Psych156A/Ling150: Psychology of Language Learning

Lecture 19

Learning Structure with Parameters

Announcements

Next class: Review session for final

- Review homework and quiz questions, come in with questions to go over
- If you want, you may email me which questions you would like to discuss in class. We'll prioritize based on how many people want to discuss any given question.
- Remember: review questions are available for the last 3 lectures ("Structure & Learning Structure"). These are fair game for the final.

HW6: average 33.2 out of 43

Language Variation: Summary

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. Then, they 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 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.

Chomsky: 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).

Learning Complex Systems Like Language

But obviously language is *learned*, not just prespecified beforehand. Children learn their native language, not just any old language.

However, we see constrained variation across languages: sounds, words, structure.

English

Navajo

Learning Complex Systems Like Language

The big point: need both innate biases & probabilistic learning abilities

We need to find a way to explicitly integrate them with each other, so that we can understand how learning language might work. It will likely involve both prior knowledge about language (which may come from the biology of our brains) as well as general-purpose learning strategies like probabilistic/statistical learning.

English

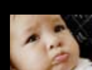
Navajo

Combining Language-Specific Biases with Probabilistic Learning

Statistics for word segmentation (remember Gambell & Yang (2006))

"Modeling shows that the statistical learning (Saffran et al. 1996) does not reliably segment words such as those in child-directed English. Specifically, precision is 41.6%, recall is 23.3%. In other words, about 60% of words postulated by the statistical learner are not English words, and almost 80% of actual English words are not extracted. This is so even under favorable learning conditions".

Unconstrained (simple) statistics: not so good.



Combining Language-Specific Biases with Probabilistic Learning

Statistics for word segmentation (remember Gambell & Yang (2006))

If statistical learning is constrained by language-specific knowledge (Unique Stress Constraint: words have only one main stress), performance increases dramatically: 73.5% precision, 71.2% recall.

Constrained statistics - much better!

Combining Statistical Learning With Language-Specific Biases

A big deal:
“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 bias

Ex: Transitional Probability

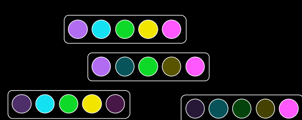
…of rhyming syllables?
…of individual sounds (b, a, p, d, …)?
…of stressed syllables?

No…any syllable sequences.

Constraints for Structure-Learning

Parameters = constraints on language variation. Only certain rules/patterns are possible.

Grammar = combination of language rules.
= combination of parameter values.



So, use statistical learning to learn which value (for each parameter) that the native language uses for its grammar.

Yang (2004): Variational Learning

Idea taken from evolutionary biology:
Individual grammars compete against each other in a child’s mind to see which grammar can best analyze the available data. A grammar’s “fitness” is determined by how well the grammar fares with native language data.

Intuition: Most successful grammar will be the native language grammar. This grammar will “win”, once the child encounters enough native language data.

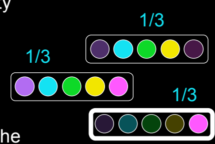
Yang (2004): Variational Learning

Initially, each grammar is equally likely to be the native language grammar.

A grammar will have a probability associated with it, which represents that grammar's likelihood of being the native language grammar.

So, initially, all grammars have the same probability.

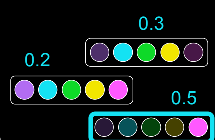
3 grammars, G = 3
Initial probability for any given grammar = 1/G = 1/3



Yang (2004): Variational Learning

After the child has encountered native language data, some grammars will have been more successful while other grammars will have been less successful.

So, the probabilities associated with these grammars will reflect that. The more successful grammars will have a higher probability associated with them.



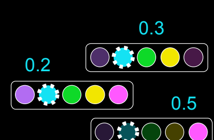
Intuition: Most successful grammar will be the native language grammar. This grammar will have a probability near 1.0 once the child encounters enough native language data.

Grammar Success

How can some grammars be successful while other grammars are not?

Example: Native language data is

Vamos
1st-pl-come
"We're coming"



One parameter may be whether it's okay to leave off or drop the subject (+/- subject-drop).

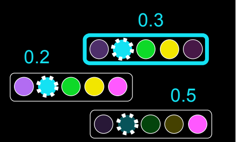
Value 1: Must always have a subject (-subject-drop)
Value 2: May optionally drop the subject (+subject-drop)

Grammar Success

How can some grammars be successful while other grammars are not?

Example: Native language data is

Vamos
1st-pl-come
"We're coming"



Suppose a grammar with the -subject-drop value tried to analyze this data point.

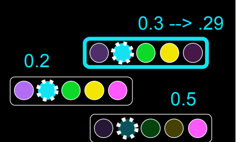
It would not be able to since this sentence does not have an overt subject. So, a -subject-drop grammar is not compatible with this data point. Its probability will go down.

Grammar Success

How can some grammars be successful while other grammars are not?

Example: Native language data is

Vamos
1st-pl-come
"We're coming"



Suppose a grammar with the -subject-drop value tried to analyze this data point.

