Psych 156A/ Ling 150: Acquisition of Language II

Lecture 16
Language Structure II

Announcements

Be working on structure review questions

A correction was made to last time's lecture notes – please make sure you have the most up-to-date version.

Read relevant sections of Yang (2011) as reference for next time ("learning as selection", "learnability and development").

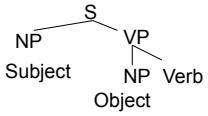
Please fill out online evaluation forms for this class! :)

Universal Grammar: Parameters

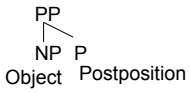
Parameters: Constrained variation across languages. Children must learn which option their native language uses.

Japanese/Navajo

Basic word order:
Subject Object Verb



Postpositions:
Noun Phrase Postposition



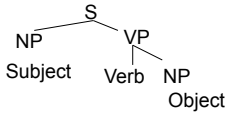
Possessor before Possessed
Possessor Possession

Universal Grammar: Parameters

Parameters: Constrained variation across languages. Children must learn which option their native language uses.

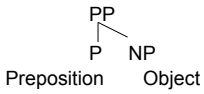
Edo/English

Basic word order:
Subject Verb Object



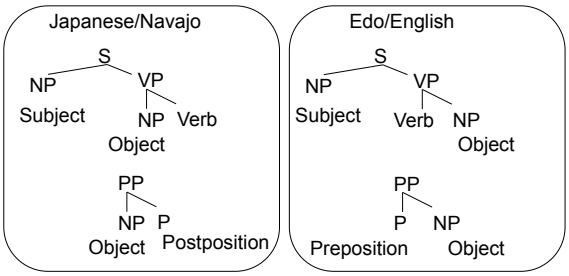
Prepositions:
Preposition Noun Phrase

Possessed before Possessor
Possession Possessor



Universal Grammar: Parameters

At this level of structural analysis (parameters), languages differ vary minimally from each other. This makes language structure much easier for children to learn. All they need to do is set the right parameters for their language, based on the data that are easy to observe.



But what are linguistic parameters really? How do they work? What exactly are they supposed to do?

Parameters

A parameter is meant to be something that can account for multiple observations in some domain.

Parameter for a statistical model: determines what the model predicts will be observed in the world in a variety of situations

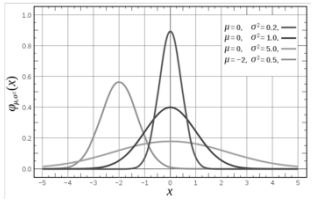
Parameter for our mental (and linguistic) model: determines what *we* predict will be observed in the world in a variety of situations

Statistical Parameters

The normal distribution is a statistical model that uses two parameters:

- μ for the mean
- σ for the standard deviation

$$\varphi_{\mu,\sigma^2}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

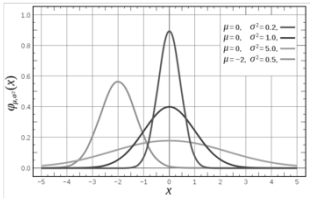


If we know the values of these parameters, we can make predictions about the likelihood of data we rarely or never see.

Statistical Parameters

Suppose this is a model of how many minutes late you'll be to class.

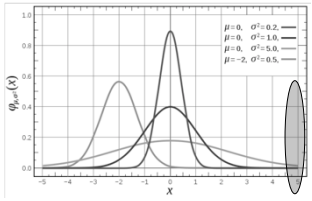
Let's use the model with $\mu = 0$, and $\sigma^2 = 0.2$. (blue line)



Statistical Parameters

Suppose this is a model of how many minutes late you'll be to class.

Let's use the model with $\mu = 0$, and $\sigma^2 = 0.2$. (blue line)

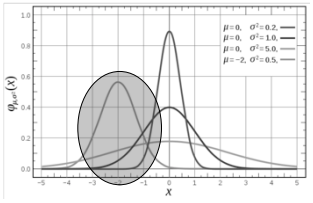


How likely are you to be 5 minutes late, given these parameters?

Not very likely! We can tell this just by knowing the values of the two statistical parameters. These parameter values allow us to infer the likelihood of some observed behavior.

