Modelling and Learning Adjuncts

Meaghan Fowlie
mfowlie@ucla.edu
http://meaghanfowlie.com

UCLA/McGill

McGill Linguistics Colloquium
November 20, 2015

Human language acquisition

Learning

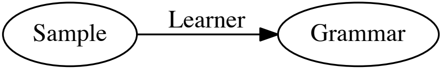
Blah blah...

Blah blah...

Photograph by Andrew Hetherington, Scientific American
July 20 2011
http://www.scientificamerican.com/article.cfm?id=benasich-baby-brains-signal-later-language-problems

Overview

How do people learn adjuncts?



1. **Grammars**: What would a grammar look like that modelled adjuncts as adjuncts, yet accounted for ordering?
2. **Models**: How do formal models of language learning learn properties of adjuncts?
3. **People**: How do people learn properties of adjuncts?
4. **Birdsong**: What do birds generate?

Adjuncts

Generally adjectives, adverbs, prepositional phrases

(1) a. My love is like a rose.
    b. My love is like a red red rose.

(2) a. I'm tired!
    b. I'm really really really really tired!

(3) He suddenly (*suddenly suddenly) smiled.

Properties of adjuncts to be captured by a grammar

(4)

a. The (bad) wolf — *optional*

b. The bad wolf — *transparent to selection*

c. The big bad wolf

d. *The bad big wolf — *strictly ordered*

e. The Alliance officer shot Kaylee in the cargo hold with a gun

f. The Alliance officer shot Kaylee with a gun in the cargo hold

Unordered

g. [bright blue] balloon — *Adjuncts of adjuncts*

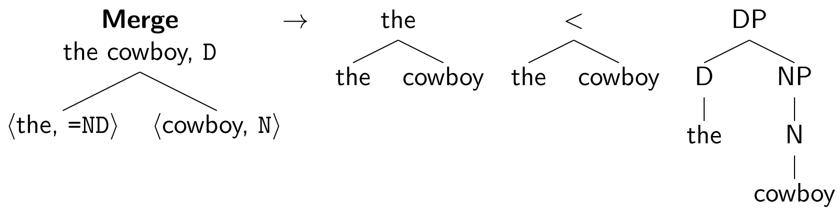
h. Kaylee is clever. — *Selectable category*

Minimalist Grammars (Stabler, 1997a)

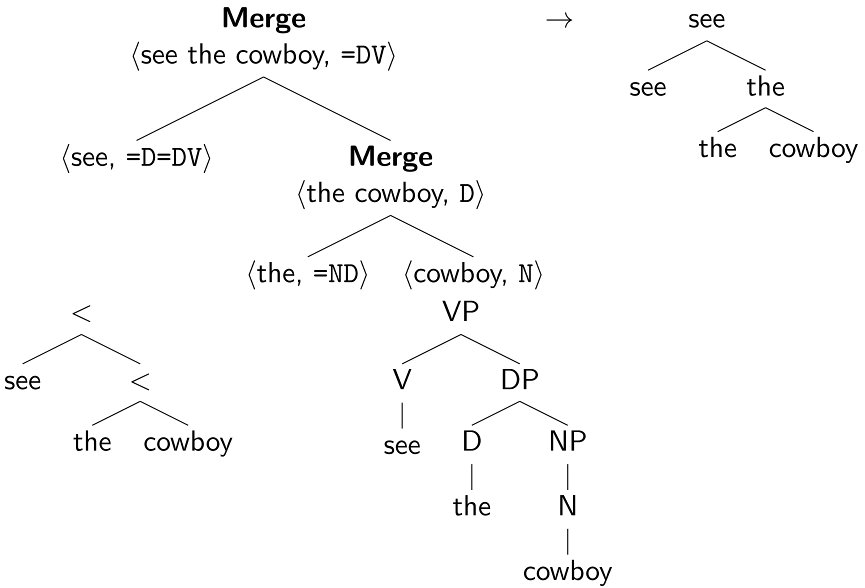
Formalisation of Chomsky (1993) etc's *Minimalist Program*

- Features on Lexical Items drive the derivation via **Merge** and **Move**
- **Features**: **sel** (for **Merge**): `=X` (positive), `X` (negative)

