Psych 156A/ Ling 150: Acquisition of Language II

Lecture 16

Structure I

Announcements

HW3 is due by the end of class today

Review questions are available for structure

Online course evaluations are available for this class - please fill them out! :)

Computational Problem: Figure out the order of words (syntax)

Jareth juggles crystals

Subject Verb Object

Noun Verb Noun

NP NP

Depends on grammatical categories like Nouns and Verbs (and their associated phrases (NP)), but also on more precise distinctions like Subjects and Objects.

Some Noun Phrase distinctions:

Subject = usually the agent/actor of the action, “doer”: Jareth

Object = usually the recipient of the action, “done to”: crystals

Computational Problem: Figure out the order of words (syntax)

Jareth juggles crystals

Subject Verb Object

Important idea: The observable word order speakers produce (like Subject Object Verb) is the result of a system of word order rules that speakers unconsciously use when they speak. This system of word order rules is called syntax.

Computational Problem: Figure out the order of words (syntax)

Jareth juggles crystals
Subject Verb Object

One way to generate Subject Verb Object order:
The linguistic system specifies that order as the general pattern of the language. An example of this kind of system is English.

English Subject Verb Object

Computational Problem: Figure out the order of words (syntax)

Jareth juggles crystals
Subject Verb Object

Another way to generate Subject Verb Object order:
The linguistic system specifies Subject Object Verb as the general pattern, but the Verb in main clauses moves to the second position and some other phrase (like the Subject) moves to the first position. An example language like this is German.

German Subject Object Verb

Computational Problem: Figure out the order of words (syntax)

Jareth juggles crystals
Subject Verb Object

Another way to generate Subject Verb Object order:
The linguistic system specifies Subject Object Verb as the general pattern, but the Verb in main clauses moves to the second position and some other phrase (like the Subject) moves to the first position. An example language like this is German.

movement rules

German ____ Verb Subject Object Verb

Computational Problem: Figure out the order of words (syntax)

Jareth juggles crystals
Subject Verb Object

Another way to generate Subject Verb Object order:
The linguistic system specifies Subject Object Verb as the general pattern, but the Verb in main clauses moves to the second position and some other phrase (like the Subject) moves to the first position. An example language like this is German.

movement rules

German Subject Verb Subject Object Verb

Computational Problem: Figure out the order of words (syntax)

Jareth juggles crystals

Subject Verb Object

A third way to generate Subject Verb Object order:
The linguistic system specifies Subject Object Verb as the general pattern, but the Object moves after the Verb in certain contexts (the Object is unexpected information). Kannada is a language like this.

Kannada Subject Object Verb

Computational Problem: Figure out the order of words (syntax)

Jareth juggles crystals

Subject Verb Object

A third way to generate Subject Verb Object order:
The linguistic system specifies Subject Object Verb as the general pattern, but the Object moves after the Verb in certain contexts (the Object is unexpected information). Kannada is a language like this.

movement rule

Kannada Subject *Object* Verb Object

Computational Problem: Figure out the order of words (syntax)

Jareth juggles crystals

Subject Verb Object

English

Subject Verb Object

German

Subject Verb *Subject* Object *Verb*

Kannada

Subject *Object* Verb Object

The learning problem: How do children know which system their language uses?

Computational Problem: Figure out the order of words (syntax)

Jareth juggles crystals

Subject Verb Object

English

Subject Verb Object

German

Subject Verb *Subject* Object *Verb*

Kannada

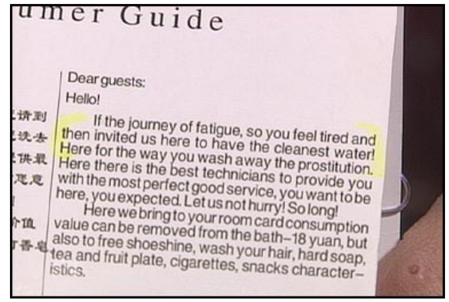
Subject *Object* Verb Object

This is a hard question!

Children only see the output of the system (the observable word order of Subject Verb Object).

Syntax: One reason why translation is so hard

Translation is not so easy: more than just word-by-word



http://www.nbc.com/nbc/The_Tonight_Show_with_Jay_Leno/headlines/

Translation is not so easy: more than just word-by-word

translate.google.com

Through dangers untold and hardships unnumbered, I have fought my way here to the castle beyond the goblin city to take back the child that you have stolen.

Hebrew

דרך סכנות עצומות וקשיים לא ממוספרים, יש לי נלחם בדרך שלי כאן לטירה מעבר לעיר גובלין לקחת בחזרה את הילד שיש לך נגנב.

Literally:
Through dangers immense and difficulties not numbered, there-is to-me fighting through my here castle transition city goblin take back you child there-was to-you stolen.

Translation is not so easy: more than just word-by-word

translate.google.com

Through dangers untold and hardships unnumbered, I have fought my way here to the castle beyond the goblin city to take back the child that you have stolen.

Haitian Creole

Atravè danje inonbrabl ak difikilte inonbrabl, mwen te goumen jan m 'isit la yo chato la pi lwen pase lavil la Goblin yo pran tounen timoun nan ke ou te vòlè li.

Literally:
Through danger countless and difficulties countless, I was fight how me here they mansion the more far than cities the Goblin they take back children of that you was thief it.

Translation is not so easy: more than just word-by-word

translate.google.com

Through dangers untold and hardships unnumbered, I have fought my way here to the castle beyond the goblin city to take back the child that you have stolen.

Hindi

अनकहा और बेशुमार कठिनाइयों खतरों के माध्यम से, मैं तुम्हें चुराया है कि बच्चे को वापस लेने के लिए भूत शहर परे महल को यहाँ अपने तरीके से लड़ाई लड़ी है.

Literally:

Untold and uncountable difficulties threats medium through, I you stole is that children back take the ghost city beyond palace the here your methods from fight fought.

