Psych 156A/ Ling 150: Acquisition of Language II

Lecture 16

Learning Language Structure

Announcements

Please pick up HW3

Work on structure review questions

For those with 88%+ in the class: Let me know if you will be writing a final paper instead of taking the final exam on June 8.

Final review this Thursday 6/3.

Consider taking more language science classes in the future! (ex: Ling 155/Psych155 this fall (Psychology of Language))

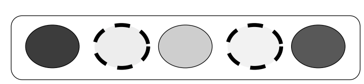
Language Variation: Recap from before

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.

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. 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.

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.

Learning Structure with Statistical Learning: The Relation Between Linguistic Parameters and Probability



Learning Complex Systems Like Language

Only humans seem able to learn human languages
Something in our biology must allow us to do this.

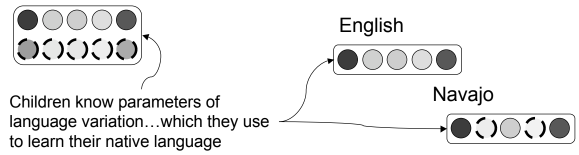
This is what Universal Grammar is: innate biases for learning language that are available to humans because of our biological makeup (specifically, the biology of our brains).

Chomsky

Learning Complex Systems Like Language

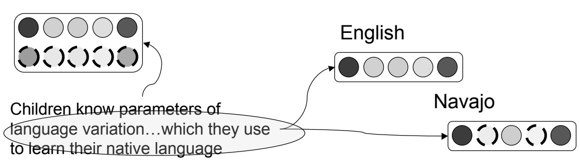
But obviously language is *learned*, so children can't know everything beforehand. How does this fit with the idea of innate biases/knowledge?

Observation: we see constrained variation across languages in their sounds, words, and structure. The knowledge of the ways in which languages vary is children's innate knowledge.



Learning Complex Systems Like Language

The big point: even if children have innate knowledge of language structure, we still need to understand how they learn what the correct structural properties are for their particular language. One idea is to remember that children are good at tracking statistical information (like transitional probabilities) in the language data they hear.



Combining Language-Specific Biases with Statistical Learning

However… remember Gambell & Yang (2006) for statistical learning and word segmentation

"Modeling shows that the statistical learning (Saffran et al. 1996) does not reliably segment words such as those in child-directed English."

Simply using transitional probability between syllables: not so good.

Combining Language-Specific Biases with Probabilistic Learning

But…what happens if statistics are used in conjunction with additional linguistic knowledge?

Gambell & Yang 2006: If statistical learning is constrained by language-specific knowledge (Unique Stress Constraint: words have only one main stress), word segmentation performance increases dramatically.

Combining Language-Specific Biases with Probabilistic Learning

But…what happens if statistics are used in conjunction with additional linguistic knowledge?

Pearl et al. 2010: If children use statistical learning with knowledge about what their lexicons should look like (words should be short, fewer words is better than more words), word segmentation performance also increases dramatically.

Combining Language-Specific Biases with Probabilistic Learning

But…what happens if statistics are used in conjunction with additional linguistic knowledge?

Statistics + linguistic knowledge: much better!

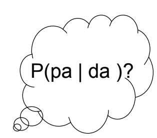
Combining Statistical Learning With Language-Specific Biases

A big deal (Yang 2004):
"Although infants seem to keep track of statistical information, any conclusion drawn from such findings must presuppose that children know *what kind* of statistical information to keep track of."

language-specific information

Ex: Transitional Probability for word segmentation

…of rhyming syllables?
…of individual sounds (b, a, p, d, …)?
…of stressed syllables?

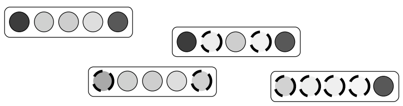


Answer: Track the transitional probability of any syllable sequences.

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.

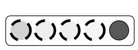
How do we translate this to learning language structure?

