Language Development: Nativist approaches

Micah B. Goldwater, PhD

University of Sydney

Outline

- Universal Grammar and nativist arguments from the complexity of syntax

- Evidence in support of UG from child language production

- Nativist arguments from sensitive periods and children generating language (in NSL)

- Simpler nativism: early abstraction accounts

Our knowledge of syntax is subtle and complex

- Some examples from Jackendoff (2002)

For a different sort of phenomenon, consider the examples in (11). The italicized elements are understood as having a role appropriate to the position marked by *t*. For instance, in (11a), *which movie* is understood as the object of the verb *saw*.

(11) a. *Which movie* does Susan imagine that Sarah saw *t* last night? [*wh*-direct question]
b.   John was wondering *who* Sarah decided she would go to the movies with *t* on Sunday. [Indirect question]
c.   I didn't like the movie *which* you said that, everyone was talking about *t* the other day. [Relative clause]
d.   You may take *whichever sandwich* you find *t* on the table over there. [Free relative]
e.   *That movie*, I wouldn't recommend that anyone consider taking their kids to *t*. [Topicalization]

Our knowledge of syntax is subtle and complex

- Some examples from Jackendoff (2002)

For a different sort of phenomenon, consider the examples in (11). The italicized elements are understood as having a role appropriate to the position marked by *t*. For instance, in (11a), *which movie* is understood as the object of the verb *saw.*

(11) a. *Which movie* does Susan imagine that Sarah saw *t* last night? [*wh*-direct question]
b. John was wondering *who* Sarah decided she would go to the movies with *t* on Sunday. [Indirect question]
c. I didn't like the movie *which* you said that, everyone was talking about *t* the other day. [Relative clause]
d. You may take *whichever sandwich* you find *t* on the table over there. [Free relative]
e. *That movie*, I wouldn't recommend that anyone consider taking their kids to *t*. [Topicalization]

(12) a. *What* did Beth eat peanut butter and *t* for dinner?
b. *Who* does Sam know a girl who is in love with *t*?
c. *Who* does Betty know which professor flunked *t*?
d. *What food* were you never aware of the hypothesis that you shouldn't eat *t*?

Where does knowledge of language come from?

Interview with Steven Pinker

Where does knowledge of language come from?

Noam Chomsky: Reinforcement learning (proposed by behaviorists) and sensorimotor learning mechanisms (as proposed by Piaget) cannot explain the acquisition of human language.

Human language is too computationally complex. Language is composed of abstract syntactic rules that can generate an infinite number of novel sentences, that don't even have to make any sense!

"Colorless green ideas sleep furiously"

Where does knowledge of language come from?

Recursion: Sentences can be embedded within sentences

The man is tall.
The man who likes cats is tall.
The man who likes cats that went to the store yesterday is tall.
The man who likes cats that went to the store yesterday to pick up a new tie to match his shirt is tall.

Chomsky argued you can't get that from Skinner's or Piaget's "general learning mechanisms." Thus, language must have special innate properties.

Only human language shows recursion
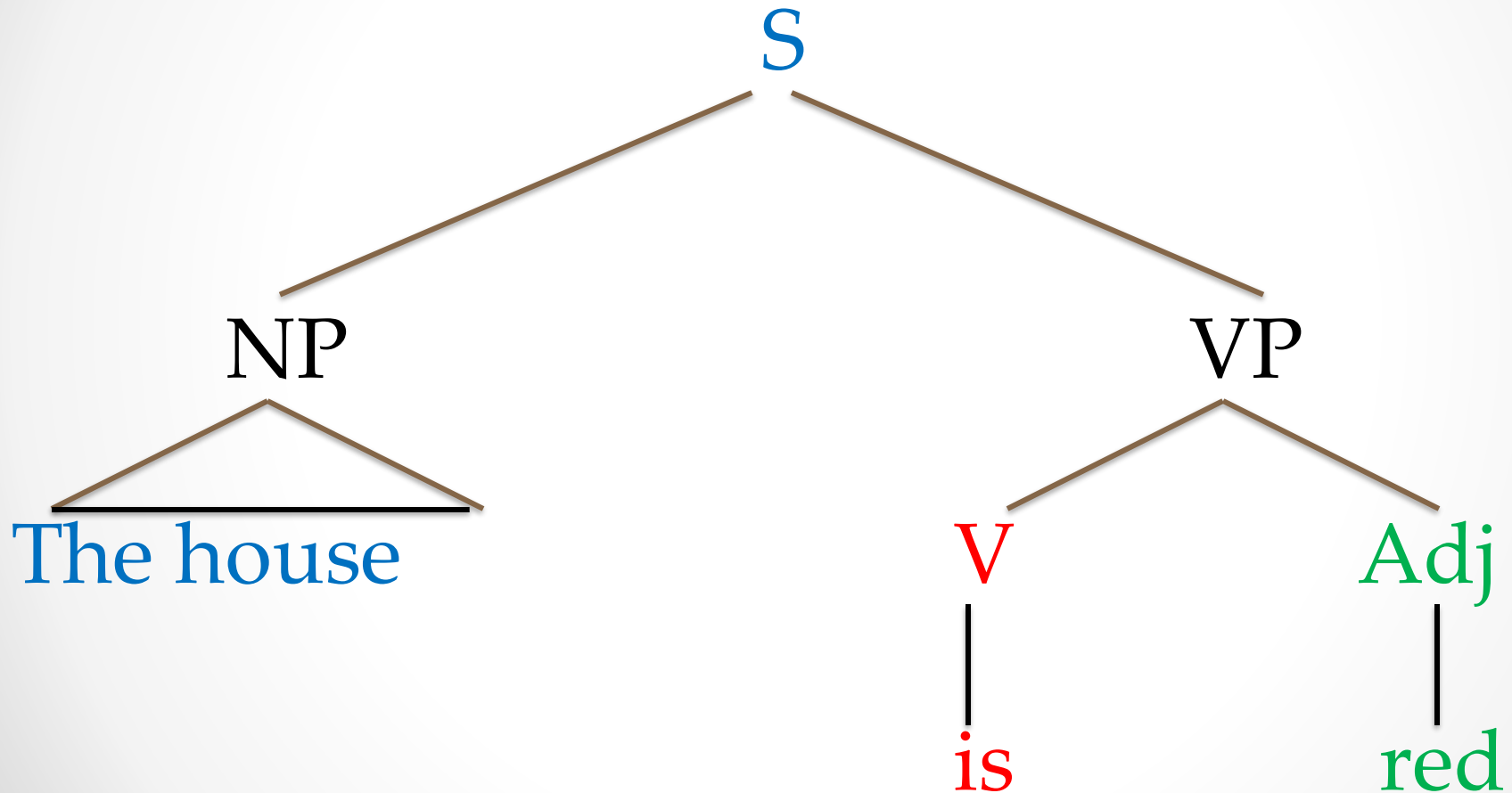(among primates. Some bird song and dolphin communication shows such structure).