About human knowledge: Language & variation

Navajo Code Talkers

Crucial cryptographic method used in World War II

http://en.wikipedia.org/wiki/Code_talker#Use_of_Navajo

"…Johnston saw Navajo as answering the military requirement for an undecipherable code. Navajo was spoken only on the Navajo lands of the American Southwest, and its syntax and tonal qualities, not to mention dialects, made it unintelligible to anyone without extensive exposure and training. One estimate indicates that at the outbreak of World War II fewer than 30 non-Navajos could understand the language.…"

https://www.youtube.com/watch?v=5rSvm3m8ZUA
(~3 min video)

Navajo Code Talker Paradox (Baker 2001)

English must be very different from Navajo

Japanese could decode English, but couldn't decode Navajo when they didn't know it was Navajo.

English must be similar to Navajo

English can be translated into Navajo and back with no loss of meaning. (Languages are not just a product of the culture - pastoral Arizona lifestyle couldn't have prepared the code talkers for Pacific Island high tech warfare. Yet, translation was still possible.)

Types of variation

Morphology (word forms)

English: invariant word forms

"the girl is crying", "I am crying"

Navajo: no invariant forms (there may be 100-200 prefixes for verb stems)

At'ééd yicha. "Girl crying"

Yishcha. "I am crying"
(yi + sh + cha)

Nináháhwiishdlaad. "I am again plowing"
(ni + náá + ho + hi + sh + l + dlaad)

Types of variation

Word order (syntax)

English: Subject Verb Object (invariant word order)

"The boy saw the girl"

Navajo: Subject Object Verb, Object Subject Verb
(varying word orders, meaning depends only on verb's form)

Ashkii at'ééd **yiyiiltsá**
boy girl saw
"The boy saw the girl"

Ashkii at'ééd **biiltsá**
boy girl saw
"The girl saw the boy"

Types of variation

wals.info: The World Atlas of Language Structures

Welcome to WALS Online

The World Atlas of Language Structures (WALS) is a large database of structural (phonological, grammatical, lexical) properties of languages gathered from descriptive materials (such as reference grammars) by a team of 55 authors.

Types of variation

Let's look at syntax…

81	Order of Subject, Object and Verb	Matthew S. Dryer	Word Order
82	Order of Subject and Verb	Matthew S. Dryer	Word Order
83	Order of Object and Verb	Matthew S. Dryer	Word Order

Chapter Order of Subject and Verb

by Matthew S. Dryer cite

Values of Map 82A. Order of Subject and Verb

	Value	Representation
●	Subject precedes verb (SV)	1193
●	Subject follows verb (VS)	194
●	Both orders with neither order dominant	110
	Total:	1497

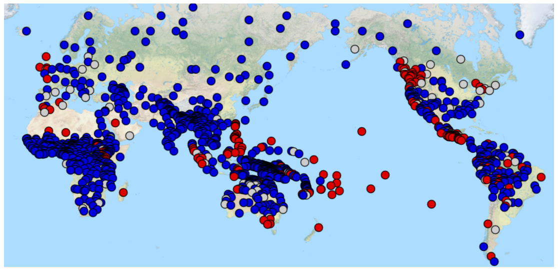
Jack laughs.

Laughs Jack.

This tells us that most languages have the Subject come before the Verb…but not all do.

Types of variation

Let's look at syntax…



Values

●	SV	1193
●	VS	194
○	No dominant order	110

How are the different Subject and Verb orders distributed around the world?

Home | Features | Chapters | Languages | References | Authors

Features

A feature is a structural property of language that describes one aspect of cross-linguist maps. Most features correspond straightforwardly to chapters, but some chapters are a

82A	Order of Subject and Verb	Matthew S. Dryer	Word Order	1497

Types of variation

Let's look at syntax…

What value does English have?

English ● SV

What about Fijian?

Fijian ● VS

What about Spanish?

Spanish ○ No dominant order

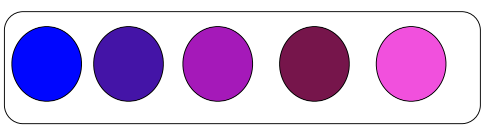
Home | Features | Chapters | Languages | References | Authors

Features

A feature is a structural property of language that describes one aspect of cross-linguist maps. Most features correspond straightforwardly to chapters, but some chapters are a

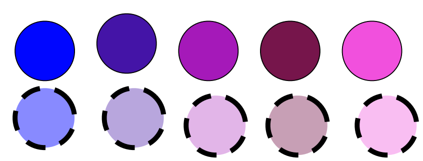
82A	Order of Subject and Verb	Matthew S. Dryer	Word Order	1497

Thinking about syntactic variation



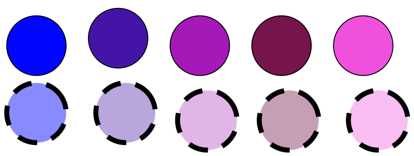
Similarities & differences: Parameters

Chomsky: Different combinations of different basic elements (parameters) would yield the observable languages (similar to the way different combinations of different basic elements in chemistry yield many different-seeming substances).



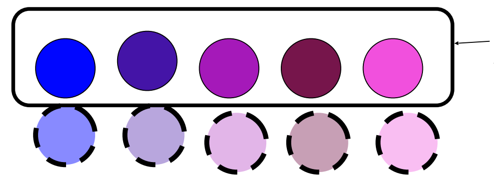
Similarities & differences: Parameters

Big Idea: A relatively small number of syntax parameters yields a large number of different languages' syntactic systems.



Similarities & differences: Parameters

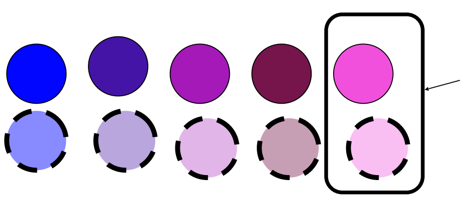
Big Idea: A relatively small number of syntax parameters yields a large number of different languages' syntactic systems.



5 different parameters of variation

Similarities & differences: Parameters

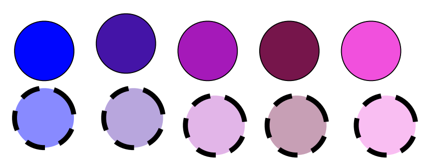
Big Idea: A relatively small number of syntax parameters yields a large number of different languages' syntactic systems.