Example: Merge

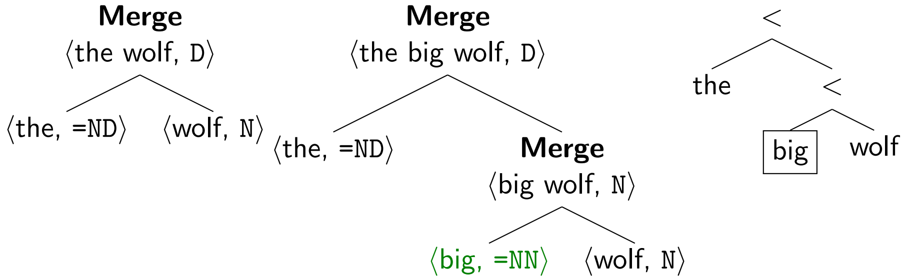


Example: Merge



Traditional MG/Categorial Grammar approach

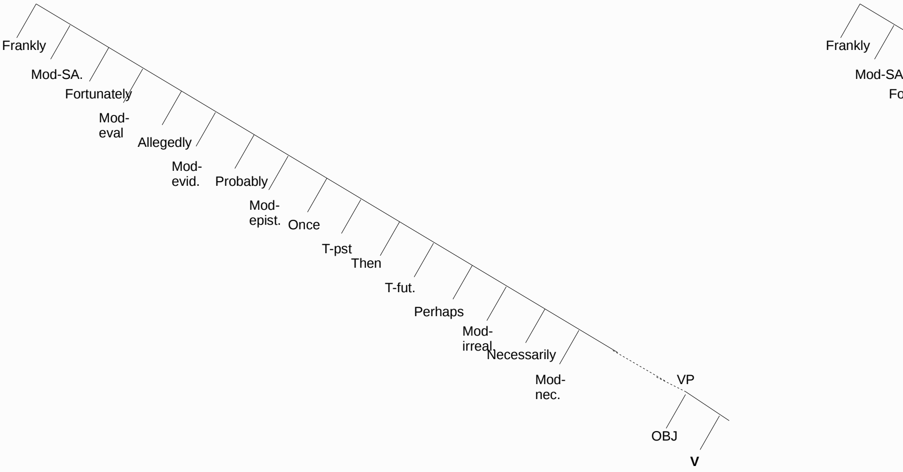
- X-Modifier features: (Categorial Grammar: X/X or X\ X) =XX; Verbal modifier: =VV; Nominal modifier: =NN etc.
- ✓Optionality
- ✗ Ordering
- ✗ Transparent to selection



Traditional MG/Categorial Grammar approach

	Trad. (`=XX`)	Cart. (`=A`$_5$`A`$_6$)	MGAs (`[X,i,j]`)
Optionality	✓		
Selector selects expected category	✓		
Adjunct does not become head	✗		
Unordered adjuncts possible	✓		
Ordered adjuncts possible	✗		
Adjuncts of adjuncts	✗		
Selectability	✗		

Cartography: adverbs

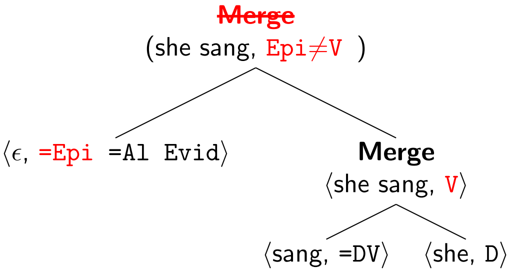


Problem

Allegedly, she sang

Lexicon:

- (Allegedly, `Al`)
- (ϵ, `=Epi =Al Evid`)
- (probably, `Pr`)
- (ϵ, $\texttt{=T}_{\texttt{pst}}$ `=Pr` $\texttt{T}_{\texttt{fut}}$)
- ...
- (ϵ, `=V =Compl` $\texttt{Asp}_{\texttt{compl}}$)
- (she, `D`)
- (sang, `=D V`)



Solution: silent, meaningless LIs

Lexicon:

- (allegedly, $[\![\text{allegedly}]\!]$, $\mathtt{Al}$)
- (ϵ, $[\![\text{evid}]\!]$, $\mathtt{=Epi\ =Al\ Evid}$)
- (ϵ, **id**, $\mathtt{=Epi\ Evid}$)
- (ϵ, $[\![\text{prob}]\!]$, $\mathtt{=T_{pst}=Prob\ Epi}$)
- (ϵ, **id**, $\mathtt{=T_{pst}Epi}$)
- (ϵ, $[\![\text{past}]\!]$, $\mathtt{=T_{fut}T_{pst}}$)
- (ϵ, **id**, $\mathtt{=T_{fut}T_{pst}}$)
- ...
- (ϵ, $[\![\text{compl}]\!]$, $\mathtt{=V=Compl\ Asp_{Compl}}$)
- (ϵ, **id**, $\mathtt{=V\ Asp_{Compl}}$)
- (she, $\mathtt{D}$)
- (sang, $\mathtt{=D\ V}$)

Cartography – Properties

	Trad. (=`X``X`)	Cart. (=$\mathtt{A}_5\mathtt{A}_6$)	MGAs (`[X,i,j]`)
Optionality	✓	✗	
Selector selects expected category	✓	✗	
Adjunct does not become head	✗	✗	
Unordered adjuncts possible	✓	✗	
Ordered adjuncts possible	✗	✓	
Adjuncts of adjuncts	✗	✗	
Selectability	✗	✗	

Cartography: summary

If we want to keep:

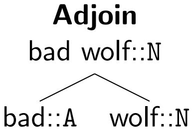
- Merge driven by features
- Adjunct ordering is syntactic
- Just Merge and Move

Then we need:
Silent, meaningless, non-specifier-selecting versions of each functional head, yielding a full Cinque hierarchy in every sentence

Minimalist Grammars with Adjunction (MGAs)

Proposal

Adjoin is optional $\rightarrow$ Add an optional operation **Adjoin** that applies to full phrases. Resulting phrase has the category of the adjoined-to phrase.