Where does knowledge of language come from?

- At some level, clearly in our genetic endowment is the ability to learn language and all it's complexities
  - Also true of everything we learn/do
- So, what specifically is in UG?
- Structured representations with abstract syntactic categories
- For example, that sentences are composed of noun phrases and verb phrases, and verb phrases are composed of verbs and noun phrases
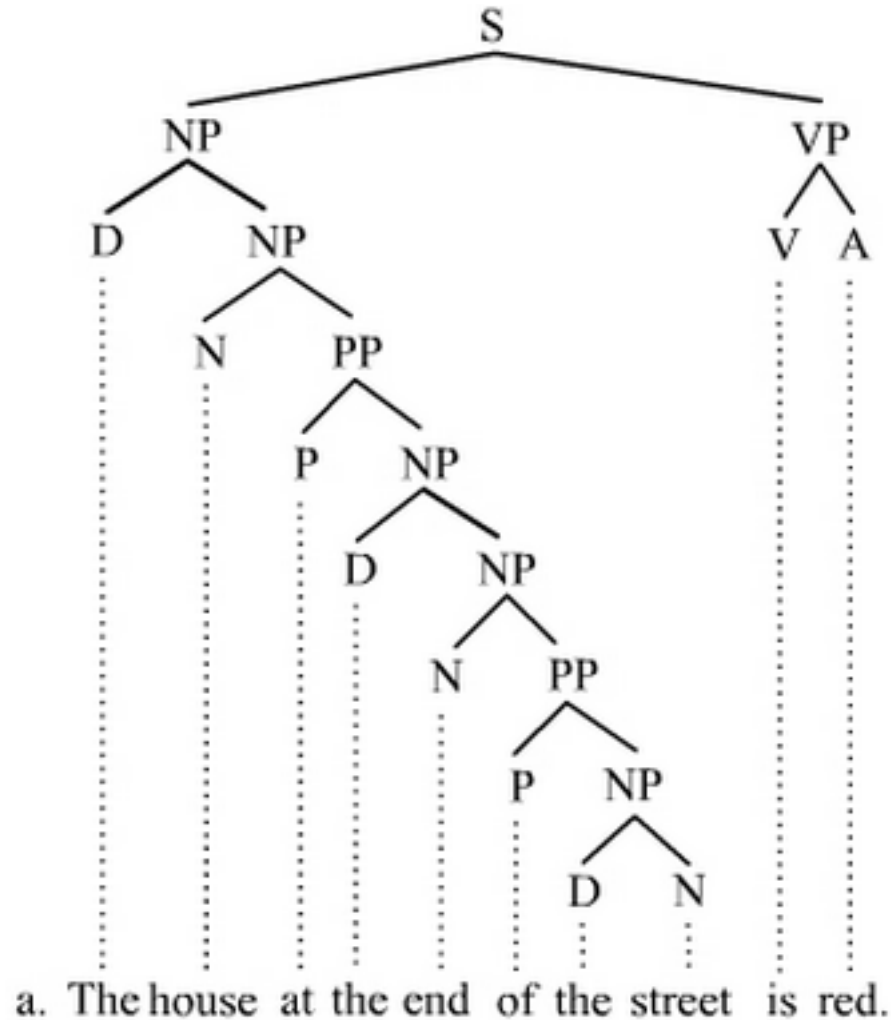- Rules operate over phrases.

What's a phrase?

- A group of words that form a functional/structural unit.

- For example, a noun phrase can consist of a determiner and a noun, such as "the dog" and one can add modifiers such as "the dog by the house."

- Those modifiers can have phrases within them, e.g., "by the house" is a prepositional phrase that has another noun phrase "the house" within it.

- Phrases are hierarchically organized.

Example hierarchically structured syntax tree

Example hierarchically structured syntax tree

a. The house at the end of the street is red.

How can you tell what is in a phrase?

1 example way: pronominal reference:

e.g., "it is red." "it" can take the place of the entire noun phrase regardless of how much stuff is embedded in it, but it can't take less than all of it, e.g., "the it is red" is bad

"A man with dark glasses is following us" →

OK= "He is following us"

Not OK = "He with dark glasses is following us"

"We watched a documentary about cheese making last week" →

OK = "We did that last week"

 Not OK = "We did that about cheese last week"

Where does knowledge of language come from?

- "People attain knowledge of the structure of their language for which no evidence is available in the data to which they are exposed as children." (Hornstein and Lightfoot 1981:9)
- What is innate and what is learned?

Where does knowledge of language come from?

- "People attain knowledge of the structure of their language for which no evidence is available in the data to which they are exposed as children." (Hornstein and Lightfoot 1981:9)
- What is innate and what is learned?
- For example, UG states VP=V+NP but NOT the order
- Children must learn the order of elements in phrases
- This can't be in UG as it's different for different languages:

- English: VP = V then NP (kicked the ball)
- Turkish: VP= NP then V (the ball kicked)

Where does knowledge of language come from?

- "People attain knowledge of the structure of their language for which no evidence is available in the data to which they are exposed as children." (Hornstein and Lightfoot 1981:9)

- What is innate and what is learned?

- For example, UG states VP=V+NP but NOT the order

- Children must learn the order of elements in phrases

- This can't be in UG as it's different for different languages:

- English: VP = V then NP (kicked the ball)

- Turkish: VP= NP then V (the ball kicked)

- This knowledge/learning is syntax specific, similar to core knowledge modules from before.

UG: a test case in child language

- Structural dependence in question formation, "a parade case of an innate constraint" – Stephen Crain.

"The boy is crazy" – declarative sentence

"Is the boy crazy?" – to form a question *is* goes to the start of the sentence.

What is the rule? Could it be, take the first verb and move it to the front? Perhaps as a leaner tracking statistics of word sequences might think?

UG: a test case in child language

- Structural dependence in question formation, "a parade case of an innate constraint" – Stephen Crain.