Statistical Parameters

Observing different quantities of data with particular values can tell us which values of μ and σ^2 are most likely, if we know we are looking to determine the values of μ and σ^2 in function $\varphi(X)$



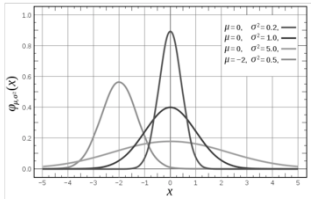
$$\varphi_{\mu,\sigma^2}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

Observing data points distributed like the green curve tells us that μ is likely to be around -2, for example.

Statistical vs. Linguistic Parameters

Important similarity:

We do not see the process that generates the data, but only the data themselves. This means that in order to form our expectations about X, we are, in effect, reverse engineering the observable data.



Our knowledge of the underlying function/principle that generates these data - $\varphi(X)$ - as well as the associated parameters - μ, and σ^2 - allows us to represent an infinite number of expectations about the behavior of variable X.

Linguistic principles vs. linguistic parameters

Both principles and parameters are often thought of as innate domain-specific abstractions that connect to many structural properties about language.

Linguistic principles correspond to the properties that are invariant across all human languages. Comparison: the equation's form– it is the statistical "principle" that explains the observed data.

$$\varphi_{\mu,\sigma^2}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

Linguistic parameters correspond to the properties that vary across human languages. Comparison: μ and σ^2 determine the exact form of the curve that represents the likelihood of observing certain data. While different values for these parameters can produce many different curves, these curves share their underlying form due to the common invariant function.

The utility of connecting to multiple properties

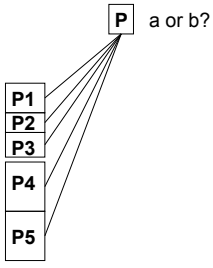
The fact that parameters connect to multiple structural properties then becomes a very good thing from the perspective of someone trying to acquire language. This is because a child can learn about that parameter's value by observing many different kinds of examples in the language.

"The richer the deductive structure associated with a particular parameter, the greater the range of potential 'triggering' data which will be available to the child for the 'fixing' of the particular parameter" – Hyams (1987)

Parameters can be especially useful when a child is trying to learn the things about language structure that are otherwise hard to learn, perhaps because they are very complex properties themselves or because they appear very infrequently in the available data

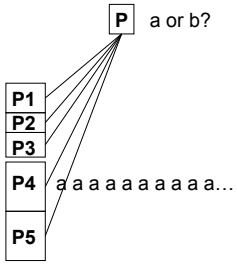
Why Hard-To-Learn Structures Are Easier

Let's assume a number of properties are all connected to parameter P, which can take one of two values: a or b.



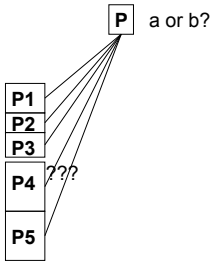
Why Hard-To-Learn Structures Are Easier

How do we learn whether P4 shows behavior a or b?
One way is to observe many instances of P4.



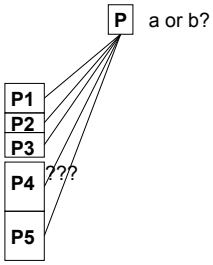
Why Hard-To-Learn Structures Are Easier

But what if P4 occurs very rarely? We might never see any examples of P4.



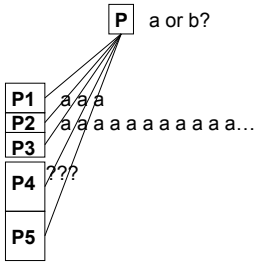
Why Hard-To-Learn Structures Are Easier

Fortunately, if P4 is connected to P, we can learn the value for P4 by learning the value of P. Also fortunately, P is connected to P1, P2, P3, and P5.



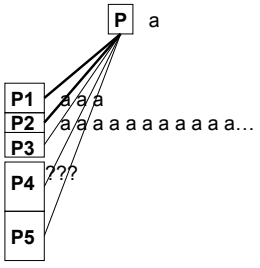
Why Hard-To-Learn Structures Are Easier

Step 1: Observe P1, P2, P3, *or* P5. In this case, all the observed examples of these structures are behavior a.