MGAs

Wait! Adjunction doesn't occur between just any to phrases $\rightarrow$ add to the grammar a set of adjuncts for each category

$\mathbf{Ad} : \mathbf{sel} \rightarrow \mathcal{P}(\mathbf{sel})$

eg $\mathbf{Ad}(\mathtt{N}) = \{\mathtt{A},\mathtt{P},\mathtt{C}\}$

MGAs

Adjoin

bad wolf::$\mathtt{N}$

bad::$\mathtt{A}$ wolf::$\mathtt{N}$

- Optional
- Category-preserving
- Applies to complete phrases
- Specifies which phrases can adjoin to which phrases
- Adjuncts have their own categories ($\rightarrow$ selectable, adjoin-able, intuitive)

MGAs

Wait! What about adjunct ordering?

Add indicies to track hierarchy level of most recent adjunct.

big::$\mathtt{A}$ $\rightarrow$big::$\mathtt{[A,5,0]}$

bad::$\mathtt{A}$ $\rightarrow$bad::$\mathtt{[A,3,0]]}$

Indices stand for ordered semantic classes of adjuncts

Example

$$[\boxed{\mathtt{X},\mathtt{i}},\boxed{\mathtt{n}}]$$

$$[\boxed{\mathtt{X},\mathtt{i}},\mathtt{j}] \qquad [\mathtt{Y},\boxed{\mathtt{n}},\mathtt{m}]$$

bad wolf: $[\boxed{\mathtt{N},\mathtt{0}},\boxed{\mathtt{3}}]$

wolf:: $[\boxed{\mathtt{N},\mathtt{0}},\mathtt{0}]$ bad:: $[\mathtt{A},\boxed{\mathtt{3}},\mathtt{0}]$

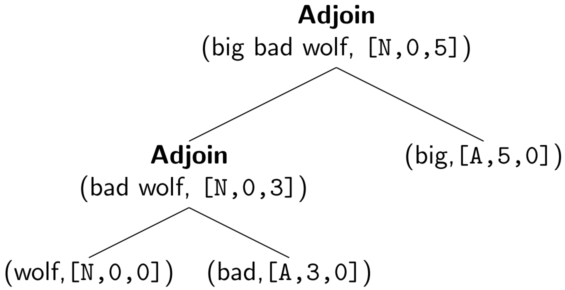
Adjoin example

big bad wolf

ad(N) = {A, P, C}

Lexicon:

- ⟨bad, [A,3,0]⟩
- ⟨big, [A,5,0]⟩
- ⟨the, =N[D,0,0]⟩
- ⟨wolf, [N,0,0]⟩



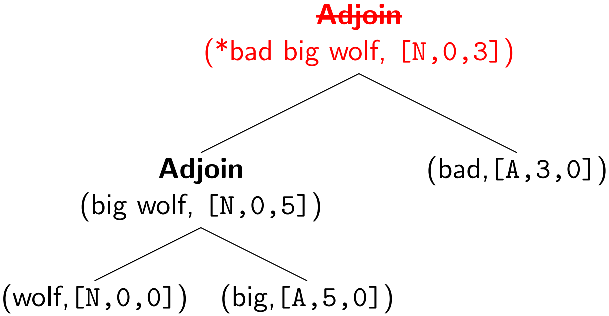
Failed example: bad adjunct order

*The bad big wolf

ad(N) = {A, P, C}

Lexicon:

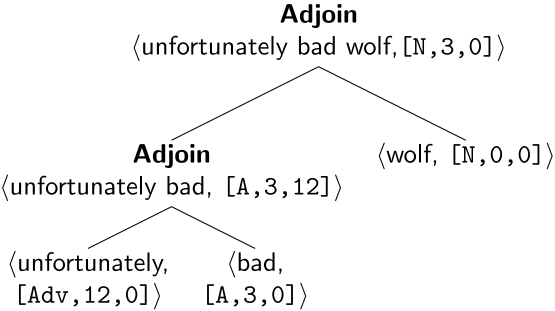
- ⟨bad, [A,3,0]⟩,
- ⟨big, [A,5,0]⟩,
- ⟨the, =N[D,0,0]⟩,
- ⟨wolf, [N,0,0]⟩,



Adjuncts of adjuncts

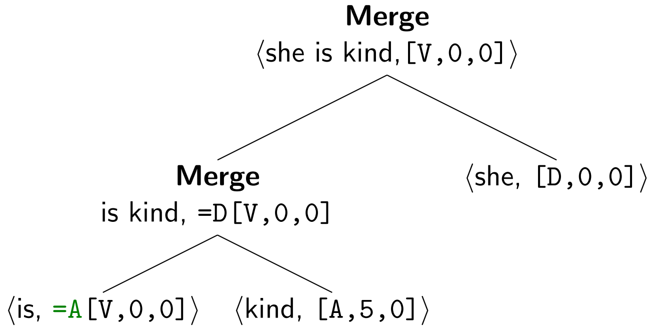
ad(N) = {A,P,C}
ad(V)={Adv,P,C}
ad(A)={Adv}

- ⟨frankly, [Adv,12,0]⟩
- ⟨unfortunately, [Adv,11,0]⟩
- ⟨allegedly, [Adv,10,0]⟩
- ⟨bad, [A,3,0]⟩
- ⟨wolf, [N,0,0]⟩