"The boy is crazy" – declarative sentence

"Is the boy crazy?" – to form a question is goes to the start of the sentence.

What is the rule? Could it be, take the first verb and move it to the front?

NO!

"The boy who is smoking is crazy" becomes

"Is the boy who is smoking crazy?" and not

"Is the boy who smoking is crazy?"

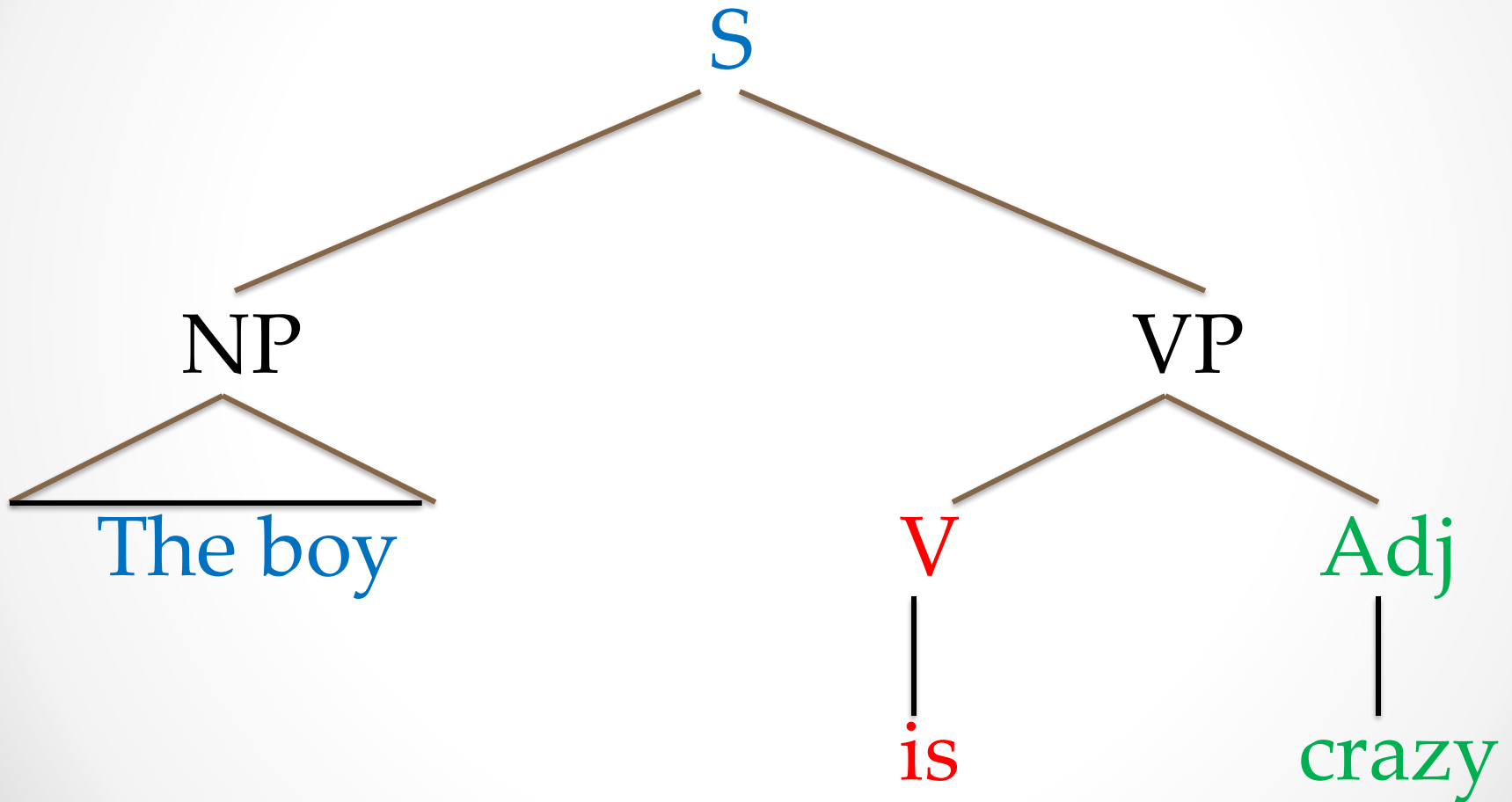# UG: a test case in child language

- Structural dependence in question formation, "a parade case of an innate constraint" – Stephen Crain.
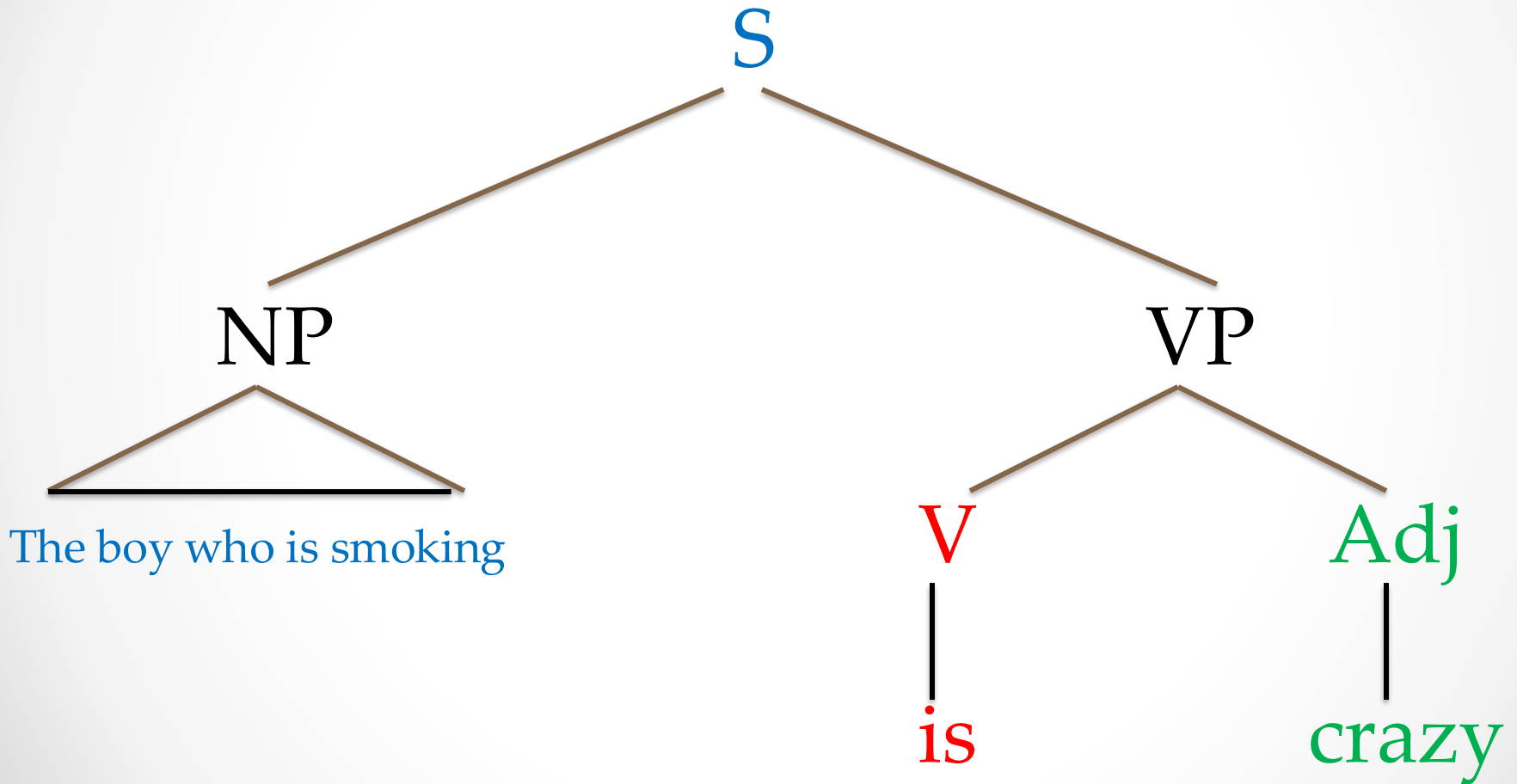
"The boy is crazy"
"The boy who is smoking is crazy"