2 different parameter values of one parameter

Similarities & differences: Parameters

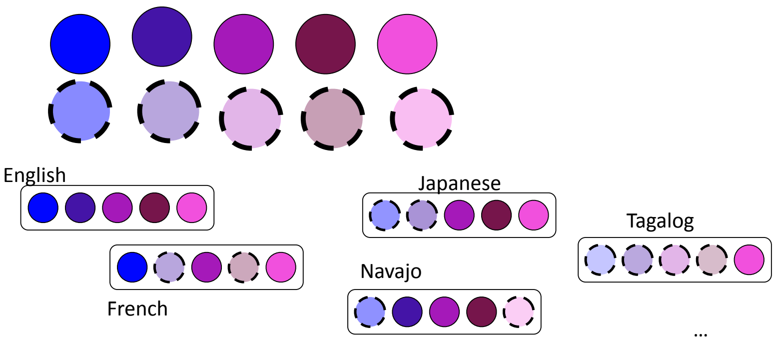
Big Idea: A relatively small number of syntax parameters yields a large number of different languages' syntactic systems.



Total languages that can be represented
$= 2*2*2*2*2$
$= 2^5$
$= 32$

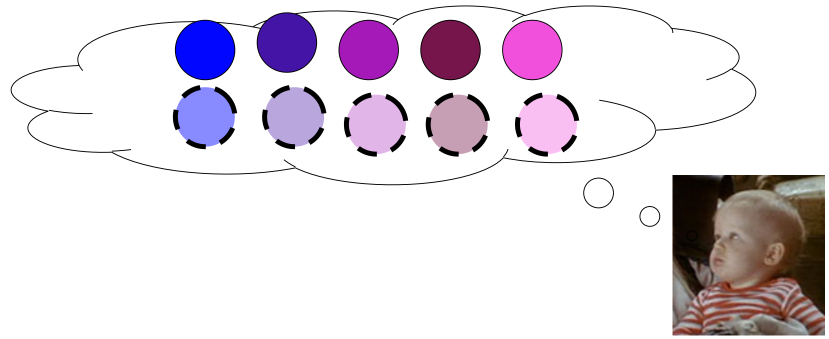
Similarities & differences: Parameters

Big Idea: A relatively small number of syntax parameters yields a large number of different languages' syntactic systems.



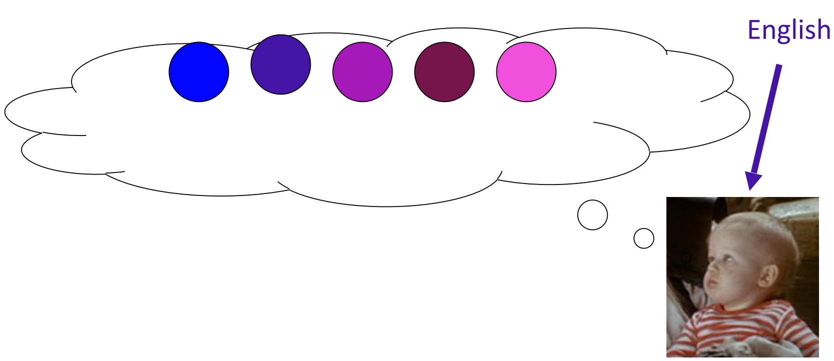
Learning language structure

Chomsky: Children are born knowing the parameters of variation. This is part of Universal Grammar. Input from the native linguistic environment determines what values these parameters should have.



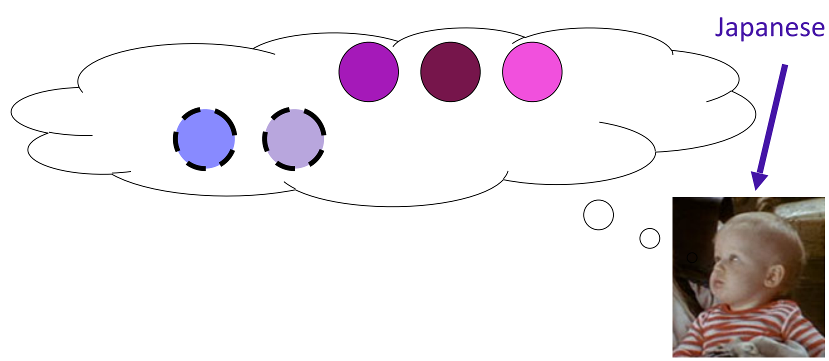
Learning language structure

Chomsky: Children are born knowing the parameters of variation. This is part of Universal Grammar. Input from the native linguistic environment determines what values these parameters should have.



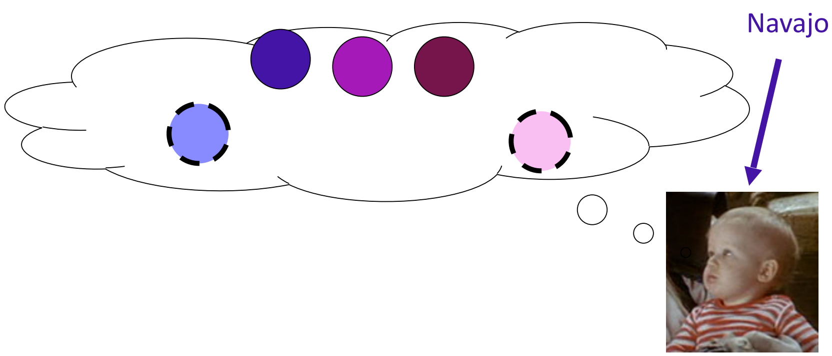
Learning language structure

Chomsky: Children are born knowing the parameters of variation. This is part of Universal Grammar. Input from the native linguistic environment determines what values these parameters should have.



Learning language structure

Chomsky: Children are born knowing the parameters of variation. This is part of Universal Grammar. Input from the native linguistic environment determines what values these parameters should have.



Generalizations about language structure

Greenberg's word order generalizations

Navajo

Japanese

Greenberg's word order generalizations

Navajo	Japanese
Basic word order: Subject Object Verb	Basic word order: Subject Object Verb
Ashkii at'éé́d yiyiiltsá	Jareth-ga Hoggle-o butta
boy girl saw	*Jareth Hoggle hit*
"The boy saw the girl"	"Jareth hit Hoggle"

Greenberg's word order generalizations

Navajo

Postpositions:
Noun Phrase Postposition

'éé' biih náásdzá
clothing into I-got-back
"I got back into (my) clothes."