Selecting adjuncts

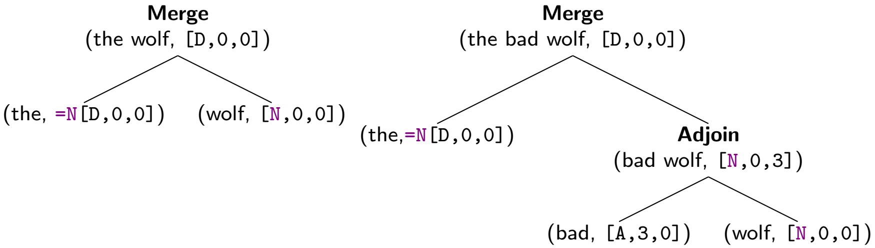
She is kind



MGA – properties

Optional

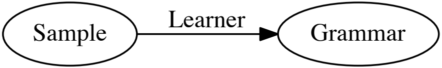
Transparent to selection



Comparison

	Trad. (=XX)	Cart. (=A_5A_6)	MGAs ([X,i,j])
Optionality	✓	✗	✓
Selector selects expected category	✓	✗	✓
Adjunct does not become head	✗	✗	✓
Unordered adjuncts possible	✓	✗	(✓)
Ordered adjuncts possible	✗	✓	✓
Adjuncts of adjuncts	✗	✗	✓
Selectability	✗	✗	✓

Learnability



1. **Grammars**: What would a grammar look like that modelled adjuncts as adjuncts, yet accounted for ordering?
2. **Models**: How do formal models of language learning learn properties of adjuncts?
3. **People**: How do people learn properties of adjuncts?
4. **Birdsong**: What do birds generate?

Learnability

A very weak claim *For some definition of "learn" and some definition of "language", humans learn language*

Learnability

Definition (Language)

A set of sequences of things, with the "things" taken from a finite set

eg: words are built out of phonemes $\rightarrow$language $=$ the words
eg: sentences are built of out morphemes $\rightarrow$language $=$ the sentences

Definition (Learner)

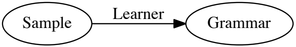
A function from samples from L to grammars

Definition (Learn)

A learner learns a class of languages if it distinguishes them from each other

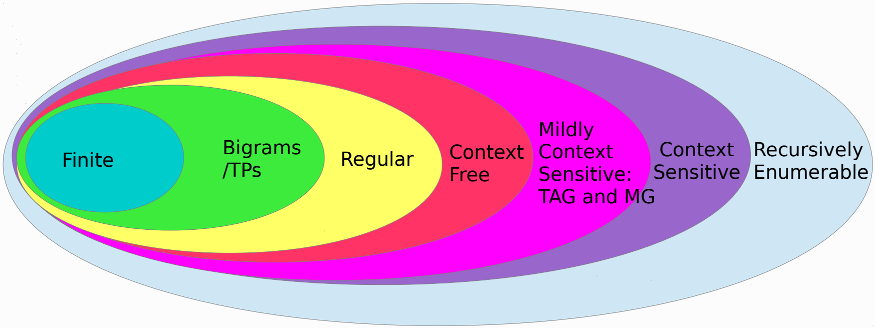
Note: semantics does not factor into this definition of learning

Learnability

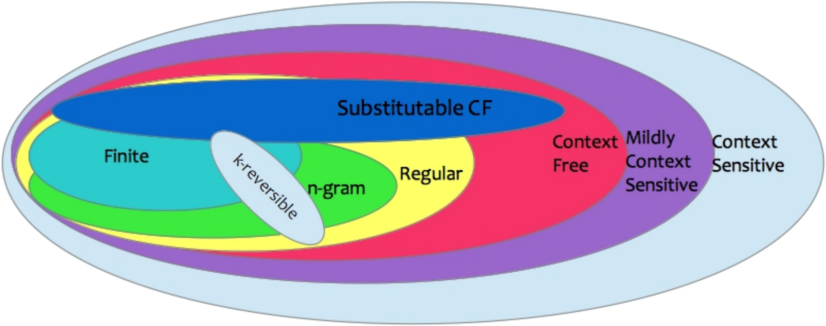


- A learner always guesses a grammar from the class of languages it learns
- So if you gave samples of English to a Substitutable CF learner, the learner will guess a Substitutable language *even though English isn't a Substitutable CF langiuge*
- In this case, we say it failed to learn the target language

Chomsky hierarchy



Chomsky hierarchy



Optionality and repetition

Definition (Optional)

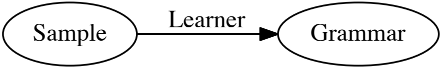
$x \in \Sigma^*$ is *optional in context u,v* iff $uv \in L$ and $uxv \in L$

Definition (Repeatable)

$x \in \Sigma^*$ is *repeatable in context u,v* iff $ux^+v \subseteq L$

(5) Mary sniffed the (red red red) rose

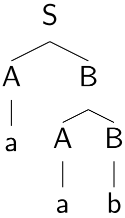
We say *red* is repeatable and optional in context (Mary sniffed the, rose)



1. **Grammars**: What would a grammar look like that modelled adjuncts as adjuncts, yet accounted for ordering?
2. **Models**: How do formal models of language learning learn properties of adjuncts?
3. **People**: How do people learn properties of adjuncts?
4. **Birdsong**: What do birds generate?