These two is's have different places in the word order, but the same structural position.

UG: a test case in child language

UG: a test case in child language

UG: a test case

- Structural dependence in question formation, "a parade case of an innate constraint" – Stephen Crain.

"The boy <u>is</u> crazy"
"The boy who is smoking <u>is</u> crazy"

These two is's have different places in the word order, but the same structural position.

Rules of syntax operate over abstract syntactic structure. Children know this innately, and thus **<u>NEVER</u>** make mistakes like "is the boy who smoking is crazy?"

Or so the claim goes…

Is this true?

- YES! No errors of the type

   Is the boy who smoking is crazy?

   ever recorded in naturalistic data
Or in an elicited production study
   (Crain & Nakayama, 1987)

Conclusion: Structure dependence is an
   "innate schematism" (Chomsky, 1971)

Outline

- Nativist arguments from sensitive periods and children generating language (in NSL)
- Simpler nativism: early abstraction accounts
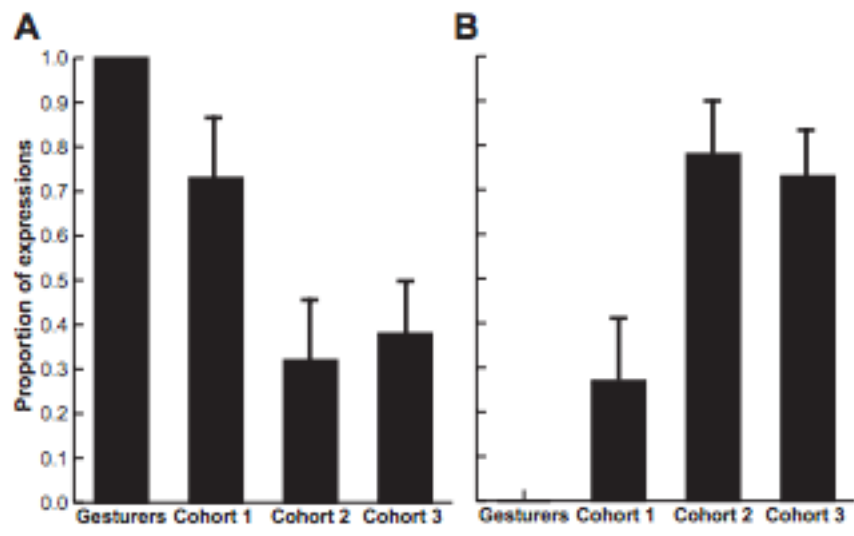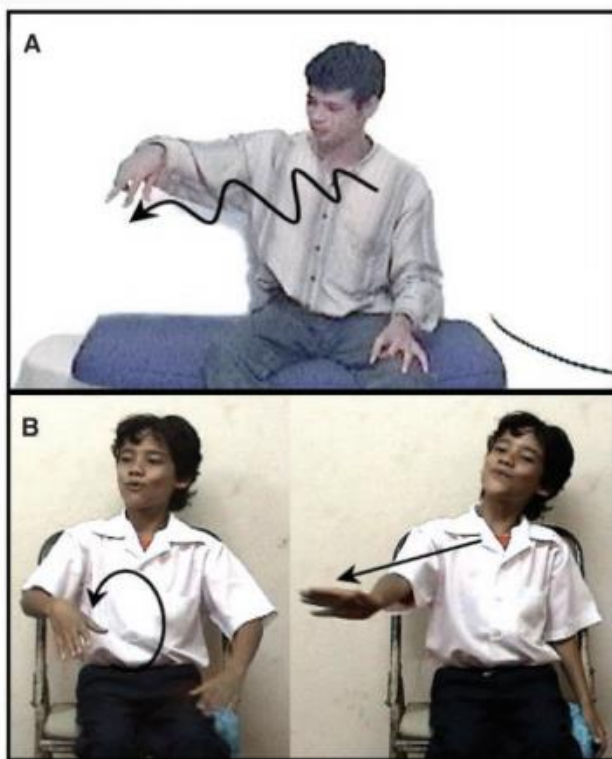
Nicaraguan Sign Language & Sensitive Periods

- Video discussed how if deaf children are not exposed to language before 7, there will be difficulties learning, and if not before puberty: no chance of true native fluency.

- Suggests genetically driven biological maturation constraints.

- Before sensitive period ends, children generate language beyond what they could have possibly been in the input.

- We discussed how language is componential and compositional.

- How quickly did NSL acquire these properties?

Compositional Structure in NSL

- Senghas, Kita, O¨zyurek (2004)
- Language splits meaning up into component parts so they can be recombined to generate more kinds of meanings.