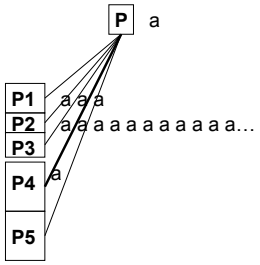
Why Hard-To-Learn Structures Are Easier

Step 2: Use this knowledge to set the value of parameter P to a.

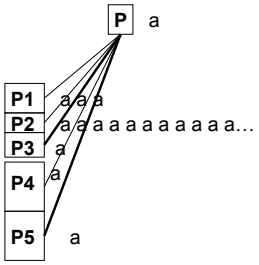


Why Hard-To-Learn Structures Are Easier

Step 3: Since parameter P is connected to P4, we can predict that P4 will also show behavior a - even though we've never seen any examples of it! (We can also infer P3 and P5 the same way.)



Why acquisition is easier



This highlights another benefit of parameters - we don't have to learn the behavior of each structure individually. Instead, we can observe some structures (ex: P1 and P2) and infer the right behavior for the remaining structures (P3, P4, and P5).

That is, instead of having to make 5 decisions (one for P1, P2, P3, P4, and P5), we actually only need to make one decision - is P a or b?

Hierarchical Bayesian learning links: Overhypotheses

Overhypotheses in hierarchical Bayesian learning are generalizations made at a more abstract level, which cover many different data types.

In this way, they are similar in spirit to linguistic parameters.

Hierarchical Bayesian learning links: Overhypotheses

Overhypothesis example

Suppose you are observing the contents of marble bags.

Hierarchical Bayesian learning links: Overhypotheses

Overhypothesis example

The first bag you look at has 20 black marbles.

20 ●

Hierarchical Bayesian learning links: Overhypotheses

Overhypothesis example

The second bag you look at has 20 white marbles.


20 ● 20 ○

Hierarchical Bayesian learning links: Overhypotheses

Overhypothesis example

The third and fourth bags you look at have 20 black marbles.




20 ● 20 ○ 20 ● 20 ●

Hierarchical Bayesian learning links: Overhypotheses

Overhypothesis example

You get a fifth bag and pull out a single marble. It's white. What do you predict about the color distribution of the rest of the marbles in the bag?

1 ○





20 ● 20 ○ 20 ● 20 ●

Hierarchical Bayesian learning links: Overhypotheses

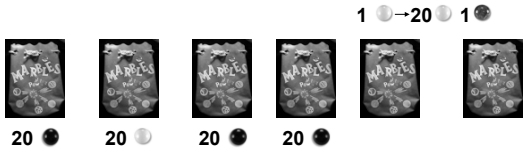
Overhypothesis example

Most adults predict this bag will contain 19 other white marbles, for a total of 20 white marbles.

Hierarchical Bayesian learning links: Overhypotheses

Overhypothesis example

What if you then get a sixth bag and pull out a single purple marble from it?



Hierarchical Bayesian learning links: Overhypotheses

Overhypothesis example

Most adults would predict that the other 19 marbles in that bag are purple too, for 20 purple marbles total.

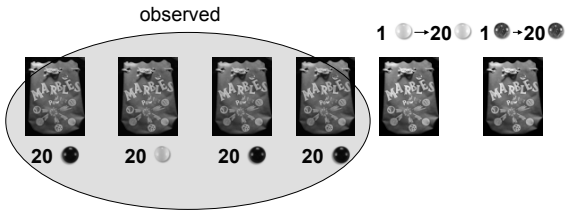
Hierarchical Bayesian learning links: Overhypotheses

Overhypothesis example

Why does this happen? It seems like you're learning something about the color distribution *in general* (not just for a particular bag): all marbles in a bag have the same color. This allows you to make predictions when you've only seen a single marble of whatever color from a bag.