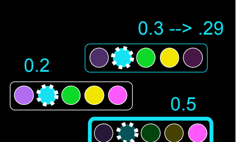
It would not be able to since this sentence does not have an overt subject. So, a -subject-drop grammar is not compatible with this data point. Its probability will go down.

Grammar Success

How can some grammars be successful while other grammars are not?

Example: Native language data is

Vamos
1st-pl-come
"We're coming"



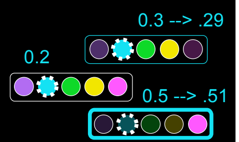
However, suppose a grammar with the +subject-drop value tried to analyze this data point.

It would be able to since it allows sentences to not have an overt subject. So, a +subject-drop grammar is compatible with this data point. Its probability will go up.

Grammar Success

How can some grammars be successful while other grammars are not?

Example: Native language data is



Vamos
1st-pl-come
"We're coming"

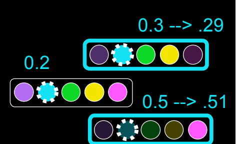
However, suppose a grammar with the +subject-drop value tried to analyze this data point.

It would be able to since it allows sentences to not have an overt subject. So, a +subject-drop grammar is compatible with this data point. Its probability will go up.

Grammar Success

How can some grammars be successful while other grammars are not?

Example: Native language data is



Vamos
1st-pl-come
"We're coming"

Key point: This data is *unambiguous* for the +subject-drop value. Only grammars with the +subject-drop parameter value will be able to successfully analyze this data point.

Unambiguous Data

Unambiguous data from the target language can only be analyzed by grammars that use the target 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 target language.

Ex: the -subject-drop value is not compatible with sentences that drop the subject subject like Vamos
1st-pl-come
"We're coming"

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

Unambiguous Data Learning Examples

Wh-fronting for questions

Wh-word moves to the front (like English)

Sarah will see who?

Unambiguous Data Learning Examples

Wh-fronting for questions

Wh-word moves to the front (like English)

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

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?

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 mos)

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

Unambiguous Data Learning Examples

Verb raising

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

Jean voit souvent $_{\text{voit}}$ Marie
Jean sees often Marie "Jean often sees Marie."

Jean voit pas $_{\text{voit}}$ Marie
Jean sees not Marie "Jean doesn't see Marie."

Unambiguous Data Learning Examples

Verb raising

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

Jean voit souvent $_{\text{voit}}$ Marie
Jean sees often Marie "Jean often sees Marie."

Jean voit pas $_{\text{voit}}$ 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.

Unambiguous Data Learning Examples

Verb raising

Parameter: +/- verb-raising

Native language value (French): +verb-raising

Unambiguous data: verb adverb/negative word data points
("Jean voit souvent Marie")

Frequency of unambiguous data to children: 7% of input

Age of +verb-raising acquisition: 1 yr, 8 months

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

Unambiguous Data Learning Examples

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."

Unambiguous Data Learning Examples

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."

Sarah das Buch liest
Sarah the book reads

Unambiguous Data Learning Examples

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."

Das Buch liest Sarah das Buch liest
The book reads Sarah "Sarah reads the book."

Unambiguous Data Learning Examples

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."

Das Buch liest Sarah das Buch liest
The book reads Sarah "Sarah reads the book."

Verb does not move (English)
Sarah reads the book.

Unambiguous Data Learning Examples

Verb Second

Parameter: +/- verb-second

Native language value (German): +verb-second

Unambiguous data: Object Verb Subject data points ("Das Buch liest Sarah")

Frequency of unambiguous data to children: 1.2% of input

Age of +verb-second acquisition: ~3 yrs

Unambiguous Data Learning Examples

Intermediate wh-words in complex questions ("scope marking")

(Hindi, German)

… wer Recht hat?
…who right has
"…who has the right?"

Unambiguous Data Learning Examples

Intermediate wh-words in complex questions ("scope marking")

(Hindi, German)

Wer	glaubst	du	wer	Recht	hat?
Who	think-2nd-sg	you	who	right	has

"Who do you think has the right?"

Unambiguous Data Learning Examples

Intermediate wh-words in complex questions ("scope marking")

(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 $_{\text{who}}$ has the right?

Unambiguous Data Learning Examples

Intermediate wh-words in complex questions ("scope marking")

(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?

Unambiguous Data Learning Examples

Intermediate wh-words in complex questions ("scope marking")

Parameter: +/- intermediate-wh

Native language value (English): - intermediate-wh

Unambiguous data: complex questions of a particular kind ("Who do you think has the right?")

Frequency of unambiguous data to children: 0.2% of input

Age of -intermediate-wh acquisition: > 4 yrs

Unambiguous Data Examples Summary

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: The time course of when a parameter is set depends on how frequent the necessary evidence is in child-directed speech. This falls out from the probabilistic learning framework, where unambiguous data for the native language parameter value punishes the non-native language value.

Predictions of variational learning:

- Parameters set early: more unambiguous data
- Parameters set late: less unambiguous data

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

Questions?