# Psych 156A/ Ling 150: Acquisition of Language II

Lecture 17
Learning Language Structure

## Announcements

Please pick up HW3

Work on structure review questions

Final review this Thursday 6/7/12

Final exam next Thursday 6/14/12 between 1:30 and 3:30pm (taken online through EEE).

Consider taking more language science classes in the future!

## Language Variation: Recap from before

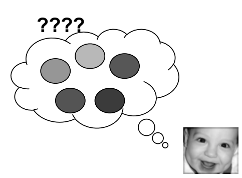
While languages may differ on many levels, they have many similarities at the level of language structure (syntax). Even languages with no shared history seem to share similar structural patterns.

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.

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.

## Issue from last time: 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!




German
Subject Verb *Subject* Object *Verb*

Kannada
Subject *Object* Verb Object

English
Subject Verb Object

## Interacting Parameters

Parameter 1: Head-directionality

Edo/English: Head first

Basic word order:
Subject Verb Object [SVO]

S
NP VP
Subject Verb NP
Object

Prepositions:
Preposition Noun Phrase

Possessed before Possessor
Possession Possessor

PP
P NP
Preposition Object

## Interacting Parameters

Parameter 1: Head-directionality

Japanese/Navajo: Head-final

Basic word order:
Subject Object Verb [SOV]

S
NP VP
Subject NP Verb
Object

Postpositions:
Noun Phrase Postposition

Possessor before Possessed
Possessor Possession

PP
NP P
Object Postposition

## Interacting Parameters

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*

*Underlying form of the sentence*

## 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."

*Observable (spoken) form of the sentence*

## 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."

Sarah das Buch liest
Sarah the book reads

*Underlying form of the sentence*

## 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."

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

*Observable (spoken) form of the sentence*

## 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."

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*

## Interacting Parameters

Data point: Subject Verb Object

Grammars available:

| | |
|---|---|
| **G1** +head-first +V2 | **G2** +head-first -V2 |
| **G3** -head-first +V2 | **G4** -head-first -V2 |

## Interacting Parameters

Data point: Subject Verb Object

**Which grammars can analyze this data point?**

**G1** +head-first +V2

**G2** +head-first -V2

**G3** -head-first +V2

**G4** -head-first -V2

## Interacting Parameters

Data point: Subject Verb *Verb* Object

**G1?** ✔ +head-first predicts SVO
✔ +V2 predicts Verb moved to second position

✔**G1** +head-first +V2

**G2** +head-first -V2

**G3** -head-first +V2

**G4** -head-first -V2

## Interacting Parameters

Data point: Subject Verb Object

**G2?** ✔ +head-first predicts SVO
✔ -V2 predicts Verb in original position

✔**G1** +head-first +V2

✔**G2** +head-first -V2

**G3** -head-first +V2

**G4** -head-first -V2

## Interacting Parameters

Data point: Subject Verb *Subject* Object *Verb*

**G3?** ✔ -head-first predicts SOV
✔ +V2 predicts Verb moved to second position

✔**G1** +head-first +V2

✔**G2** +head-first -V2

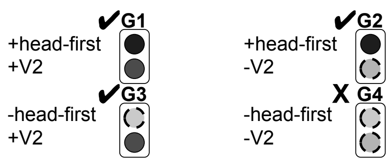
✔**G3** -head-first +V2

**G4** -head-first -V2

## Interacting Parameters

Data point: Subject Verb Object

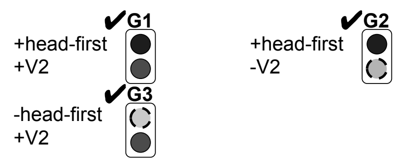
**G4?** X -head-first predicts SOV
✔ -V2 predicts Verb in original position




## Interacting Parameters

Data point: Subject Verb Object

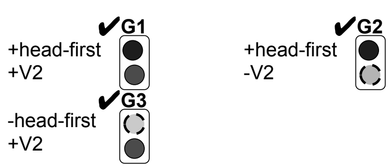
What do the grammars that can analyze this data point have in common?