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)

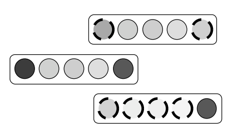


Fitness = how well a grammar can analyze the data the child encounters

Yang (2004): Variational Learning

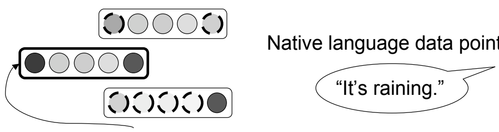
Idea taken from evolutionary biology:
A child's mind consists of a population of grammars that are competing to analyze the data in the child's native language.

Population of Grammars



Yang (2004): Variational Learning

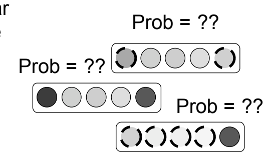
Intuition: The most successful (fittest) grammar will be the native language grammar because it can analyze all the data the child encounters. This grammar will “win”, once the child encounters enough native language data because none of the other competing grammars can analyze all the data.



This grammar can analyze the data point while the other two can’t.

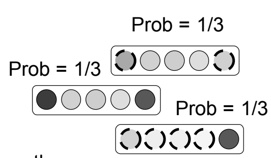
Variational Learning Details

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

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.

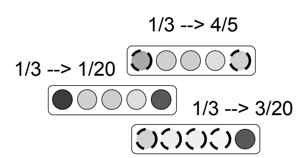


If there are 3 grammars, the initial probability for any given grammar = 1/3

Variational Learning Details

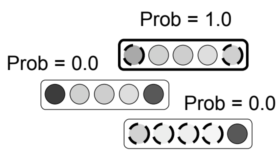
As the child encounters data from the native language, some of the grammars will be more fit because they are better able to account for the structural properties in the data.

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



Variational Learning Details

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.



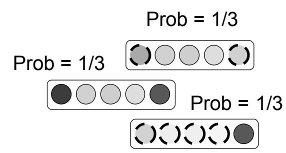
Variational Learning Details

How do we know if a grammar can successfully analyze a data point or not?

Example: Suppose is the subject-drop parameter.

is +subject-drop, which means the language may optionally choose to leave out the subject of the sentence, like in Spanish.

is -subject-drop, which means the language must always have a subject in a sentence, like English.



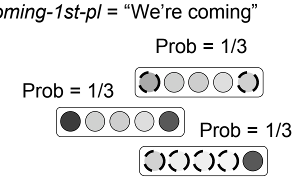
Here, one grammar is +subject-drop while two grammars are -subject-drop.

Variational Learning Details

How do we know if a grammar can successfully analyze a data point or not?

Example data: Vamos = *coming-1st-pl* = “We’re coming”

The +subject-drop grammar is able to analyze this data point as the speaker optionally dropping the subject.



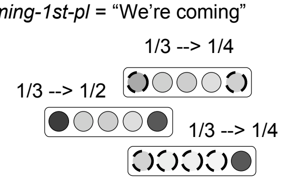
The -subject-drop grammars cannot analyze this data point since they require sentences to have a subject.

Variational Learning Details

How do we know if a grammar can successfully analyze a data point or not?

Example data: Vamos = *coming-1st-pl* = “We’re coming”

The +subject-drop grammar would have its probability increased if it tried to analyze the data point.



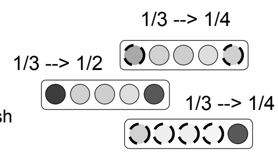
The -subject-drop grammars would have their probabilities decreased if either of them tried to analyze the data point.

Variational Learning Details

Important idea: From the perspective of the subject-drop parameter, certain data will only be compatible with +subject-drop grammars. These data will always reward grammars with +subject-drop and always punish grammars with -subject-drop.

Certain data always reward +subject-drop grammar(s).

Certain data always punish -subject-drop grammar(s).



These are called unambiguous data for the +subject-drop parameter value because they unambiguously indicate which parameter value is correct (here: +subject-drop) for the native language.

The Power of Unambiguous Data

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 it is incompatible with the parameter value that is incorrect for the native language.

Ex: the -subject-drop parameter value is not compatible with sentences that drop the subject. So, these sentences are unambiguous data for the +subject-drop parameter value.