- "The ball rolled down the hill"

- the rolling (manner of motion), and the down (path of motion), are happening simultaneously, but expressed separately in language

- Allows for either to be highlighted or modified or recombined independently

- Does NSL do this?

Compositional Structure in NSL

- Senghas, Kita, O¨zyurek (2004)
- 1st cohort show componential structure somewhat, 2nd cohort show full systematic linguistic structure
  - All children exposed by 6 years of age to NSL. First cohort invented it.
- A: Simultaneous expression. B: Componential expression

Spanish speakers' gestures vs. NSL signers

Outline

- Simpler nativism: early abstraction accounts

Early abstraction accounts

- Many psychologists have abandoned theorizing that much of the rich specific structure (e.g., that VP = V +NP) is part of a language-specific genetic endowment

- But they do posit innate constraints that interpret their language input abstractly and have biases to link syntax and semantics

- That is, at a minimum children interpret language in terms of abstract classes, not on a word-by-word basis

- Specifically, abstract notions of *agents* and *patients*, and general notions of how they are linked to nouns constrains language learning from the start

  o That is, each semantic role needs a noun, and each noun needs a semantic role

Early abstraction accounts

- English has systematic mappings between syntax and semantics.

- Generally, agents are expressed before the verb, and patients afterwards.

The boy broke the vase

I threw the ball

etc.

Children can learn the abstract word order rules from their innate links from event roles to nouns

Early abstraction accounts: Evidence from verb learning

- Innate links between roles & nouns → learn word order rules

- Links between roles & nouns combined with word order rule knowledge → use these cues to lean novel verbs

- Will test this account with how children can use word order rules alone to learn verbs

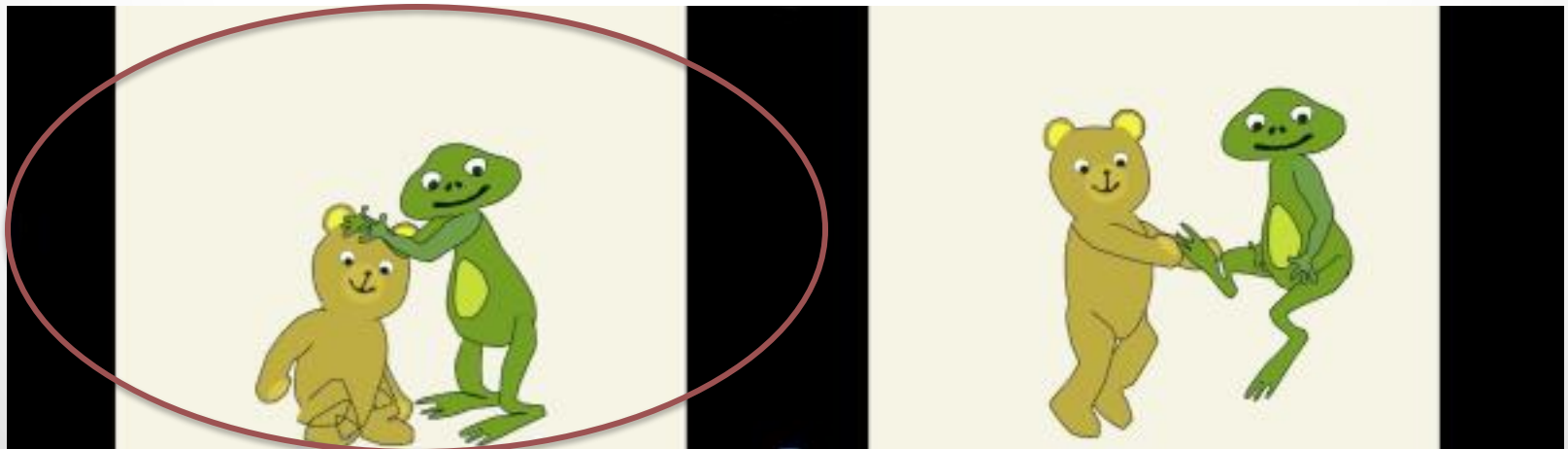# Gertner, Fisher, & Eisengart, 2006: Abstract knowledge in 2 year olds

- "The frog is gorping the bear"
- Gertner et al., both scenes show events with the same two participants, only verb general knowledge can guide children.

Gertner, Fisher, & Eisengart, 2006: Abstract knowledge in 2 year olds

- "The frog is gorping the bear"
- Children look longer to the frog as causal actor
- Must have some sort of links between abstract syntax and semantics to interpret the new verb

Nativist Accounts: Summary

- Chomsky argues: hierarchical phrase structure and recursion is too computationally complex for general learning mechanisms: thus they must be innate

- Evidence that children never violate structural dependence in question formation, supporting SD as part of UG

- Deaf people not exposed to sign language before puberty never learn language while deaf children simply put in the same place invent language!: suggests genetically driven maturational sensitive period

- Early abstraction accounts assume less innate knowledge than UG, but still posit innate biases to represent language structure in abstract terms and link syntax with semantics.

Language Development: Constructionist Approaches

Micah B. Goldwater, PhD

University of Sydney

If interested in trying to make connections across concept and language development:

Goldwater, M. B. (2017). Grammatical constructions as relational categories. *Topics in cognitive science*, 9(3), 776-799.

Ambridge, B. (2020). Against stored abstractions: A radical exemplar model of language acquisition. *First Language*, 40(5-6), 509-559.

Ambridge, B. (2020). Abstractions made of exemplars or 'You're all right, and I've changed my mind': Response to commentators. *First Language*, 40(5-6), 640-659.

Outline

- Contemporary constructionist accounts and more powerful domain-general learning mechanisms

- Statistical learning & structural dependence

- Gradual abstraction

- Response from nativism: Early abstraction accounts, and then response from constructivism

Constructionist Approaches

- Clearly humans are endowed genetically with the ability to learn and create language.

- But that is also true of baseball. Does anyone think there is innate Universal Baseball?

- Do we have a language/syntax specific endowment? Or does language emerge from the interaction of more domain-general cognitive mechanisms?

Constructionist Approaches

- Operating assumptions:

- We do not start out with any endowment of pre-specified abstract grammatical categories and phrase structure rules

- Input is extremely rich, despite claims of the "poverty of the stimulus,"
  - And thus the inability of the child to learn from it

- More powerful learning mechanisms have been explored since Chomsky refuted Skinner & Piaget such as

1. Statistical driven learning via the structure of neural networks (as in Cohen/Smith)

2. Analogical/relational abstraction (as in Gentner).

3. Also social reasoning, see ToM lectures

Statistical Learning

- Hypothesis: Infants and children will track the transitional probabilities between sounds, words, and phrases, and how they are distributed more globally to learn grammatical categories and phrase structure rules.