## Interacting Parameters

Data point: Subject Verb Object

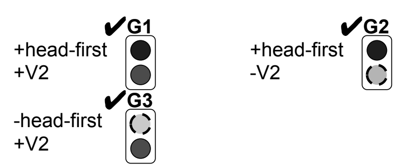
We don't know whether it's +head-first or -head-first since there's a grammar of each kind.




## Interacting Parameters

Data point: Subject Verb Object

We don't know whether it's +V2 or -V2 since there's a grammar of each kind.



## Interacting Parameters

Data point: Subject Verb Object

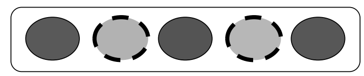
This data point doesn't help us choose the parameter values for either of these parameters.

✔G1
+head-first
+V2

✔G2
+head-first
-V2

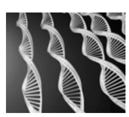
✔G3
-head-first
+V2

## 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.

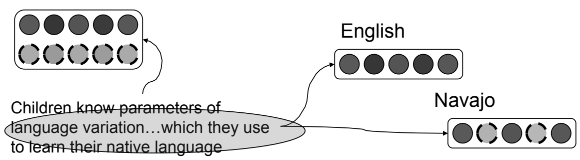
English

Navajo

Children know parameters of language variation…which they use to learn their native language

## Learning Complex Systems Like Language

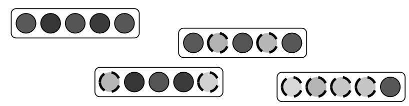
The big point: Even if children have innate knowledge of language structure, we still need to understand how they learn what the correct structural properties are for their particular language. One idea is to remember that children are good at tracking statistical information (like transitional probabilities)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




## 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.

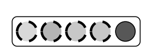
How do we translate this to learning language structure?

## 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)

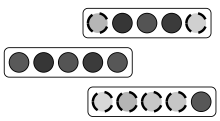


Fitness = how well a grammar can analyze the data the child encounters

## Yang (2004): Variational Learning

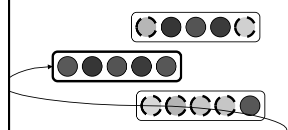
Idea taken from evolutionary biology:
A child's mind consists of a population of grammars that are competing to analyze the data in the child's native language.

Population of Grammars



## Yang (2004): Variational Learning

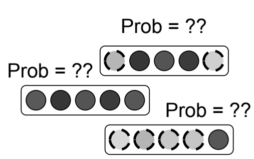
Intuition: The most successful (fittest) grammar will be the native language grammar because it can analyze all the data the child encounters. This grammar will "win", once the child encounters enough native language data because none of the other competing grammars can analyze all the data.



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

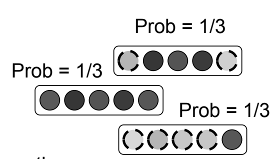
## Variational Learning Details

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.



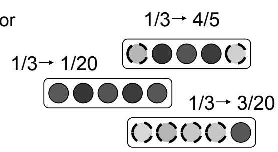

If there are 3 grammars, the initial probability for any given grammar = 1/3

## 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.

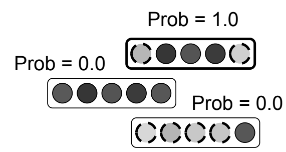
Other grammars will be less fit because they cannot account for some of the data encountered.

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.




## 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 these data are incompatible with the parameter value that is incorrect for the native language.

## Unambiguous data

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.