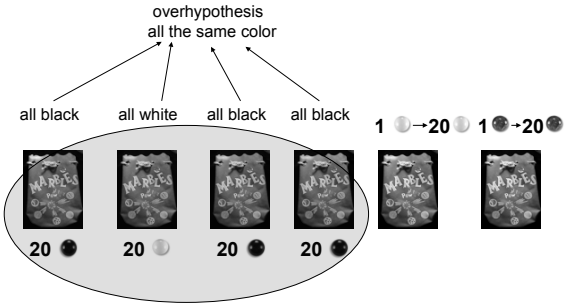
Hierarchical Bayesian learning links: Overhypotheses

Overhypothesis example



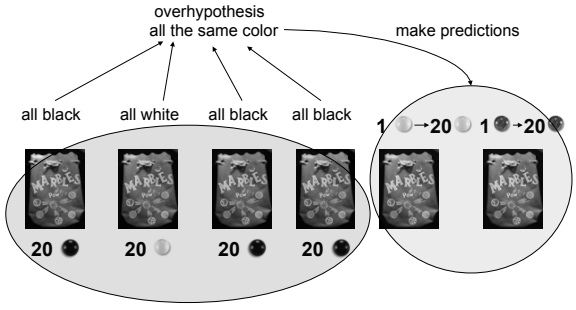
Hierarchical Bayesian learning links: Overhypotheses

Overhypothesis example



Hierarchical Bayesian learning links: Overhypotheses

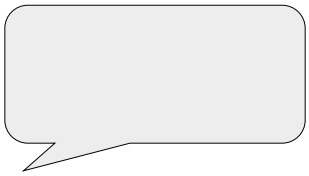
Overhypothesis example



Hierarchical Bayesian learning links: Overhypotheses

Linguistic overhypothesis example

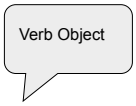
Suppose you are observing the word order of sentences you hear.



Hierarchical Bayesian learning links: Overhypotheses

Linguistic overhypothesis example

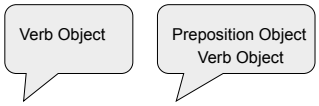
The first sentence you hear has the Verb before the Object ("See the penguin?")



Hierarchical Bayesian learning links: Overhypotheses

Linguistic overhypothesis example

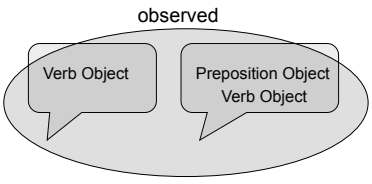
The second sentence you hear has the Preposition before the Object ("I like the penguin on the iceberg") and also the Verb before the Object ("I like the penguin on the iceberg").



Hierarchical Bayesian learning links: Overhypotheses

Linguistic overhypothesis example

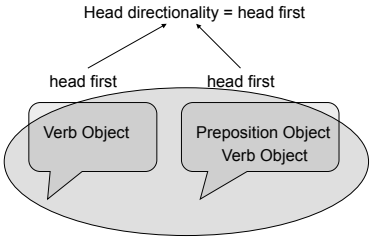
These data tell you about word order for verbs and objects and also about word order for prepositions and their objects.



Hierarchical Bayesian learning links: Overhypotheses

Linguistic overhypothesis example

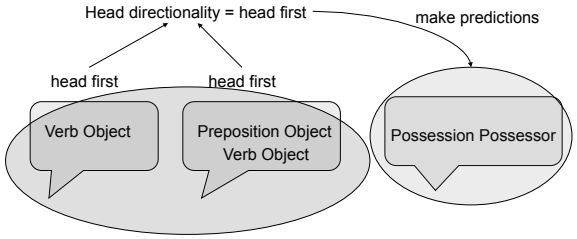
In addition, they are related via the head directionality parameter, which functions as an overhypothesis.



Hierarchical Bayesian learning links: Overhypotheses

Linguistic overhypothesis example

Knowing the value of this parameter allows you to predict other word order properties of the language.



Hierarchical Bayesian learning links: Overhypotheses

Learning Overhypotheses

Bayesian learner computational models are able to learn overhypotheses, provided they know what the parameters are and the range of values those parameters can take (ex: Kemp, Perfors, & Tenenbaum 2006).

What about real learners?

Learning overhypotheses: Dewar & Xu (2010)

9-month-olds

Question:

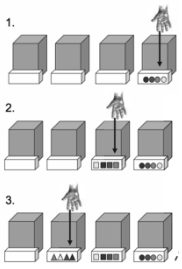
When provided with partial evidence about a few objects in a few categories, can infants form a more abstract generalization (an overhypothesis) that then applies to a new category?

