# Language learnability: How do learners evaluate candidate grammars?

Janet Dean Fodor

The Graduate Center, City University of New York

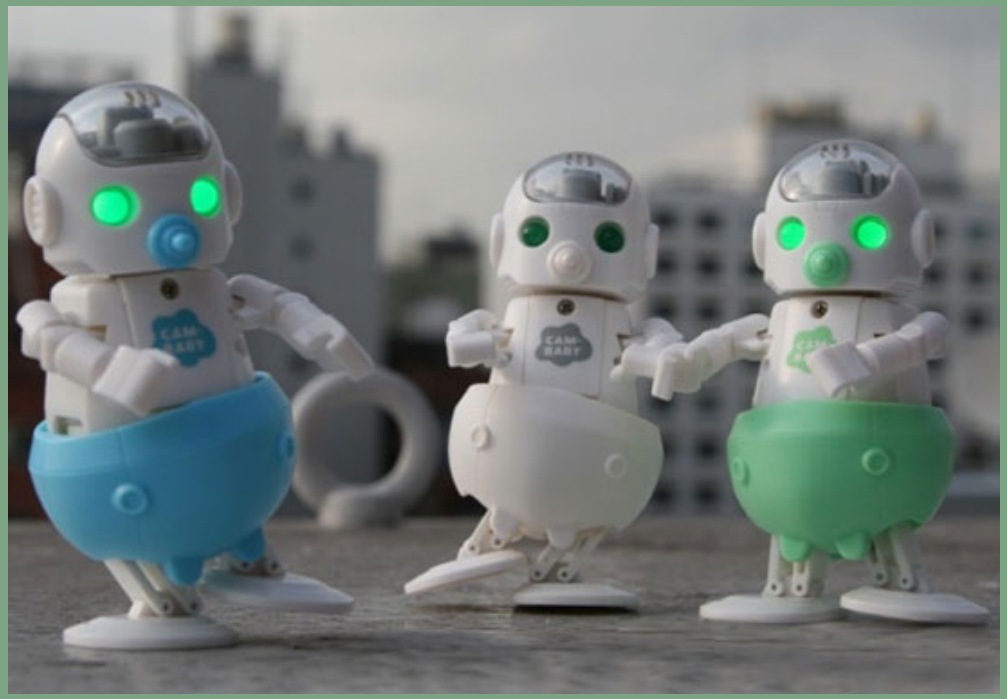
May 2021
Queen Mary University of London

# What's become of the evaluation measure?

- As conceived by Chomsky (1965), an evaluation measure (EM) guides a learner's selection of a grammar when more than one is compatible with the available input.
- In other words, EM resolves ambiguities between the E-language sample that the learner is exposed to and the I-language (grammar) that he or she will develop.
- Our CUNY Computational Language Acquisition Group (CUNY-CoLAG, co-director William Sakas) has been examining the types of ambiguities that learners may confront, and the kinds of strategies they might deploy to choose between them.

# Real children, real language acquisition


# E-children, models of language acquisition



How do children do it?

How could it be done?

# A proposed EM aims to predict the grammar choices children actually make

- "…a theory of linguistic structure that aims for explanatory adequacy must contain…a way of evaluating alternative proposed grammars." (*Aspects,* p.31)
- Chomsky makes it clear that EM is not required to select grammars that are 'really' good in any absolute sense.
- All that's required is…

(a) that EM be shared (universal, innate) in order that learners exposed to comparable input will make comparable grammar choices;

(b) that EM, applied to (typical) input from any human language, favors the very grammar choices that child learners make for that language.

# Is EM still needed, even for P&P? for MP?

- The evaluation measure gets hardly a mention these days.
- It may indeed play a lesser role now than was envisaged back in *Aspects*, because a parametric theory defines a finite and orderly set of grammars for learners to choose between.
- But some sort of EM is almost certainly still needed, because at any stage during learning, there's likely to be more than one set of parameter values compatible with the learner's limited input sample.
- How does the learning system make its choices among all the multiple candidate grammars? ($2^n$ for n binary parameters.)
- Different learning models answer differently, often quite unlike Chomsky's original proposal for a formal ranking of grammar complexity. Emphasis now is on how well they perform.

