambiguous data learning examples

Intermediate wh-words in complex questions

Parameter: +/- intermediate-wh

Native language value (English): -intermediate-wh

Unambiguous data: complex questions of a particular kind that show the absence of a wh-word at the beginning of the embedded clause (“_Who do you think has the right?_”)

Frequency of unambiguous data to children: 0.2% of input

Age of -intermediate-wh acquisition: > 4 yrs

---

**Yang 2004, 2011:**

Unambiguous data learning examples

<table>
<thead>
<tr>
<th>Parameter value</th>
<th>Frequency of unambiguous data</th>
<th>Age of acquisition</th>
</tr>
</thead>
<tbody>
<tr>
<td>+wh-fronting (English)</td>
<td>25%</td>
<td>Before 1 yr, 8 months</td>
</tr>
<tr>
<td>+topic-drop (Chinese)</td>
<td>12%</td>
<td>Before 1 yr, 8 months</td>
</tr>
<tr>
<td>+subject-drop (Italian)</td>
<td>10%</td>
<td>Before 1 yr, 8 months</td>
</tr>
<tr>
<td>+verb-raising (French)</td>
<td>7%</td>
<td>1 yr, 8 months</td>
</tr>
<tr>
<td>+verb-second (German)</td>
<td>1.2%</td>
<td>3 yrs</td>
</tr>
<tr>
<td>-subject-drop (English)</td>
<td>1.2%</td>
<td>3 yrs</td>
</tr>
<tr>
<td>-intermediate-wh (English)</td>
<td>0.2%</td>
<td>&gt; 4 yrs</td>
</tr>
</tbody>
</table>

The quantity of unambiguous data available in the child’s input seems to be a good indicator of when they will acquire the knowledge. The more there is, the sooner they learn the right parameter value for their native language.

---

**Summary: Linguistic structure**

Even with parameters, acquisition of linguistic structure can be hard because a child has to figure out which parameter values produce the observable data. This isn’t always easy because parameters interact.

Variational learning leverages the fact that grammars can be divided into parameters, and a data point can be informative for one parameter but not others.
Big idea: When a parameter is set depends on how frequent the unambiguous data are in the data the child encounters. This can be captured easily with the variational learning, since unambiguous data are very influential: They always reward the native language grammar and always punish grammars with the non-native parameter value.

Predictions of variational learning:
Parameters set early: more unambiguous data available
Parameters set late: less unambiguous data available

These predictions seem to be born out by available data on when children learn certain structural patterns (parameter values) about their native language.

Questions?

You should be able to do all the questions on the structure review questions. Remember to bring questions to the final exam review next class!