Important to remember: To use the information in these data, the child must know the subject-drop parameter exists.

Unambiguous Data

Idea from Yang (2004): The more unambiguous data there is, 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: 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

Yang 2004: Unambiguous Data Learning Examples

Wh-fronting for questions

Wh-word moves to the front (like English)

Who will Sarah $_{will}$ see $_{who}$?

Observable (spoken) form of the question

Yang 2004: Unambiguous Data Learning Examples

Wh-fronting for questions

Wh-word moves to the front (like English)

Who will Sarah $_{will}$ see $_{who}$?

Wh-word stays "in place" (like Chinese)

Sarah will see who?

Observable (spoken) form of the question

Yang 2004: Unambiguous Data Learning Examples

Wh-fronting for questions

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)

Yang 2004: Unambiguous Data Learning Examples

Verb raising

Verb moves "above" (before) the adverb/negative word (French)

Jean souvent voit Marie
Jean often sees Marie

Jean pas voit Marie
Jean not sees Marie

Underlying form of the sentence

Yang 2004: Unambiguous Data Learning Examples

Verb raising

Verb moves “above” (before) the adverb/negative word (French)
Jean voit souvent voit Marie
Jean sees often Marie “Jean often sees Marie.”

Jean voit pas voit Marie
Jean sees not Marie “Jean doesn’t see Marie.”

Observable (spoken) form of the sentence

Yang 2004: Unambiguous Data Learning Examples

Verb raising

Verb moves “above” (before) the adverb/negative word (French)
Jean voit souvent voit Marie
Jean sees often Marie “Jean often sees Marie.”

Jean voit pas voit Marie
Jean sees not Marie “Jean doesn’t see Marie.”

Verb stays “below” (after) the adverb/negative word (English)
Jean often sees Marie.
Jean does not see Marie.

Observable (spoken) form of the sentence

Yang 2004: 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: Unambiguous Data Learning Examples

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

Underlying form of the sentence

Yang 2004: Unambiguous Data Learning Examples

Verb Second

Verb moves to second phrasal position, some other phrase moves to the first position (German)

Sarah liest $_{\text{Sarah}}$ das Buch $_{\text{liest}}$
Sarah reads the book "Sarah reads the book."

Observable (spoken) form of the sentence

Yang 2004: Unambiguous Data Learning Examples

Verb Second

Verb moves to second phrasal position, some other phrase moves to the first position (German)

Sarah liest $_{\text{Sarah}}$ das Buch $_{\text{liest}}$
Sarah reads the book "Sarah reads the book."

Sarah das Buch liest
Sarah the book reads

Underlying form of the sentence

Yang 2004: Unambiguous Data Learning Examples

Verb Second

Verb moves to second phrasal position, some other phrase moves to the first position (German)

Sarah liest $_{\text{Sarah}}$ das Buch $_{\text{liest}}$
Sarah reads the book "Sarah reads the book."

Das Buch liest Sarah $_{\text{das Buch}}$ $_{\text{liest}}$
The book reads Sarah "Sarah reads the book."

Observable (spoken) form of the sentence

Yang 2004: Unambiguous Data Learning Examples

Verb Second

Verb moves to second phrasal position, some other phrase moves to the first position (German)

Sarah liest $_{\text{Sarah}}$ das Buch $_{\text{liest}}$
Sarah reads the book "Sarah reads the book."

Das Buch liest Sarah $_{\text{das Buch}}$ $_{\text{liest}}$
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: 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: Unambiguous Data Learning Examples

Intermediate wh-words in complex questions

(Hindi, 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: Unambiguous Data Learning Examples

Intermediate wh-words in complex questions

(Hindi, 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: 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: Unambiguous Data Learning Examples

Parameter value	Frequency of unambiguous data	Age of acquisition
+wh-fronting (English)	25%	Before 1 yr, 8 months
+verb-raising (French)	7%	1 yr, 8 months
+verb-second (German)	1.2%	3 yrs
-intermediate-wh (English)	0.2%	> 4 yrs

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: Variational Learning for Language Structure

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 idea, 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?

Bring questions for the final review!