# Performance-based rankings

- The Triggering Learning Algorithm (TLA, Gibson & Wexler 1994): On encountering a novel input *i* ('novel' = not generated by $G_{current}$), shift to a grammar that differs from $G_{current}$ with respect to the value of any one parameter, if, on subsequent testing, it is found to generate *i*.

  This ranks all grammars that license i and differ by one p-value from $G_{current}$ equally high, and all others totally low.

- The Variational Learner (Yang 2000): Test one grammar at a time. If it succeeds, for each parameter nudge a pointer (in memory) toward the successful p-value. If the grammar fails, nudge the pointers away from those p-values.

  This gives a rich ranking of grammars, by aggregating the success rates of all their P-values. (But can be slow to converge.)

# CUNY-CoLAG's metric: Adopt a grammar that assists the parser

- Use $G_{current}$ in combination with innate sentence processing routines to assign structure to an input sentence.
- That is: build structure over an incoming word, to connect it into the growing parse tree for the sentence. As adults do.
- If $G_{current}$ provides no way to integrate a word into the tree, search UG for a parameter value (in the form of a 'treelet' – examples below) which can fill that specific hole in the tree.
- If that is successful, adopt that parametric treelet into $G_{current}$. (Or: increase its activation level.)
- In case of structural ambiguity, apply Minimal Attachment, Late Closure, etc. Just as adults do.

# We call this "learning-by-parsing"

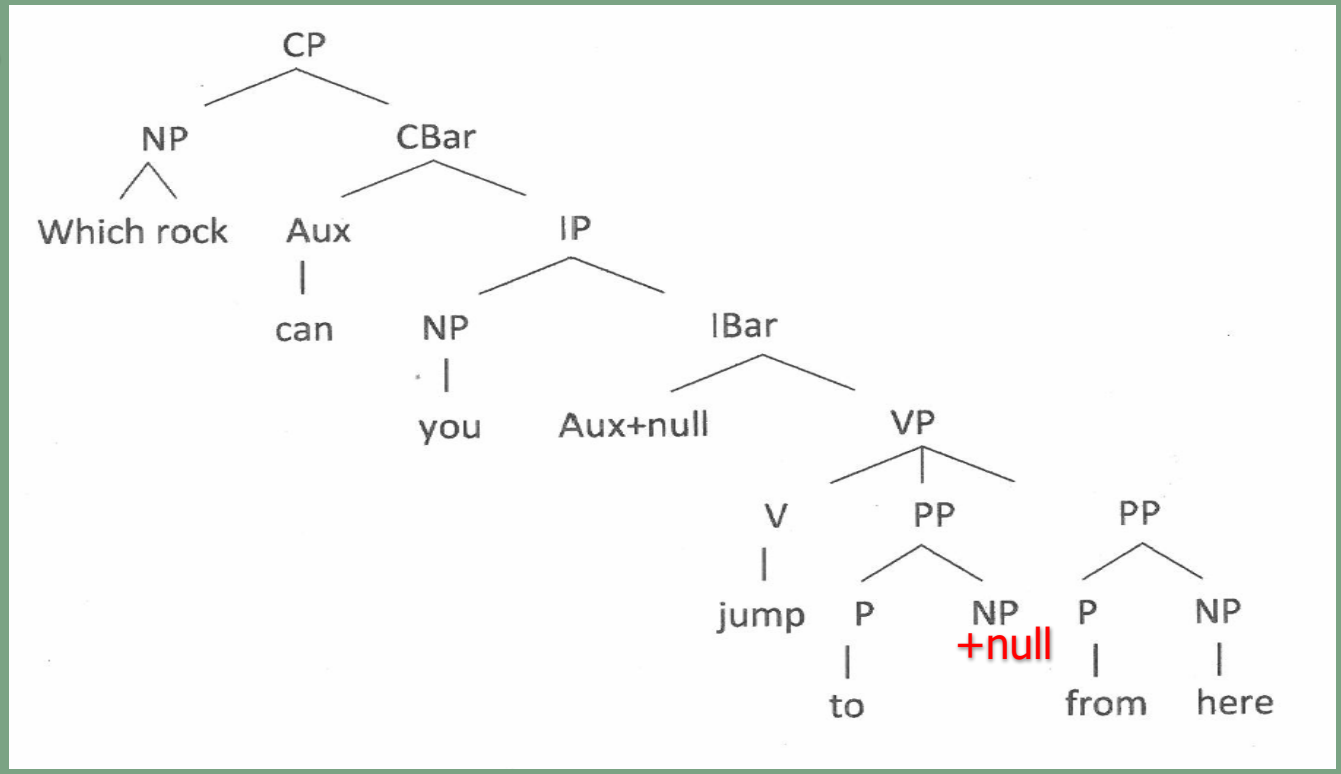
- A child's aim is to understand what people are saying.
- So, just like adults, children try to parse the sentences they hear: Assign structure to word string, → semantic interpretn.
- When a child's $G_{current}$ licenses an input, her parsing routines function exactly as in adult sentence processing.
- When the sentence lies beyond $G_{current}$, the parsing mechanism can process parts of the sentence but not all. It seeks a way to complete the parse tree. (Not just yes/no.)
- To do so, it consults the store of parameter values that UG makes available, seeking one that can solve the problem.
- If a parameter value succeeds in rescuing the parse, that means it's useful! So it is adopted (or strengthened).