- First test case: infant word segmentation and transitional probabilities.

Statistical Learning

- What is a transitional probability? The probability that, after a given syllable, a particular syllable will occur
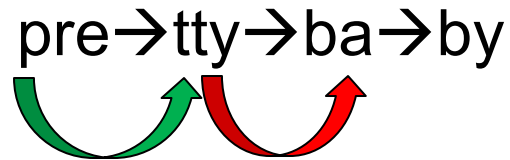   - Or whatever relevant unit, not just syllable

e.g.,

pre→tty   0.5
pre→dict   0.25
pre→cise   0.25

- If, after pre, the next syllable is tty 50% of the time, the TP of pre→tty is 0.5.
- How does this help the learner?

Statistical Learning

- If you hear a string such as prettybaby, you can use TPs to find the words
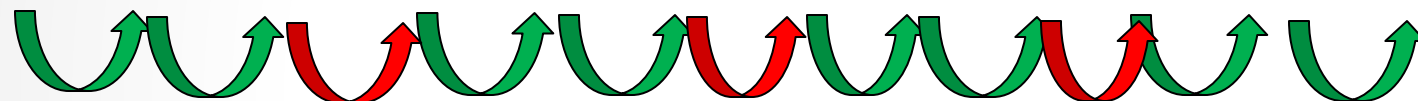
pre→tty→ba→by

TP of pre→tty is high, so don't posit a word boundary

TP of ty→ba is very low (almost zero) so DO posit a word boundary

Statistical Learning

So can children actually do this? Test with "artificial grammar" Saffran, Aslin & Newport (1996): 8m olds

bi-da-ku-pa-do-ti-go-la-bu-bi-da-ku

1.0 1.0 0.33 1.0 1.0 0.33 1.0 1.0 0.33 1.0 1.0

Only cue to word boundaries is the transitional probability information: High between syllables within a "word" (e.g., bi→da); low between syllables across a "word" (ku→pa). No gaps in stream!

Statistical Learning

- Infants listened to this continuous stream of speech sounds for 2 minutes. Then their listening time was measured for:

A) Repetitions of words: "bidaku, bidaku, bidaku"
B) Repetitions of non-word "kupado, kupado, kupado"

- As defined by the transitional probabilities with a ½ second gap between each

- They showed a novelty preference for B.

Statistical Learning

- Mintz 2003

- Grammatical categories are revealed through their distributions, analysing corpora of child-directed speech

- e.g., the ___ (is usually a noun), is__ing (usually a verb), and the "the ___" is often followed by "is __ing"

- Showed infants are sensitive to forming categories of words through shared distribution in artificial grammar

Statistical Learning & Structural Dependence

- The constructionist claim is that the learning of grammatical structures can be driven via these sorts of mechanisms, tracking transitional probabilities and the distributions of words and phrases.

- Can this perspective re-explain previous findings interpreted as evidence for UG, such as the Crain's work on children's questions?

- Reminder:
  - given "The man who is smoking is crazy" children always turn it into a question by respecting phrase structure, as in "Is the man who is smoking crazy?" and not "Is the man who smoking is crazy?"

But….

- Whether or not children have innate knowledge of SD, such errors would not be expected to occur given (implicit) knowledge of the transitional probabilities of particular

  lexical items or categories thereof.

Is the boy who smoking is crazy?

AUX DET NOUN COMP PARTICIPLE AUX ADJ

- That is, no ever says anything like "who smoking" in any context, question or otherwise

(besides auxiliary verb dropping dialects…).

So,

- If children are building up their knowledge of syntax by tracking the statistics/distributions of words and categories of words, they also shouldn't make such errors.
- How can we test the two accounts?
- Ambridge, Rowland, & Pine (2008) came up with a way

Constructionist approach to development of question formation

- Unlike for *is* questions, for *can* questions, a SD error results in a possible word (category) pair

"The boy who can smoke can drive"

"Can the boy who can smoke drive?

SD error "Can the boy who smoke can drive?"

"Who smoke" unlike "who smoking" is a grammatical pair of words in other contexts. e.g., "people who smoke die young"

If children construct rules from complex sentences from simpler smaller units, they might make such errors.

Constructionist approach to development of question formation

- Ambridge et al. (2008) used an elicited production method similar to Crain & Nakayama, 1987
  - gave situations, made 7 year olds ask questions about them
- Unlike with *is* questions, where there are 0 SD errors, with *can* questions, they found 7% SD errors, with children ranging from 0% SD errors up to 43%!

- UG would not predict such word-specific effects. Structural dependence is a universal constraint that applies equally to all words.

Outline

- Gradual abstraction
- Response from nativism: Early abstraction accounts

Gradual Abstraction (Tomasello, 2003)

- Abstract grammatical categories and structures allow for free production of words, independent of how individual words have been used before

- However, children start by only using words in phrases they hear them are used in the input

- Slowly children start noticing commonalities in patterns of words use and generalize individual words to novel contexts
  - Hypothesized mechanism: Gentner's analogical learning

- Children's earliest grammatical structures are based on individual words, and more abstract adult-like grammars are built from there.

- More specifically…

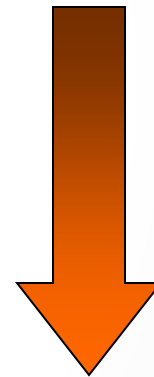# Constructionist approach to development of abstract structure more generally

I kick it, I kick ball

I kick [OBJECT], I hit [OBJ]

I [ACTION] [OBJECT],
Mummy [ACTION] [OBJ]

[SUBJ] [VERB] [OBJ]

Wholly concrete
(substantive)

Wholly abstract
(schematic)

Constructionist approach to development of abstract structure more generally

- Schematization

Break utterances into component parts and generalize across them to form partially-productive, lexically specific schema

Kick it + Kick ball + Kick Mummy = Kick X

- This creates verb specific schemas.
- Then across verb specific schemas, find the commonalities in these schema/utterances to form abstract constructions

(X kick Y, P kiss Q → SUBJ, VERB, OBJ)

What's the evidence for this verb-specific to verb-general pattern?