Japanese

Postpositions:
Noun Phrase Postposition

Jareth-ga Sarah to kuruma da
Jareth Sarah with car by

London ni itta
London to went

"Jareth went to London with Sarah by car."

Greenberg's word order generalizations

Navajo

Possessor before Possessed

Possessor Possession

Chidí bi-jáád
Car its-leg

"the car's wheel"

Japanese

Possessor before Possessed

Possessor Possession

Toby-no imooto-ga
Toby's sister

"Toby's sister"

Greenberg's word order generalizations

Navajo

Basic word order:
Subject Object Verb

Postpositions:
Noun Phrase Postposition

Possessor before Possessed
Possessor Possession

Japanese

Basic word order:
Subject Object Verb

Postpositions:
Noun Phrase Postposition

Possessor before Possessed
Possessor Possession

Despite the differences in the languages (and their cultural histories), both Japanese and Navajo are very similar when viewed through these three structural descriptions.

Greenberg's word order generalizations

English

Edo (Nigeria)

Greenberg's word order generalizations

English

Basic word order:
Subject Verb Object

Sarah found Toby

Edo (Nigeria)

Basic word order:
Subject Verb Object

Òzó mién Adésuwá

Ozo found Adesuwa

Greenberg's word order generalizations

English

Prepositions:
Preposition Noun Phrase

Jareth gave the crystal to Sarah

Edo (Nigeria)

Prepositions:
Preposition Noun Phrase

Òzó rhié néné ebé né Adésuwá
Ozo gave the book to Adesuwa

Greenberg's word order generalizations

English

Possessed before Possessor

Possession Possessor

quest of Sarah

(alternative: Sarah's quest)

Edo (Nigeria)

Possessed before Possessor

Possession Possessor

Omo Ozó
child Ozo

"child of Ozo"

Greenberg's word order generalizations

English

Basic word order:
Subject Verb Object

Prepositions:
Preposition Noun Phrase

Possessed before Possessor
Possession Possessor

Edo (Nigeria)

Basic word order:
Subject Verb Object

Prepositions:
Preposition Noun Phrase

Possessed before Possessor
Possession Possessor

Again, despite the differences in the languages (and their cultural histories), both English and Edo are very similar when viewed through these three structural descriptions.

Greenberg's word order generalizations

Greenberg found forty-five "universals" of languages - patterns overwhelmingly followed by languages with unshared history (Navajo & Japanese, English & Edo)

Not all combinations are possible - some patterns rarely appear
Ex: Subject Verb Object language (English/Edo-like) + postpositions (Navajo/Japanese-like)

Moral: Languages may be more similar than they first appear "on the surface", especially if we consider their structural properties.

One potential parameter

English	Italian
Subject Verb	Subject Verb Jareth verrá *Jareth will-come*
"Jareth will come."	"Jareth will come."
grammatical	grammatical

One potential parameter

English	Italian
*Verb Subject	Verb Subject Verrá Jareth *Will-arrive Jareth*
**Will arrive Jareth*	
	"Jareth will arrive"
ungrammatical	grammatical

One potential parameter

English	Italian
*Verb	Verb Verrá *He-will-come*
Will come	
	"He will come"
ungrammatical	grammatical

One potential parameter

English	Italian
Subject Verb	Subject Verb
*Verb Subject	Verb Subject
*Verb	Verb

These word order patterns might be fairly easy to notice. They involve the combinations of Subject and Verb that are grammatical in the language. A child might be able to notice the prevalence of some patterns and the absence of others.

One potential parameter

Expletive subjects: words without content (may be more difficult to notice)

English	Italian
Raining.	Piove. It-rains.
"It's raining."	"It's raining."
Not okay to leave out expletive subject "it".	Okay to leave out expletive subject "it".

One potential parameter

That-trace effect for subject questions

English

Italian

Who do you think (*that) will come?

Requires no "that" in embedded clause, despite allowing "that" in declaratives and object questions

I think (that) Hoggle will save Sarah.
Who did you think (that) Hoggle would save?

One potential parameter

That-trace effect for subject questions

English

Italian

Credi che Jareth verrá.
You think that Jareth will-come.
"You think that Jareth will come."

Che credi che __ verrá?
Who think-you that will-come?
"Who do you think will come?"

Allows "that" in the embedded clause of a subject question (and declarative clauses).

One potential parameter

English	Italian
Subject Verb	Subject Verb
*Verb Subject	Verb Subject
*Verb	Verb
Not okay to leave out expletive subject "it".	Okay to leave out expletive subject "it".
Requires special action for embedded subject questions.	Does not require special action for embedded subject questions.

All these involve the subject in some way - coincidence?
Idea: No! There's a language parameter involving the subject.

The Value of Parameters: Learning the hard stuff by noticing the easy patterns

English vs. Italian: Subject Parameter

English	Italian
Subject Verb	Subject Verb
*Verb Subject	Verb Subject
*Verb	Verb

Easier to notice

Hard to notice

Expletives

English	Italian
It rains	Piove. It-rains.

Embedded Subject-question formation (easy to miss)

English	Italian
Who do you think (*that) will come?	Che credi che __ verrá? *Who think-you that will-come?*

The Value of Parameters: Learning the hard stuff by noticing the easy patterns

English vs. Italian: Subject Parameter

Big idea: If all these structural patterns are generated from the same linguistic parameter (e.g. a "subject" parameter), then children can learn the hard-to-notice patterns (like the patterns of embedded subject questions) by being exposed to the easy-to-notice patterns (like the optional use of subjects with verbs). The hard-to-notice patterns are generated by one setting of the parameter, which children can learn from the easy-to-notice patterns.

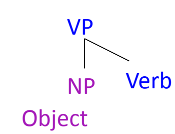
Children's knowledge of language structure variation is believed by linguistic nativists to be part of Universal Grammar, which children are born with.