# If so – a parameter value must be something the parser can use

- What a parsing mechanism (adult or child) needs is some linkage of syntactic nodes and branches to connect an incoming word into the tree structure currently being built.
- This is why CUNY-CoLAG takes p-values to be UG-specified 'treelets'. (Not switch-settings.)
- A treelet is a sub-structure of larger sentential trees (typically underspecified in some respects).
- Example: a PP node immediately dominating a preposition and a nominal trace. Indicates a positive value for the preposition-stranding parameter (*Who are you talking with now?* vs. **Qui parles-tu avec maintenant?*). <INSERT TREELET>

# A child's sentence processing differs from an adult's, only in the need to draw on new treelets from UG

- Imagine an English learner who already knows wh-movement but not yet preposition stranding (Sugisaki & Snyder, 2006).
- In English, a preposition can have a null (trace) object. Learners will one day discover they need this treelet, to parse with.
- E.g.

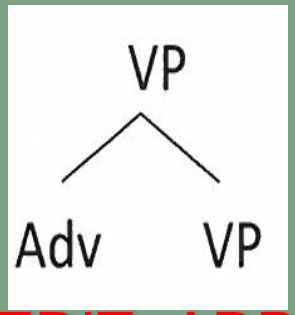



# Learning = patching up parse trees with new treelets

- FR: Pierre compte souvent ses poulets.
  EN: Peter often counts his chickens.
- In FR, the verb moves up out of VP, into TP.
  In EN, the Tense moves down onto the verb in VP.
  (See trees – next slides.)
- Suppose a child is just beginning to acquire adverb placement. She knows what 'souvent' / 'often' means but not how to position it in a sentence.
- The FR learner can parse as far as 'Pierre compte' but is expecting the object right after the verb. So she needs a treelet that lets her attach the Adv *souvent* into the tree between the verb and the object.
- An EN learner expects the verb to follow the subject. So she needs a treelet that lets her attach the Adv *often* into the tree between the subject and the verb.

# Treelets for VP-adverbs

- Both children need a treelet in which the Adv is left-adjoined to the VP (because this adverb precedes the object in both FR & EN).

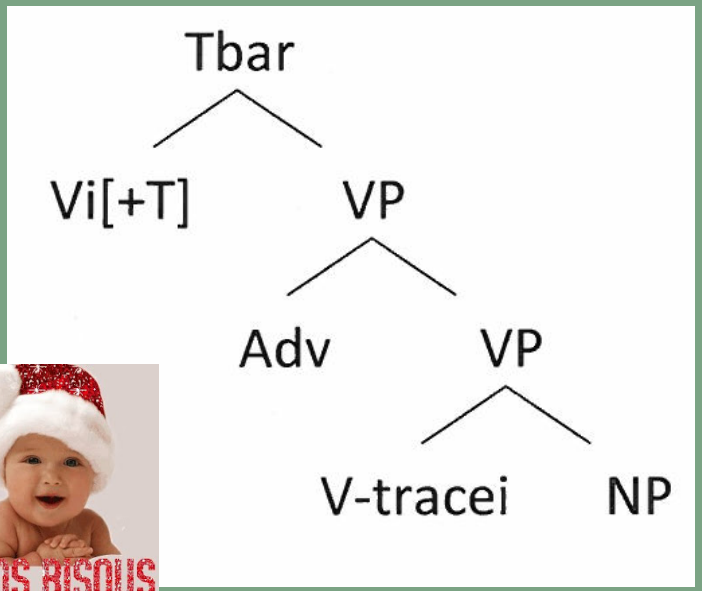



<TO EDIT- ADD NODES>

- But this adjunction treelet must fit inside a larger treelet which makes it possible for the verb to acquire tense even though Adv blocks adjacency to the T head of the clause.
- UG offers two different treelets for achieving this.