Learning overhypotheses: Dewar & Xu (2010)

9-month-olds

Training trials:

Observe four different objects pulled out by experimenter who had her eyes closed - the objects are different colors but always have the same shape.



Learning overhypotheses: Dewar & Xu (2010)

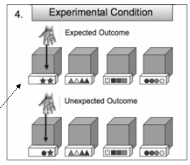
9-month-olds

Experimental trials:

Expected outcome (assuming infants had the overhypothesis that all the objects from a single box should be the same shape) =

Experimenter pulls out two objects with the same new shape.

Infants should not be surprised.



Learning overhypotheses: Dewar & Xu (2010)

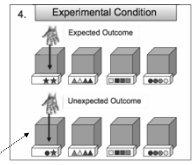
9-month-olds

Experimental trials:

Unexpected outcome (assuming infants had the overhypothesis that all the objects from a single box should be the same shape) =

Experimenter pulls out two objects with different shapes, one which is new and one which is old.

Infants should be surprised.



Learning overhypotheses: Dewar & Xu (2010)

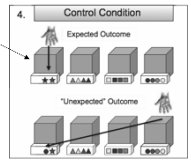
9-month-olds

Control trials:

Expected outcome (assuming infants had the overhypothesis that all the objects from a single box should be the same shape) =

Experimenter pulls out two objects with the same new shape.

Infants should not be surprised.



Learning overhypotheses: Dewar & Xu (2010)

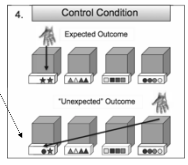
9-month-olds

Control trials:

Unexpected outcome control =

Experimenter pulls out two objects, one with a new shape that came from the new box and one with an old shape that came from an old box that contained that shape.

Infants should not be surprised if they have the overhypothesis.



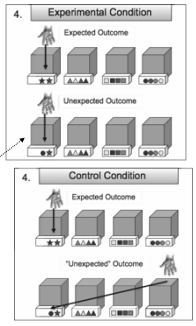
Learning overhypotheses: Dewar & Xu (2010)

9-month-olds

Results:

Infants in the experimental condition looked longer at the unexpected outcome (~14.28s) when compared to the expected outcome (~11.32s).

They were surprised at the evidence that didn't support the overhypothesis!



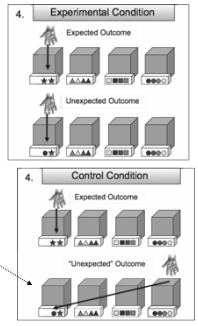
Learning overhypotheses: Dewar & Xu (2010)

9-month-olds

Results:

Infants in the control condition did not look longer at the expected outcome as compared to the unexpected outcome control that had the same objects present (~10.3-11.0s).

They were not surprised at the evidence that was compatible with the overhypothesis, even if the evidence involved two differently shaped objects.



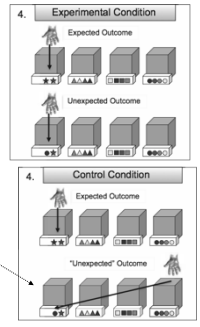
Learning overhypotheses: Dewar & Xu (2010)

9-month-olds

Overall result:

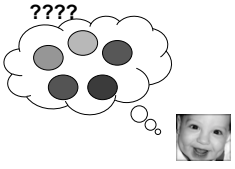
9-month-olds appear able to form overhypotheses from very limited data sets.

Hopefully, this means they can also use linguistic parameters to learn, since parameters are similar to overhypotheses about language!



Remaining problem even if infants have linguistic parameters

The observable data are often the result of a combination of 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

Summary: Linguistic Parameters

Linguistic parameters are similar to statistical parameters in that they are abstractions about the observable data. For linguistic parameters, the observable data are language data.

Parameters make acquisition easier because hard-to-learn structures can be learned by observing easy-to-learn structures that are connected to the same parameters.

Parameters may be similar to overhypotheses, which Bayesian learners and 9-month-olds are capable of learning.

Even with parameters, acquisition can be hard because a child has to figure out which parameter values produce the observable data, which isn't always easy.

Questions?

Be working on structure review questions – you should be able to do up through question 12.