Another possible parameter

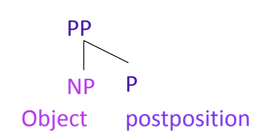
Syntax: the Head Directionality parameter (Baker 2001, Cook & Newson 1996): heads of phrases (ex: Nouns of Noun Phrases, Verbs of Verb Phrases, Prepositions of Preposition Phrases) are consistently in either the leftmost or rightmost position

Japanese/Navajo: Head-Last

Verb Phrase:
Object Verb



Postpositions:
Noun Phrase Postposition

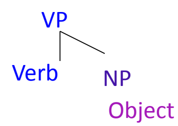


Another possible parameter

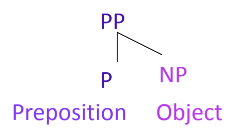
Syntax: the Head Directionality parameter (Baker 2001, Cook & Newson 1996): heads of phrases (ex: Nouns of Noun Phrases, Verbs of Verb Phrases, Prepositions of Preposition Phrases) are consistently in either the leftmost or rightmost position

Edo/English: Head-First

Verb Phrase:
Verb Object



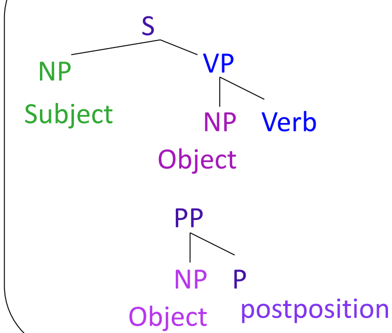
Prepositions:
Preposition Noun Phrase



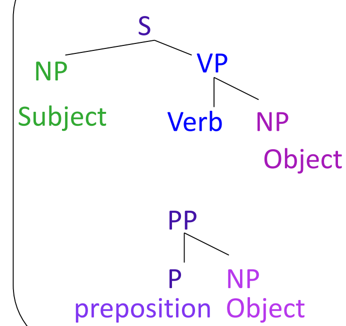
Universal Grammar: Parameters

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 parameter values for their language, based on the data that are easy to observe.

Japanese/Navajo



Edo/English



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

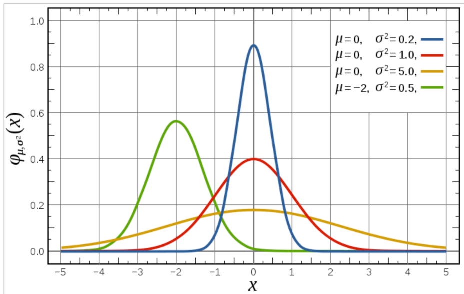
Parameter for our mental (and linguistic) model: determines what *we* predict will be observed in the world in a variety of situations

Statistical parameters

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

- μ for the mean
- σ for the standard deviation

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

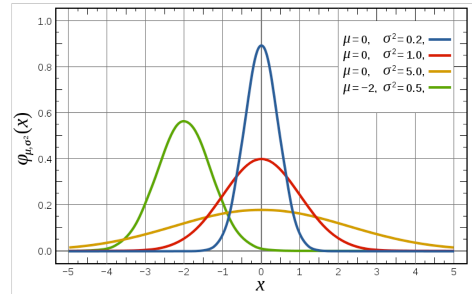


If we know the values of these parameters, we can make predictions about the likelihood of data we rarely or never see.

Statistical parameters

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

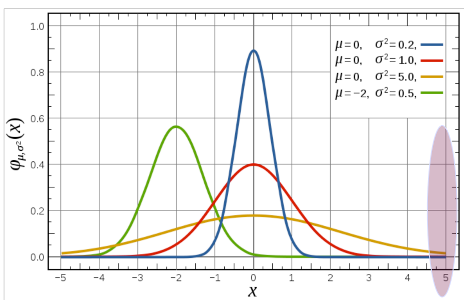
Let's use the model with μ = 0, and σ^2 = 0.2. (blue line)



Statistical parameters

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

Let's use the model with μ = 0, and σ^2 = 0.2. (blue line)



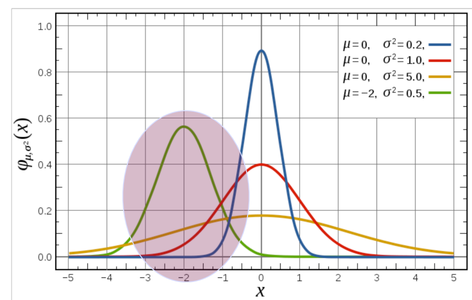
How likely are you to be 5 minutes late, given these parameters?

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 probability 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 $\phi(X)$

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

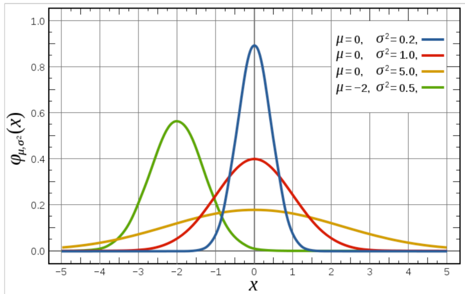


Observing data points distributed like the green curve tells us that μ is likely to be around -2, for example.

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 - ϕ(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.

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

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 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.

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

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.

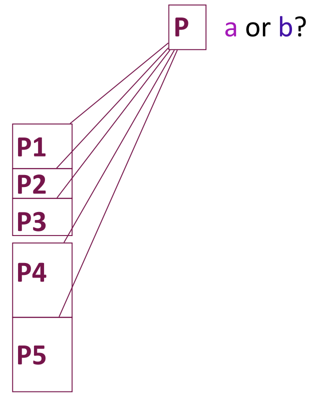
"The richer the deductive structure associated with a particular parameter, the greater the range of potential 'triggering' data which will be available to the child for the 'fixing' of the particular parameter" – Hyams (1987)

The utility of connecting to multiple properties

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

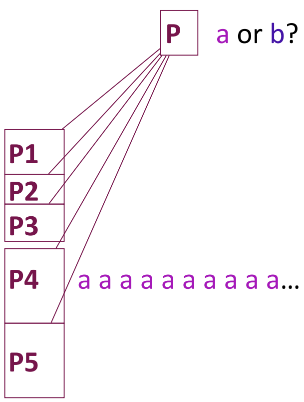
Let's assume a number of properties are all connected to parameter P, which can take one of two values: a or b.



Why hard-to-learn structures are easier

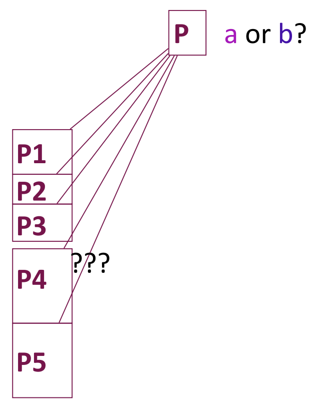
How do we learn whether P4 shows behavior a or b?