# Parametric treelets for V-raising & Tns-lowering

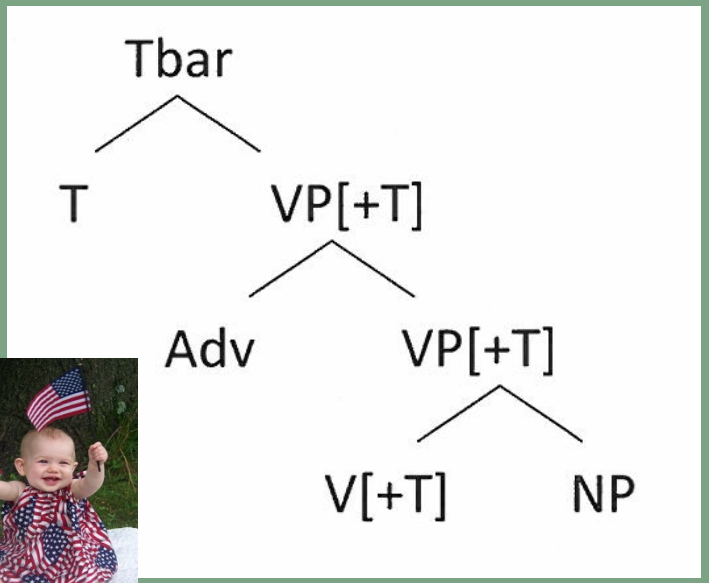
- French: Verb raising







English: Affix-hopping
(= tense lowering)






- The French learner needs a treelet that puts puts ‘souvent’ after ‘compte’. Only the Verb Raising treelet can do that.
- The English learner needs a treelet that puts ‘often’ before ‘counts’. Only the Tns-lowering treelet can do that.

# Merits of the learning-by-parsing approach

- L-by-P makes maximum use of the information the input already contains. A parse tree is created, spotlighted for where new structure is needed.
- Input guidance: The properties of input sentences provide a specific word-by-word guide to the adoption of relevant p-values. No time wasted trying out arbitrarily chosen p-values.
- Unlike trial-and-error learners, which first pick a grammar on some other basis, and then find out whether it succeeds or fails. L-by-P can find a grammar that parses the novel sentence.
- No false adoption of p-values that didn't contribute to the parse.
- So: Learning-by-parsing predicts faster convergence on the target grammar (confirmed in simulation studies, e.g., Fodor & Sakas, 2004) and fewer errors of 'comission' (Snyder, 2007) en route to the target.

# More on: L-by-P is good because...

- The challenge for any learner is to infer I-language structure from E-language word strings. But this is precisely what the sentence parsing mechanism is biologically designed to do, and does every day for the rest of your life.
- The parsing mechanism comes 'for free': it exists in any case, and is generally regarded as innate.
- So L-by-P requires no specifically-designed learning mechanism for language (no LAD) – nothing other than is inherent in the ability to produce and understand language.
- This is compatible with the eliminative ambitions of recent linguistic theory, and evolutionary speculations.
- Also, a potential source of 'third factor' influences, since the parser exhibits economy tendencies, e.g., Minimal Attachment, Minimal Chain Principle; also frequency sensitivity, etc.

# Wow! But is L-by-P's EM the right EM?

- It gives priority to grammars (p-values) that the human language processing routines favor. This increases the chances that all learners with comparable input will make the same grammar choices. So it meets requirement (a). ☺
- Can it also meet requirement (b): that EM should favor the very grammar choices that child learners make?
- Identifying the true EM calls for simulation studies and comparison with child language development, as by Yang.
- But there's also another approach to try:
  Note potential pitfalls that child learners don't fall for.
  The child's EM evidently disfavors the wrong grammar choices in those cases.

  (Where 'wrong' = unlike other child learners, unlike the adult lg.)