Context Free Grammars

- LHS $\rightarrow$RHS
- S$\rightarrow$A B
- A $\rightarrow$a
- B $\rightarrow$A B
- B$\rightarrow$b



Substitutable Context Free learner

- Learnable!

Definition (Substitutable context free language)

L is SCF iff for all $u, v, s, t, x_1, x_2 \in \Sigma^*$, if
$ux_1v \in L$ and
$ux_2v \in L$ and
$sx_1t \in L$ then
$sx_2t \in L$
i.e if two strings share one **context**, they share all contexts

Substitutable Context Free learner

Learning algorithm:

1. create a graph of substitution classes
2. create a grammar from that graph.

Definition (Substitution Graph)

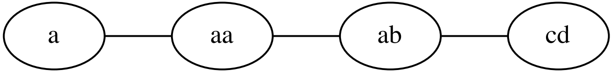
Given a finite sample S, define the substitution graph $SG(S) = \langle V, E \rangle$ as follows:
$V =$ the set of substrings of the sample

Nodes $= V$

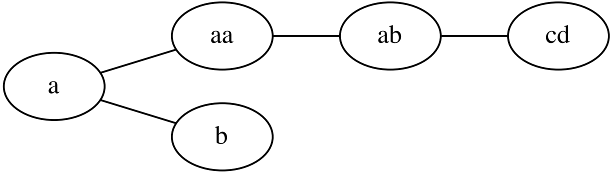
Edges connect substrings that share a context

Substitutable Context Free learner

Sample=$\{a, aa, ab, cd\}$. $V = \{a, b, aa, ab, cd, c, d\}$. a, aa, ab, cd share context (ϵ, ϵ) so we have

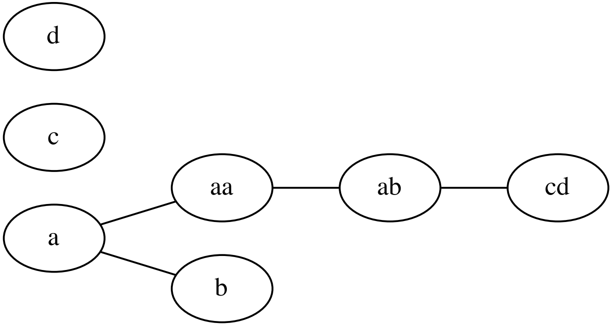


a, b share (a, ϵ), so we extend the graph:

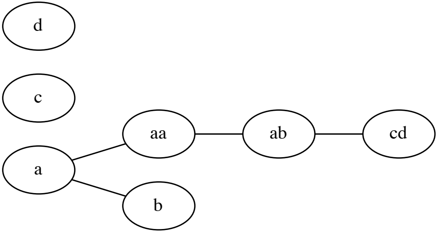


Substitutable Context Free learner

c is just in (ϵ, d) and d is just in (c, ϵ) so they don't connect to anything.



Grammar →Graph Algorithm



$V = \{a, b, c, d, aa, ab, cd\}$

Categories: [a], [c], [d]

[a] →a

[a] →b

[c] →c

[d] →d

[a] →[a] [a]

[a] →[c] [d]

Substitutable Context Free learner

Theorem (Optionality → Repetition)

Let $u, v, x \in \Sigma^*$ and $uv, uxv \in L$ Then $ux^*v \subseteq L(G_i)$.

$yuv \quad yuxv$

$yuxv \quad \rightarrow yuxxv$

Substitutable Context Free learner

Theorem (Repetition $\rightarrow$ Optionality)

Let $u, v, x \in \Sigma^*$ *and* $ux^n v, ux^{n+1} v \in T[i]$ *Then* $uv \subseteq L(G_i)$.

$yuxv \quad yuxxv$

$yuxv \quad \rightarrow yuv$

Substitutable CF – summary

- repetition $\leftrightarrow$ optionality
- one repetition $\rightarrow$indefinite repetition

HL is not substitutable CF

(6)
a. I hear John slept
b. I hear Mary slept
c. I hear Mary was kicking herself
d. *I hear John was kicking herself

CFG$_{F,K}$

Context-free languages with the finite kernel and finite context properties:

- Loosely, these are CF languages such that you can make a context-free grammar using just sets of contexts substrings can appear in.
- CFGs have a finite set of category names.
- If a CFL has the finite kernel and context properties, a grammar built out of just a finite subset of the infinite possible substrings and contexts is the right grammar.
- Better candidate for human language class

Clark et al. (2010)

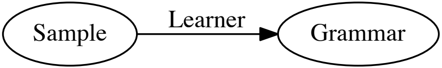
$CFG_{F,K}$

Already, just having heard abc, we guess our language is ab*c!