One way is to observe many instances of P4.



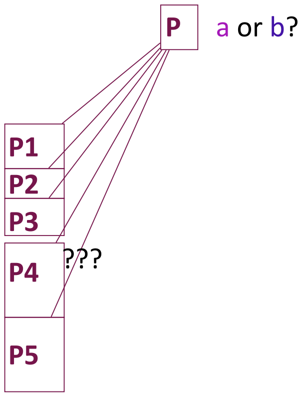
Why hard-to-learn structures are easier

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



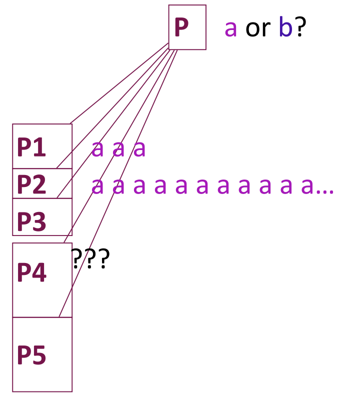
Why hard-to-learn structures are easier

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.



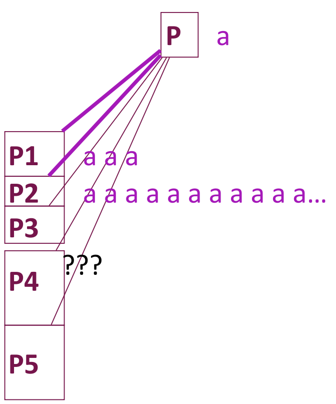
Why hard-to-learn structures are easier

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



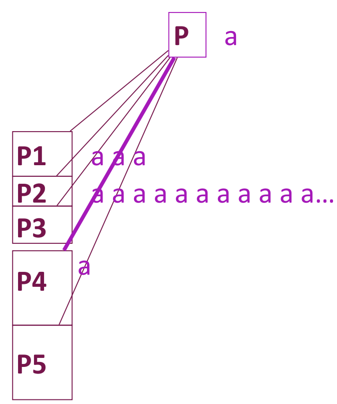
Why hard-to-learn structures are easier

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

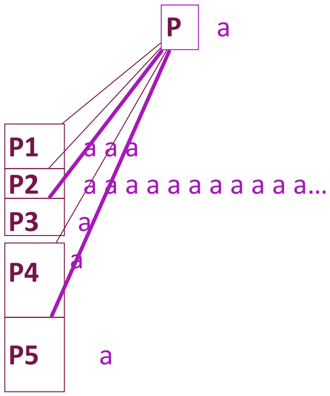


Why hard-to-learn structures are easier

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.

Hierarchical Bayesian learning links: Overhypotheses

Overhypothesis example

The second bag you look at has 20 white marbles.

Hierarchical Bayesian learning links: Overhypotheses

Overhypothesis example

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

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?

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.

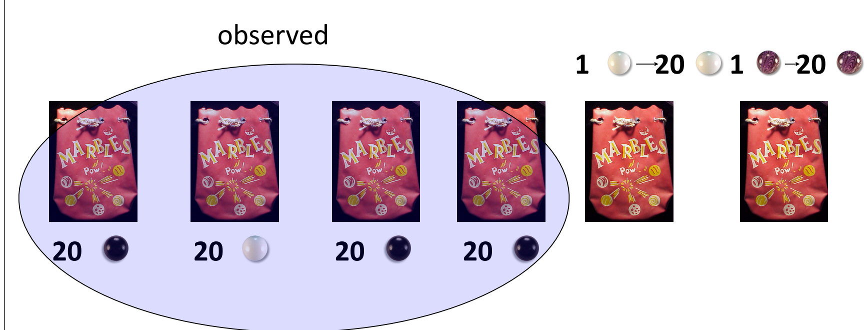
Hierarchical Bayesian learning links: Overhypotheses

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.

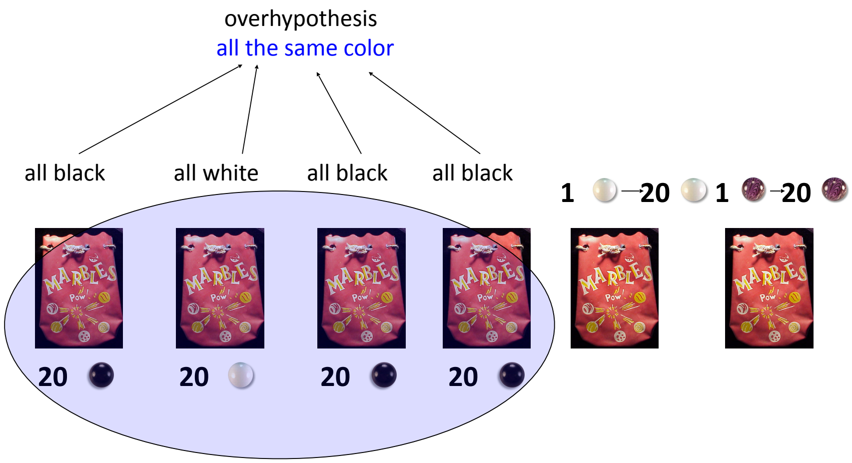
Hierarchical Bayesian learning links: Overhypotheses

Overhypothesis example



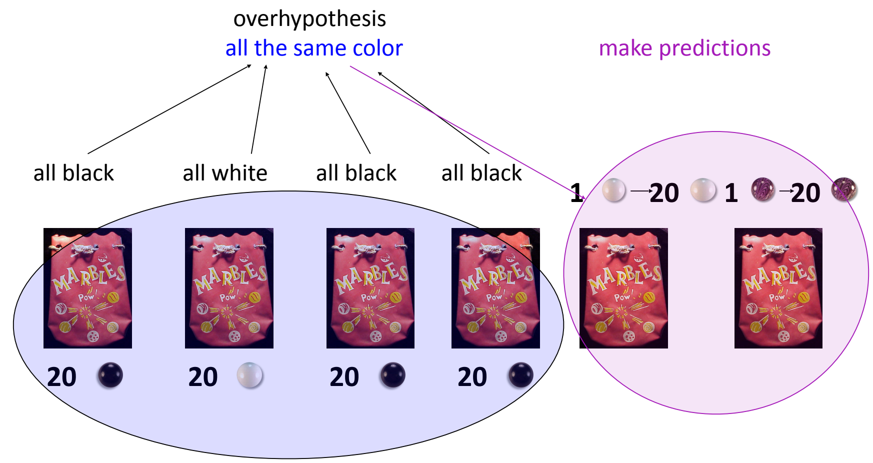
Hierarchical Bayesian learning links: Overhypotheses