# 4 types of input ambiguity

I. Everyday structural ambiguities:
Tickle the bunny with the feather.

II. Missing triggers: A parameter both of whose values lack unambiguous triggers in some languages. What choice to make? Both values can't be default. (Han, Lidz & Musolino 2007)

III. Ambiguity between a subset grammar and a superset grammar. If negative evidence is lacking, it is essential for learners to make subset choices. But how does a learner know which grammars generate subsets of which others?

IV. Ambiguity between peripheral constructions (essentially unconstrained) and the triggers for core grammar parameters. E.g., Have you a dime? *Lost/want you a dime?

# Our working assumption: Little or no parallel parsing

- A learner could avoid ambiguity only if it could detect ambiguity.
- A parallel parsing mechanism could. Compute all possible parse-trees for the sentence, using any UG-available treelets.
  If there's more than one parse, the sentence is p-ambiguous.
- But even adults don't have the capacity to fully parallel parse.☹
- More modestly, within the capacity of a serial parser: A sentence is unambiguous if the parser finds no choice-points in the parse.
- For safety: After a detected choice-point, do not use the remainder of the sentence to set any further parameters.
- But very wasteful! Many inputs would be discarded due to an early choice-point, even if it's only a temporary ambiguity, resolved later. E.g. *Sue doesn't know Bill (is a librarian)*.
- Conclusion: Even if the language domain does contain fully unambiguous triggers, the learner may fail to spot them. ☹

# I. Structural ambiguity

- A parsing ambiguity, but parametrically harmless:
  Put the frog on the napkin in the box. (Trueswell et al. 1999)

  Two correct analyses. In both, *put* has a locative complement, and a PP can modify a noun. (But see below!)

- Parametrically ambiguous, could induce error.
  Puppies love children.

  Two structures with different p-setting implications: +V2 or –V2. An English-learning child who guessed +V2 would overgenerate *Cookies love children.

- This is an error that English learners make rarely if ever. Perhaps it would be blocked by the Minimal Chain Principle.

- Nice! The parser's preferences select the safe learning choice.

# II. Garden paths: Parametrically dangerous, if the correct analysis is overlooked

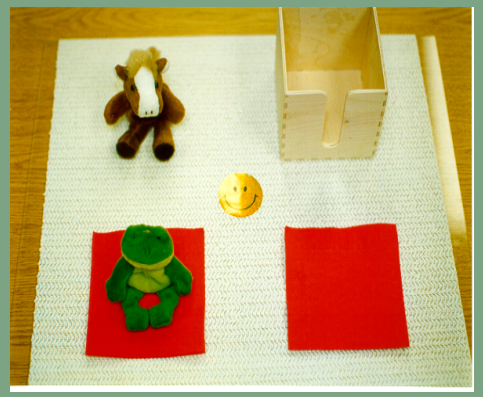
- In garden-path sentences, the parser tends to overlook the correct analysis; true parsing fails. But a learner may tap into UG and find a treelet which 'resolves' the problem – but is wrong for the target language.

- The horse raced past the barn fell.
  Overlooking the relative clause analysis of 'raced' could mis-trigger the P-value for asyndetic coordination (as in Spanish):
  (≡ …past the barn and fell.)

- While Mary was mending the sock fell off her lap.
  The Early Closure analysis (intransitive mending) is often overlooked. A child might 'fix' it by setting +NullSubj.
  While Mary was mending the sock, *pro* fell off her lap.

- Adults know these aren't legitimate parses. A child doesn't.
  <Of course, these examples aren't realistic for child learning.>

# These example sentences are more sophisticated than children ever hear. So → no problem?



- But parametrically dangerous ambiguity can occur even in child-directed speech.
- Put the frog on the napkin in the box. (Trueswell, Sekerina, Hill & Logrip 1999.)
- Adults & 5 year-olds preferentially parse the first PP as the destination of *put*. So *in the box* must modify *napkin*.
- But what if there's no napkin in the box? Adults self-correct → *box* is destination, so *napkin* must now be frog location.
- But 5-year olds are less flexible. Some put frog on napkin and into the box. (This is the "hopping" error in Trueswell et al.)
- A child-size misleading trigger for asyndetic coordination!

# How does EM help children avoid these pitfalls, if they can't reliably recognize ambiguity?

- Fodor (1998) was vaguely optimistic. No permanent damage is done as long as p-setting is gradual and frequency sensitive.
- Not instant triggers! Typically the tricky g-path cases are rare. And a child will eventually encounter unambiguous examples:

  Put the gift from Grandma beside the tree.
  The horse ridden past the barn fell.
  Lena owns the horse raced past the barn.