- bc is a substring of abc and (ab,ϵ) is a context of abc, so we got to ask the oracle "is abbc ok?"
- In real life, this means that the baby must have heard abbc from time to time.
- But, we never asked "is abbbbbbbc ok?" yet we still built a grammar that generates it.
- Starting with just abc, the learner asks whether ac and abbc are okay.
- This means the baby only has to hear ac, abc, abbc to infer ab*c

Summary of learners

	sub CFL	CFL$_{F,K}$
opt→rep	✓	✗
rep→opt	✓	✗
ac,abc,abbc→ab*c	✓	✓
HL-like	somewhat	closer



1. **Grammars**: What would a grammar look like that modelled adjuncts as adjuncts, yet accounted for ordering?
2. **Models**: How do formal models of language learning learn properties of adjuncts?
3. **People**: How do people learn properties of adjuncts?
4. **Birdsong**: What do birds generate?

Artificial language learning

- We can gather evidence to choose between learning models using real learning data from humans
- Acquisition data
- Artificial language learning experiments

Artificial language learning

The paradigm:

Training phase Participants are exposed to grammatical items from the target language

Testing phase Participants are tested on new items to see what they learned. Data like reaction time and grammaticality judgments are gathered to infer what the participants learned

Artificial language learning

Question: Does ALL tell us anything at all about real language learning?
Hopeful answers:

1. Yes, because learning is largely unconscious/implicit (eg Reber (1967))
2. Yes, because of neuroimaging data (Opitz and Friederici, 2007)

(Opitz and Friederici, 2007):

- Participants were trained on human-language-like sentences
- In an MRI, participants listened to grammatical and ungrammatical sentences of the artificial language.
- The higher their proficiency with the artificial language, the more the fMRI of their brains looked like they were processing their native language (Broca's area activation)

Artificial language learning

Goal: to use ALL to find out how people learn adjuncts

- Step 1 Do people generalise from limited to indefinite word repetition?
- Step 2 Do people generalise from optionality to word repetition?
- Step 3 Do people generalise from limited to indefinite category repetition?
- Step 4 What about when you embed the repetition in a highly complex grammar?
- Step 5 ...

Artificial language learning experiment 1: Tagalog

(7) natulog ang babae
sleep D woman
'The woman is sleeping/slept'

(8) natulog ang malaki babae
sleep D big woman
'The big woman is sleeping/slept'

(9) natulog ang malaki malaki babae
sleep D big big woman
'The big big woman is sleeping/slept'

(10) natulog siguro ang babae
sleep maybe D woman
'Maybe the woman is sleeping/slept'

V (Adv) D Adj* N

Artificial language learning experiment 1: Tagalog

Research Question: In learning language, do people generalise from limited to indefinite repetition?

Training stimuli V (Adv) D (Adj) (Adj) (Adj) N

Testing stimuli also ungrammatical and V (Adv) D Adj^4 (Adj) N

To answer the research question: compare responses to ungrammatical and generalised stimuli. If they like generalised stimuli more than ungrammatical, they've generalised repetition

Artificial language learning experiment 1: Tagalog

Testing stimuli

- 100 grammatical sentences
  - 62 familiar (no more than 3 adjectives)
  - 38 generalised (20 with 4 adjectives, 18 with 5)
- 78 ungrammatical sentences
  - 58 with the noun repeated instead of the adjective
  - 20 scrambled up grammatical sentences

Artificial language learning experiment 1: Tagalog

Training examples:

- natulog ang pusa'
- natulog siguro ang mapula kotse
- natulog siguro ang matanda matanda pusa'
- natulog ang masaya masaya masaya kotse

Artificial Language Learning experiment 1: Tagalog

Testing examples:
Generalised:

- umalis ang malaki malaki malaki malaki babae
- umalis ang malaki malaki malaki malaki malaki kotse
- umalis siguro ang mapula mapula mapula mapula babae

Repeated noun:

- *natulog ang malaki babae babae
- *umalis ang matanda kotse kotse kotse kotse

Scrambled:

- *siguro matanda matanda babe ang natulog
- *babae ang umalis

Artificial language learning experiment 1

Method:

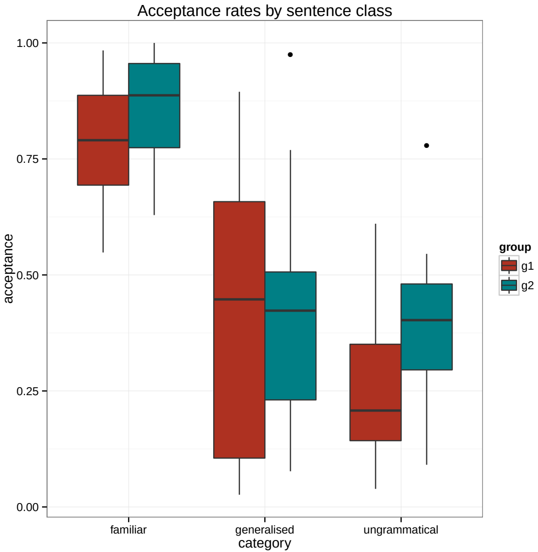
- **Training:** Participants listen to randomised training stimuli over headphones
- While they listen they see a red dot on the screen
- They respond with a keypress according to which side of the screen it's on. (distractor task)
- **Testing:** Participants listen to new stimuli
- They respond with a keypress whether they think it's a real or fake sentence of Tagalog (forced choice)

Experiment 1: Preliminary Results

- 51 UCLA undergrads
- For the main question, we dropped 6 peoples' results because they didn't learn the basics well enough (Accepted Familiar less than 0.15 more than Ungrammatical)
- 29 Grammar 1: D A* N
- 22 Grammar 2: D N A*