## Unambiguous issues

Parameter 1: subject-drop

Spanish: +subject-drop

Patterns allowed:

Vamos
*go-1$^{st}$-pl-pres*
“We go”

Subject dropped

Nosotros vamos
*1$^{st}$-pl* *go-1$^{st}$-pl-pres*
“We go”

Subject spoken

## Unambiguous issues

Parameter 1: subject-drop

English: -subject-drop

Patterns allowed:

*go-1$^{st}$-pl-pres*
“go” ≠ “we go”

Subject dropped

Subject spoken

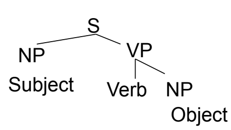
*1$^{st}$-pl* *go-pres*
“We go”

## Unambiguous issues

Parameter 2: Head-directionality

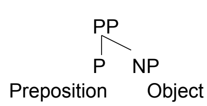
Edo/English: Head first

Basic word order:
Subject Verb Object [SVO]




Prepositions:
Preposition Noun Phrase

Possessed before Possessor
Possession Possessor

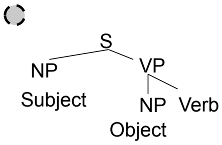



## Unambiguous issues

Parameter 2: Head-directionality

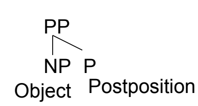
Japanese/Navajo: Head-final

Basic word order:
Subject Object Verb [SOV]




Postpositions:
Noun Phrase Postposition

Possessor before Possessed
Possessor Possession



## Unambiguous issues

Data point: Subject Object Verb

Grammars available:

| | | | |
|---|---|---|---|
| +subj-drop<br>+head-first | G1 | +subj-drop<br>-head-first | G2 |
| -subj-drop<br>+head-first | G3 | -subj-drop<br>-head-first | G4 |

## Unambiguous issues

Data point: Subject Object Verb

**Which grammars can analyze this data point?**

| | | | |
|---|---|---|---|
| +subj-drop<br>+head-first | G1 | +subj-drop<br>-head-first | G2 |
| -subj-drop<br>+head-first | G3 | -subj-drop<br>-head-first | G4 |

## Unambiguous issues

Data point: Subject Object Verb

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

| | | | |
|---|---|---|---|
| +subj-drop<br>+head-first | X G1 | +subj-drop<br>-head-first | G2 |
| -subj-drop<br>+head-first | G3 | -subj-drop<br>-head-first | G4 |

## Unambiguous issues

Data point: Subject Object Verb

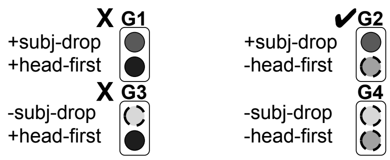
**G2?** ✔ +subj-drop allows Subject to be spoken
✔ -head-first predicts SOV

| | | | |
|---|---|---|---|
| +subj-drop<br>+head-first | X G1 | +subj-drop<br>-head-first | ✔G2 |
| -subj-drop<br>+head-first | G3 | -subj-drop<br>-head-first | G4 |

## Unambiguous issues

Data point: Subject Object Verb

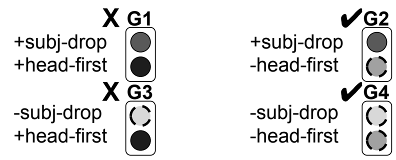
**G3?** ✔ -subj-drop requires Subject to be spoken
X +head-first predicts SVO




## Unambiguous issues

Data point: Subject Object Verb

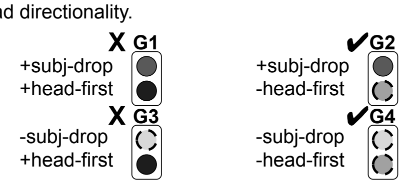
**G4?** ✔ -subj-drop requires Subject to be spoken
✔ -head-first predicts SOV




## 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 be informative for head directionality.

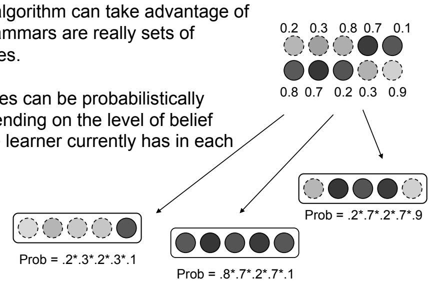



## Using parameters

Parameterized Grammars

Yang (2004)'s algorithm can take advantage of the fact that grammars are really sets of parameter values.