- Suppose every parametric treelet has a level of activation – the higher the level, the more accessible it is for parsing with.
- Suppose a treelet's activation level starts low and is slightly increased every time it is employed in a successful parse.
  Just like the activation level of a lexical entry.
- Then a wrong treelet due to an occasional mis-parse will gain little activation, while the correct treelet, encountered more often, will gain more robustly and will eventually outweigh it. <Needs testing!>

# III. Subset-superset ambiguities

- If negative evidence is lacking, it is essential that learners make subset choices. Does the EM enforce this?
- Even if EM demands subset choices, how would a learner know which grammars generate subsets of which others?
- If the ambiguity is between two values of the same parameter (e.g., obligatory vs. optional topicalization), UG could designate the subset value as the default. (Manzini & Wexler 1987)
- How? Suppose the activation level of the default treelet is higher than that of the marked value, at birth. So it will be used when either would do, so it will grow steadily stronger.
- OK. But is that just a happy accident?? For all subset-superset parameters!!
- No. But if the superset value were the stronger one at birth, that parameter would never be active – no problem!

# IV. Between-parameter subset-superset ambiguities

- When negative evidence is lacking, learners must make subset choices. But can the EM enforce this if a subset-superset relation holds between two or more different parameters?
- E.g., XP-scrambling (including wh-XPs) yields a superset of wh-movement.

  So EM would prioritize wh-movement over scrambling - OK.
- However, even in our simplified CoLAG language domain, over 42% of subset relations between grammars are not predictable from the subset / superset values of any single parameter.
- How could EM cope with these between-parameter ambiguities?
- Unclear! If no reliable strategy, this could result in permanently inconsistent p-values within the language community! ☹
- But maybe: the parser's own innate preferences just happen to favor the subset treelet. Could be! (See next.)

# Example from our CoLAG language domain

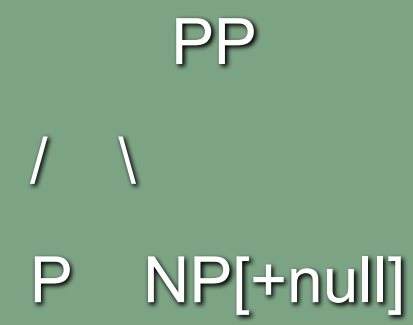
- Our CoLAG constructed language domain has the parameters:
  +/-Optional Topic  +/-Null Topic
- Each is itself a subset-superset parameter. The + value yields a superset. So the – value should be the default.
- A sentence with no overt topic (e.g., verb-initial) could be generated by the positive value of either parameter. Learner must choose between grammars +Opt –Null and –Opt +Null.

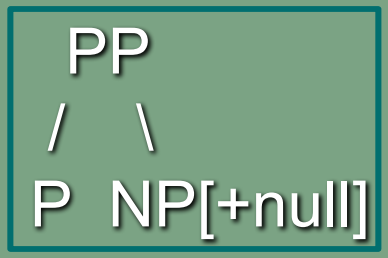
  (A constraint in the CoLAG domain forbids both to be positive; a language is either subject-oriented or topic-oriented.)
- +Null yields a superset of +Opt, because +Null permits 'omission' of obligatory items while +Opt does not.
- So learner must favor +Opt unless/until finds clear evidence for +Null. +OptTop is favored by parsing strategies: It allows no topic → preferred by Minimal Chain Principle & Avoid EC. ☺

# So now – what (if anything) is the evaluation measure?

- Broad current consensus: EM is not a matter of innate notational conventions that favor simple over complex grammar rules, as in *Aspects*. (That doesn't fit with parameter-setting.)
- Nevertheless, learners face innumerable choices, and they (eventually) choose correctly = the same way as each other. That's all that's required!
- To model this, current psycho-computational models of syntax learning favor procedural evaluation measures, based on how well they perform in licensing input sentences.
- Here I have outlined one such model. Because it takes parameter values to be 'treelets', the preferences of the parsing mechanism (presumed innate) will choose between grammars.
- Simply: A child's brain is not in the business of evaluating grammars, but is just doing its best to parse and comprehend what people are saying.

# Thank You


PP
/ \
P NP[+null]


PP
/ \
P NP[+null]