Experiment 1: Results

- People learned the grammar
- People do not completely generalise repetition
- People can recognise generalised repetition as grammatical



Experiment 1: Analysis

- Why did only Adjective-medial participants generalise?
- Why such a wide spread of responses?
    - Different strategies?
    - Some representative comments from the exit survey:
        - *Repeating never at the end, only in the middle* (Accepted 62% of generalised)
        - *I remembered specific words that were repeated in the first portion. If they were repeated too many times, I thought it sounded grammatically incorrect.* (Accepted 12% of generalised)
        - *If he repeated something too many times, I said it wasn't logical* (Accepted 44% of generalised. 3/18 with 5 As, 13/20 with 4 As)

Experiment 1: Analysis

- I'm trying to get people to learn unconsciously, but maybe word repetition is just too salient
- In English, some adjuncts are indefinitely repeatable, but maybe people's acceptance drops the more there are.
- ie Maybe this isn't just grammatical.

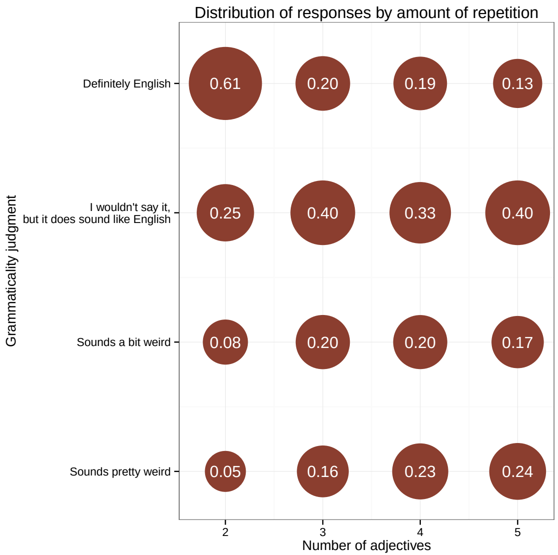
Experiment 2: English survey

- Survey Monkey survey on English word repetition
- 30 sentences
- 5 choices:
  1. Doesn't sound like English
  2. Sounds pretty weird
  3. Sounds a bit weird
  4. I wouldn't say it, but it does sound like English
  5. Definitely English

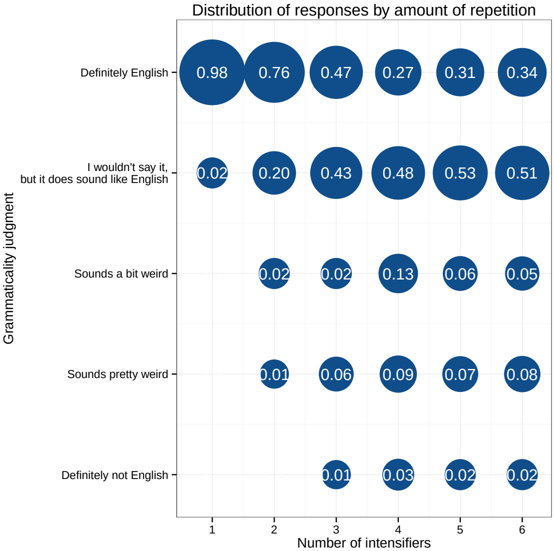
Experiment 2: English survey

- 4: 2 Adjectives (*Can someone help me tear out this itchy itchy tag?*)
- 2: 3 Adjectives (*What a stupid stupid stupid idea!*)
- 3: 4 Adjectives (*I haven't seen her in a long long long long time*)
- 2: 5 Adjectives (*The big big big big big elephant stomped.*)
- 1-6 *reallys* (*I really* like her*)
- 1: 4 *sos* (*Marie ate so so so so much food!*)
- 1: 6 *evers* (*I'll never ever ever ever ever ever ever leave you*)
- 11 fillers

Experiment 2: English survey results



Experiment 2: English survey results

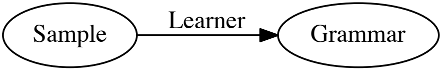


Experiment 2: English survey results

- Negative correlation between # repetitions and grammaticality
- All: Kendall $\tau = -0.2756566, p \ll 0.0001$
- Adjectives: Kendall $\tau = -0.6881024, p \ll 0.0001$
- Intensifiers Kendall $\tau = -0.396297, p \ll 0.0001$

Next

- Category repetition (*the big mean nasty bully*)
- Embed repetition in a more complex grammar



1. **Models**: How do formal models of language learning learn properties of adjuncts?
2. **People**: How do people learn properties of adjuncts?
3. **Birdsong**: What do birds generate?

Summary

- Some learning algorithms generalise from optionality to repetition and vice versa
- The learning algorithm for the language class that corresponds most closely to human languages predicts that given a minimal amount of optionality and repetition in the input, the learner should assume indefinite repetition
- People can (but don't necessarily) generalise from limited to indefinite repetition

Adger, David. 2008. A minimalist theory of feature structure. MSS, May 2008.

Applegate, Richard B. 1972. Ineseño Chumash Grammar. University of California at Berkeley.