Schematization Experiments

Novel verb studies: Teach children a novel (made up) verb (e.g.: meeking, tamming) in one construction, and see if they will use it in another

If so, this is evidence that they have a VERB-general rule (e.g., SUBJECT VERB OBJECT)

Schematization Experiments

Tomasello & Brooks (1998)

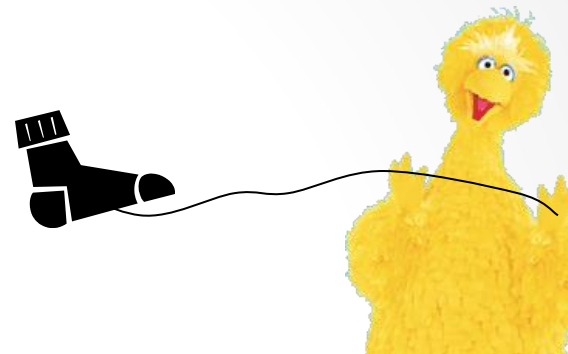Taught children a novel verb in an intransitive [SUBJECT] [VERB] construction:

The sock is tamming

Then tried to elicit use of this verb in a transitive [SUBJECT] [VERB] [OBJECT] construction

Schematization Experiments

What's Big Bird doing?

Answer He's tamming the sock

Only 3/16 children aged 2;0 (7/16 aged 2;6) would use the verb in the non-attested transitive construction. Doing so is no problem for 4 year-olds.

Suggests children formed [THING TAMMING] TAM **verb island**, so could use TAM only in this construction

Schematization Experiments

But when children are taught a new noun (Tomasello & Olguin, 1993) they readily use it in new constructions (so it's not just "shyness")

Experimenter: This is a toma
Child: I want toma, I see toma

Suggests they have small scale schemas such as I want X, I see X **NOT** abstract constructions
such as SUBJECT VERB OBJECT

Constructionist Approach: Summary

- Claim 1: Children can learn syntactic categories through the distributional/statistical patterns of words in their input.

- Claim 2: Children's initial syntactic knowledge is item-based (based on individual words) and abstract constructions are formed slowly by generalizing across such item-based constructions.

Early abstraction accounts

- Many psychologists have abandoned theorizing that much of the rich specific structure (e.g., that VP = V +NP) is part of a language-specific genetic endowment, but they also reject the constructionist notion that early language knowledge is all item-specific

- Specifically, verb-general notions of *agents* and *patients*, and general notions of how they are linked to nouns constrains language learning from the start
  - That is, each semantic role needs a noun, and each noun needs a semantic role

- From there abstract knowledge of word-order sequences is built
  - While word order needs to be learned, they are utilizing abstract verb-general representations from the beginning.

Early abstraction accounts and verb learning

- English has systematic mappings between syntax and semantics.
- Generally, agents become before the verb, and patients come afterwards.

The boy broke the vase

I threw the ball

etc.

If children can use such verb-general cues to assist them in learning verbs, evidence against item-specific syntactic knowledge
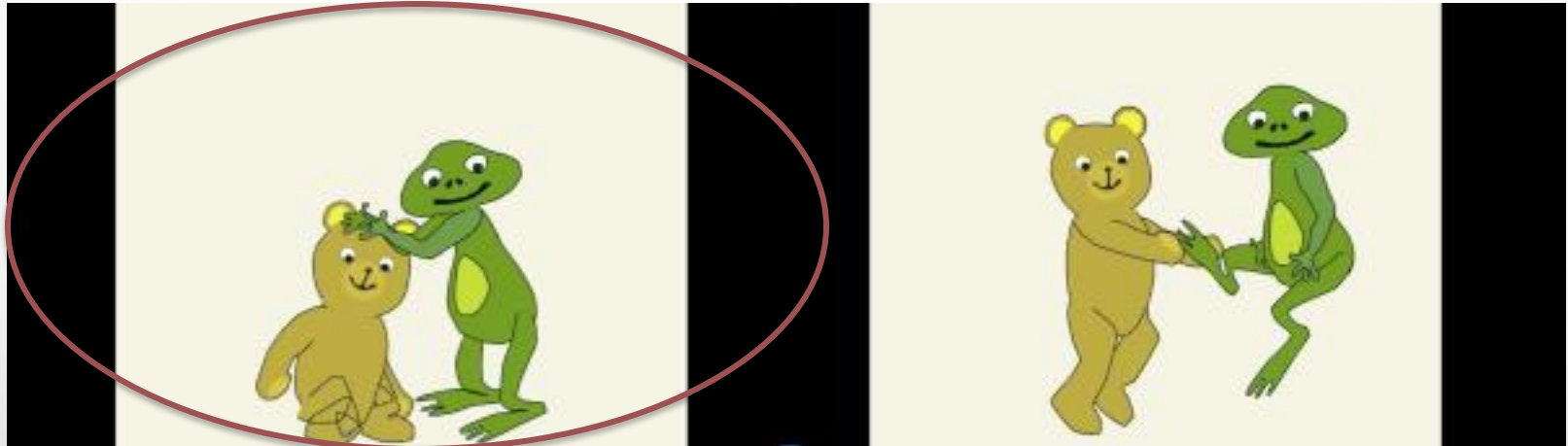
Gertner, Fisher, & Eisengart, 2006: Abstract knowledge in 21 month-olds

- "The frog is gorping the bear"
- Gertner et al., both scenes show events with the same two participants, only verb general knowledge can guide children.

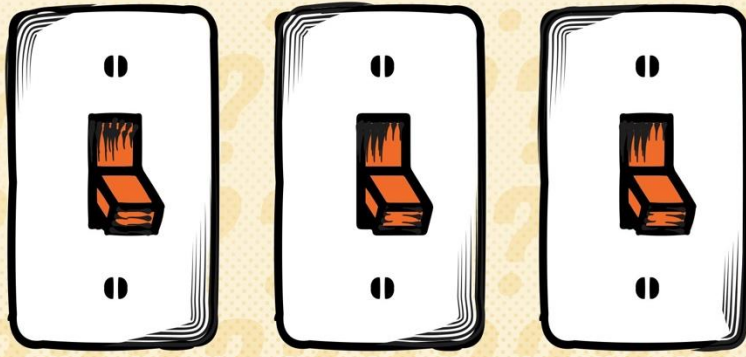# Gertner, Fisher, & Eisengart, 2006: Abstract knowledge in 21 month-olds

- "The frog is gorping the bear"
- Children look longer to the frog as causal actor
- Must have some sort of verb-general knowledge of syntactic structure

Reply from Constructionism

- Dittmar, Abbot-Smith, Lieven, & Tomasello (2008)
- Noticed that in Gertner et al., children were warmed up to the task with familiar verbs and similar kinds of causal events, using transitive syntax and the exact same nouns used at test (e.g., "frog" and "bear")
  - "The frog is washing the bear"
- Dittmar et al. replicated their procedure and added a condition that simply did not use the nouns or transitive syntax as part of the warm up.
  - "This is called washing"
- Then all children had the same test, as in Gertner et al.
- In the latter condition, children did not succeed at test! They required a relevant prior exposure.
- How abstract are their representations?