Overhypothesis example



Hierarchical Bayesian learning links: Overhypotheses

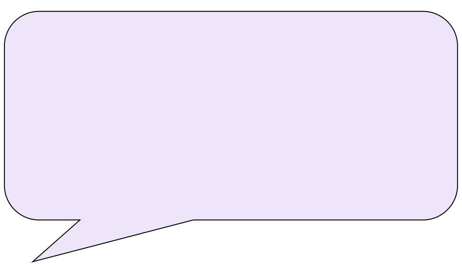
Overhypothesis example



Hierarchical Bayesian learning links: Overhypotheses

Linguistic overhypothesis example

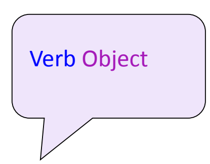
Suppose you are observing the word order of sentences you hear.



Hierarchical Bayesian learning links: Overhypotheses

Linguistic overhypothesis example

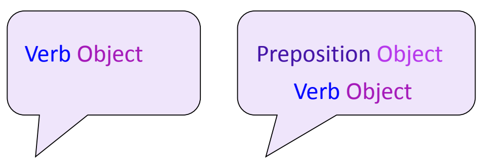
The first sentence you hear has the Verb before the Object ("See the penguin?")



Hierarchical Bayesian learning links: Overhypotheses

Linguistic overhypothesis example

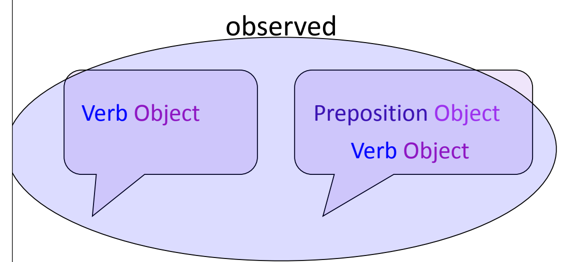
The second sentence you hear has the Preposition before the Object ("I like the penguin on the iceberg") and also the Verb before the Object ("I like the penguin on the iceberg").



Hierarchical Bayesian learning links: Overhypotheses

Linguistic overhypothesis example

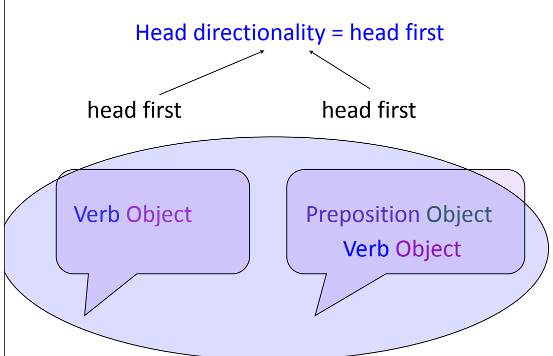
These data tell you about word order for verbs and objects and also about word order for prepositions and their objects.



Hierarchical Bayesian learning links: Overhypotheses

Linguistic overhypothesis example

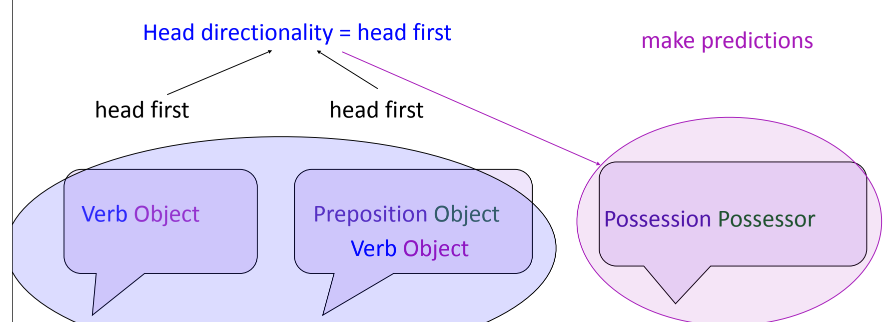
In addition, they are related via the head directionality parameter, which functions as an overhypothesis.



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:

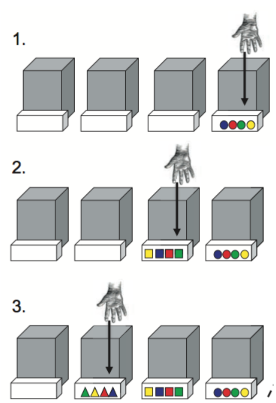
When provided with partial evidence about a few objects in a few categories, can infants form a more abstract generalization (an overhypothesis) that then applies to a new category?

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

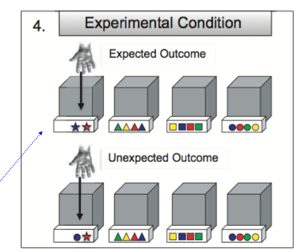
Experimental trials:

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.



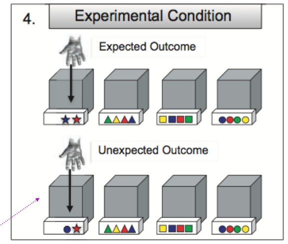
Learning overhypotheses: Dewar & Xu (2010)

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.



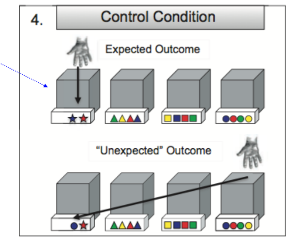
Learning overhypotheses: Dewar & Xu (2010)

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.



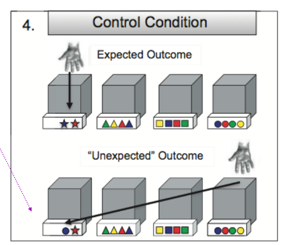
Learning overhypotheses: Dewar & Xu (2010)

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 this outcome is compatible with the overhypothesis. (The overhypothesis is actually irrelevant here.)