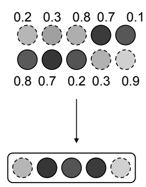
Parameter values can be probabilistically accessed, depending on the level of belief (probability) the learner currently has in each one.



## The Learning Algorithm

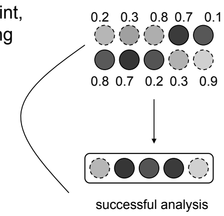
For each data point *d* encountered in the input

Choose a grammar to test out on a particular data point by generating a grammar from individual parameters, based on the probabilities associated with each parameter value.



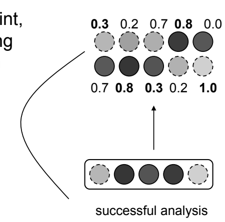

## The Learning Algorithm

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



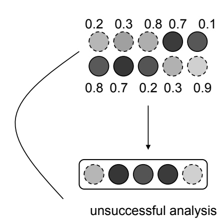

## The Learning Algorithm

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



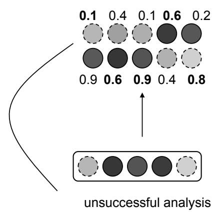

## The Learning Algorithm

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



## The Learning Algorithm

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



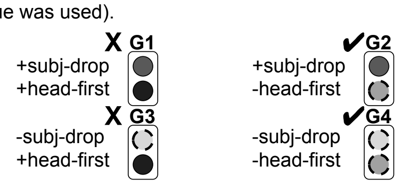

## Unambiguous data

Problem ameliorated!
Unambiguous data are much more likely to exist for individual parameter values instead of entire grammars.

## Unambiguous issues – no more!

Data point: Subject Object Verb

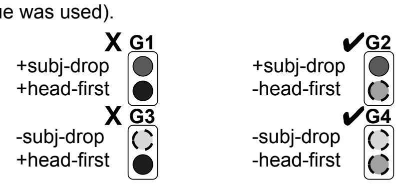
In this case, if either G2 or G4 were selected, -head-first would be rewarded (in addition to whichever subj-drop value was used).




## Unambiguous issues – no more!

Data point: Subject Object Verb

In this case, if either G1 or G3 were selected, +head-first would be punished (in addition to whichever subj-drop value was used).



## Unambiguous issues – no more!

Data point: Subject Object Verb

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

X G1
+subj-drop
+head-first

✔G2
+subj-drop
-head-first

X G3
-subj-drop
+head-first

✔G4
-subj-drop
-head-first

## Unambiguous issues – no more!

Data point: Subject Object Verb

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

X G1
+subj-drop
+head-first

✔G2
+subj-drop
-head-first

X G3
-subj-drop
+head-first

✔G4
-subj-drop
-head-first

## 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

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*

## Yang 2004: 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 guiji hui xiayu.
Tomorrow estimate will rain
'It is tomorrow that (*Jareth*) believes it will rain'

*Speaker doesn't have to say "Jareth"*

## Yang 2004: 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 guiji 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.

*Speaker has to say "Jareth"*

## Yang 2004: 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: 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: 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: Unambiguous Data Learning Examples

Subject drop

Parameter: +/- subject-drop

Native language value (English): -subject-drop

Unambiguous data: Expetive 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

## 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 often sees Marie*

Jean voit pas voit Marie
*Jean not sees 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 often sees Marie* "Jean often sees Marie."

Jean voit pas voit Marie
*Jean not sees 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 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.

*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

Observable (spoken) form of the question

(Hindi, German)
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 |
| +topic-drop (Chinese) | 12% | Before 1 yr, 8 months |
| +subject-drop (Italian) | 10% | Before 1 yr, 8 months |
| +verb-raising (French) | 7% | 1 yr, 8 months |
| +verb-second (German) | 1.2% | 3 yrs |
| -subject-drop (English) | 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?

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!