Is there a constructivist explanation for Sensitive Periods?

65%   25%   10%

Is there a constructivist explanation for Sensitive Periods?

Ramscar, Michael, and Nicole Gitcho. "Developmental change and the nature of learning in childhood." *Trends in cognitive sciences* 11.7 (2007): 274-279.

Elman, J. L. (2001). Connectionism and language acquisition. *Language development: The essential readings*, 295-306.
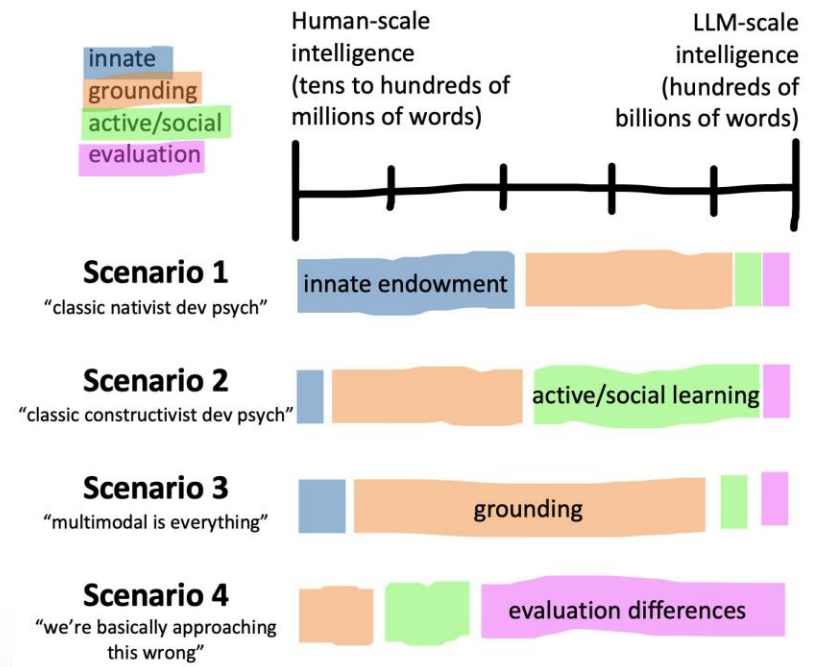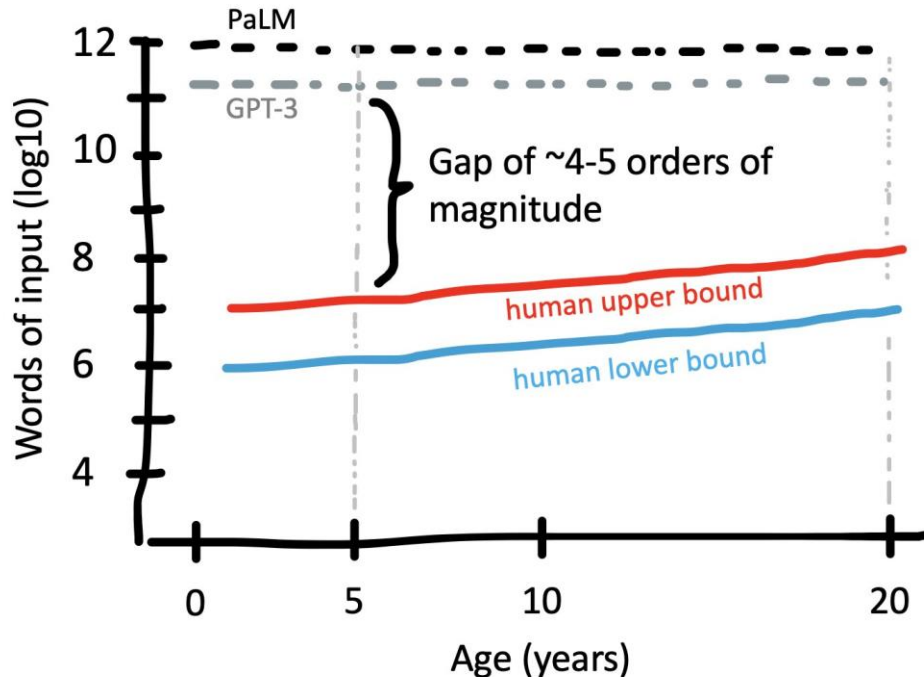
Check out Simon Kirby and who models the evolution of language, e.g.

- Kirby, S., & Hurford, J. (1997). The evolution of incremental learning: language, development and critical periods. *Edinburgh Occasional Papers in Linguistics*, 97(2), 1-33.

- The Language Game: How Improvisation Created Language and Changed the World
  - by Morten H. Christiansen &Nick Chater

What do large language models (e.g., ChatGPT) mean for theories of language development?

- On the one hand, old nativist arguments said that human syntax can not be learned just from example sentences, without innate constraints

- LLM's seem to show that is wrong, but do they learn like children?- Images via Mike Frank

What do large language models (e.g., ChatGPT) mean for theories of language development?

- If interested in more takes: https://lingbuzz.net/lingbuzz/007180

(argues this shows Chomsky is wrong)

- Chomsky thinks LLM's prove nothing about humans

- A longer discussion with Chomsky and others

-

Constructionist Approaches: Summary

- Argues that domain-general mechanisms such as statistical learning and analogical abstraction are more powerful than the GDM's Chomsky first argued against, and can in fact account for language learning.

- Infants can segment words and form grammatical categories via statistics/distributions

- Children do make structural dependence errors when the statistics support such errors, despite UG claims

- Children's early syntactic productions are verb-specific

- Early abstraction accounts (trimmed down nativism) rejects item-specific syntax

- However, constructionist have countered this rejection.

- Sensitive periods are related to domain-general brain
- maturation, not language-module changes