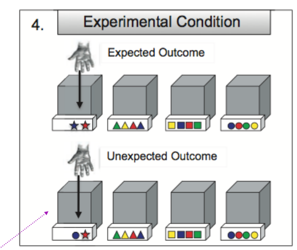
Learning overhypotheses: Dewar & Xu (2010)

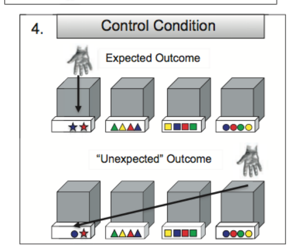
9-month-olds

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!





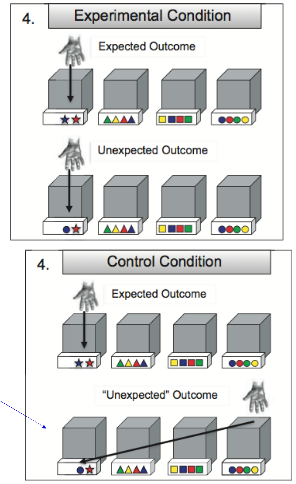
Learning overhypotheses: Dewar & Xu (2010)

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.



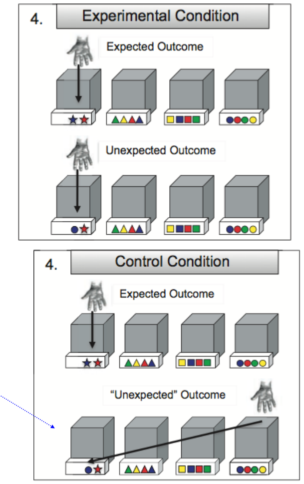
Learning overhypotheses: Dewar & Xu (2010)

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!



Summary: Linguistic parameters

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.

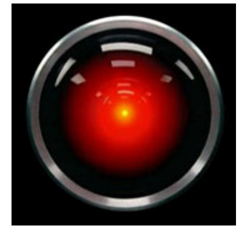
Linguistic parameters are similar to statistical parameters in that they are abstractions about the observable data. For linguistic parameters, the observable data are language data.

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

Questions?

You should be able to do up through question 11 on the structure review questions.

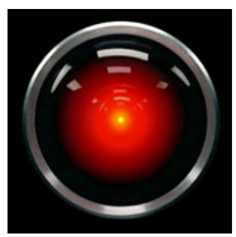
Extra Material



Syntax: One reason why natural language comprehension is so hard for computers

Solving the Language Problem (Artificial Intelligence)

HAL 9000 from 2001: A Space Odyssey (1968)



Perfect production and comprehension of English.

1960s: Language not considered one of the “hard” problems of artificial intelligence.

2010: Getting better but still not perfect.

http://www.research.att.com/~ttsweb/tts/demo.php

Solving the Language Problem (Artificial Intelligence)

2012: Apple’s Siri is getting closer, though still has problems …

http://bits.blogs.nytimes.com/2012/07/15/with-apple’s-siri-a-romance-gone-sour/?_php=true&_type=blogs&_r=0

Late last summer, I was introduced to a new special someone. I wasn’t looking to meet this new muse; it all just kind of happened.

We met at an Apple product announcement in Cupertino, Calif. She was helpful, smart and even funny, cracking sarcastic jokes and making me laugh. What more could a guy ask for?

Since then, we have had some major communication issues. She frequently misunderstands what I’m saying. Sometimes she is just unavailable. Often, she responds with the same, repetitive statement.

Her name is Siri.

Solving the Language Problem (Artificial Intelligence)

Contrast: Chess-playing.

In 1997, a program named Deep Blue beat the reigning world champion in chess. It did this by having enough computational resources to investigate every move option before it actually made the chess move. This shows that computers' poor performance on language is not about insufficient computational power, since there is enough computational power to solve the chess-playing problem (which some people might consider a very difficult problem).

Solving the Language Problem (Artificial Intelligence)

Update for 2011 on a machine's abilities to do what humans do:

Man vs. Machine (Watson) in Jeopardy
& how hard a problem language comprehension and production is

http://www.youtube.com/watch?v=dr7IxQeXr7g
(approximately 9 min video)

Watson vs. all humanity
https://www.youtube.com/watch?v=WFR3lOm_xhE
(approximately 4 min video)

Solving the Language Problem (Artificial Intelligence)

2013: True on-the-fly language comprehension is still pretty hard, as well as determining the answer to "commonsense" questions that are phrased naturally.

http://www.sciencedaily.com/releases/2013/07/130715151059.htm

"One of the hardest problems in building an artificial intelligence, Sloan said, is devising a computer program that can make sound and prudent judgment based on a simple perception of the situation or facts-the dictionary definition of commonsense.

Commonsense has eluded AI engineers because it requires both a very large collection of facts and what Sloan calls implicit facts — things so obvious that we don't know we know them. A computer may know the temperature at which water freezes, but we know that ice is cold." - Jeanne Galatzer-Levy

"We're still very far from programs with commonsense-AI that can answer comprehension questions with the skill of a child of 8," said Sloan. He and his colleagues hope the study will help to focus attention on the "hard spots" in AI research.

Types of variation

Vocabulary
English "think" verbs: think, know, wonder, suppose, assume, …

Multiple types of the action verb "think". Each has certain uses that are appropriate.

"I wonder whether the girl saved her little brother from the goblins." [grammatical]

* "I suppose whether the girl saved her little brother from the goblins." [ungrammatical]

Types of variation

Vocabulary

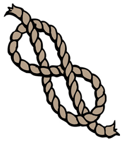
English "think" verbs: think, know, wonder, suppose, assume, …

Navajo "carry" verbs: depends on object being carried

aah (carry a solid round-ish object)

kaah (carry an open container with contents)

lé (carry a flexible object)



Types of variation

Sounds: Each language uses a particular subset of the sounds in the International Phonetic Alphabet, which represents all the sounds used in all human languages. There's often overlap (ex: "m", "p" are used in many languages), but languages also may make use of the less common sounds.

less common English sounds: "th" [θ], "th" [ð]

less common Navajo sounds: "whispered l", "nasalized a", …

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			