Bach, Emmon. 1975. Long vowels and stress in Kwakiutl. *Texas Linguistic Forum* 2:9–19.

ten Cate, Cael, and Kazuo Okanoya. 2012. Revisiting the syntactic abilities on non-human animals. *Philosophical Transactions of the Royal Society B* 367:1984–1994.

Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.

Chomsky, Noam. 1993. A minimalist program for linguistic theory. In *The view from building 20: Essays in linguistics in honor of sylvain bromberger*, ed. Kenneth Hale and Samuel Jay Keyser, 1–52. Cambridge, MA: MIT Press.

Cinque, Gugliemo. 1999. *Adverbs and functional heads: a cross-linguistic perspective*. Oxford studies in comparative syntax. Oxford: Oxford University Press.

Clark, Alexander, Rémi Eyraud, and Amaury Habrard. 2010. Using contextual representations to efficiently learn context-free languages. *The Journal of Machine Learning Research* 11:2707–2744.

Fitch, W. Tecumseh, Angela D. Frederici, and Peter Hagoort. 2012. Pattern perception and computational complexity: Introduction. *Philosophical Transactions of the Royal Society B* 367:1925–1932.

Fitch, W Tecumseh, and Marc D Hauser. 2004. Computational constraints on syntactic processing in a nonhuman primate. *Science* 303:377–380.

Frank, Robert, and Giorgio Satta. 1998. Optimality theory and the generative complexity of constraint violability. *Computational Linguistics* 24:307–315. http://aclweb.org/anthology/J/J98/J98-2006.pdf.

Graf, Thomas. 2013. Local and transderivational constraints in syntax and semantics. Doctoral Dissertation, UCLA.

Harkema, Henk. 2001. A characterization of minimalist languages. In *Logical Aspects of Computational Linguistics*, ed. P. de Groote, G. Morrill, and C. Retoré, LNCS 2099, 193–211. NY: Springer.

Heinz, Jeffrey. 2010. Learning long-distance phonotactics. *Linguistic Inquiry* 41:623–661.

Kepser, Stephan, and Uwe Mönnich. 2006. Properties of linear context free tree languages with an application to optimality theory. *Theoretical Computer Science* 354:82–97.

Kershenbaum, Arik, Ann E. Bowles, Todd M. Freeberg, Dezhe Z. Jin, Adriano R. Lameira, and Kirsten Bohn. 2014. Animal vocal sequences: Not the Markov chains we thought they were. *Proceedings of the Royal Society of London B: Biological Sciences* 281.

Kobele, Gregory M. 2006. Generating copies: An investigation into structural identity in language and grammar. Doctoral Dissertation, UCLA.

Lipkind, Dina, Gary F. Marcus, Douglas K. Bemis, Kazutoshi Sasahara, Nori Jacoby, Miki Takahasi, Kenta Suzuki, Olga Feher, Primoz Ravbar, Kazuo Okanoya, and Ofer Tchernichovski. 2013. Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. *Nature* 498:104–108. URL `http://dx.doi.org/10.1038/nature12173`.

Manaster-Ramer, Alexis. 1987. Dutch as a formal language. *Linguistics and Philosophy* 10:221–246.

Michaelis, Jens. 2001. Transforming linear context free rewriting systems into minimalist grammars. In *Logical Aspects of Computational Linguistics*, ed. Philippe de Groote, Glyn Morrill, and Christian Retoré, Lecture Notes in Artificial Intelligence, No. 2099, 228–244. NY: Springer.

Michaelis, Jens, and Marcus Kracht. 1997. Semilinearity as a syntactic invariant. In *Logical Aspects of Computational Linguistics*, ed. Christian Retoré, 37–40. NY: Springer-Verlag.

Miller, George A. 1967. Project grammarama. In *Psychology of communication*. NY: Basic Books.

Mönnich, Uwe. 2007. Minimalist syntax, multiple regular tree grammars, and direction preserving tree transductions. In *Model Theoretic Syntax at 10. ESSLLI'07 Workshop Proceedings*, ed. James Rogers and Stephan Kepser. http://cs.earlham.edu/esslli07mts/.

Opitz, Bertram, and Angela Friederici. 2007. Neural basis of processing sequential and hierarchical syntactic structures. *Human Brain Mapping* 585–592.

Reber, Arthur S. 1967. Implicit learning of artificial grammars. *Journal of Verbal learning and Verbal Behaviour* 6:855–863.

Seki, Hiroyuki, Takashi Matsumura, Mamoru Fujii, and Tadao Kasami. 1991. On multiple context-free grammars. *Theoretical Computer Science* 88:191–229.

Shieber, Stuart M. 1985. Evidence against the context-freeness of natural language. *Linguistics and Philosophy* 8:333–344.

Sportiche, Dominique. 2005. Division of labor between merge and move: Strict locality of selection and apparent reconstruction paradoxes. *LingBuzz* .

Stabler, Edward. 1997a. Derivational minimalism. *Logical Aspects of Computational Linguistics* 68–95.

Stabler, Edward. 1997b. Derivational minimalism. In *Logical aspects of computational linguistics*, ed. C. Retoré, LNCS 1328, 68–95. NY: Springer-Verlag.

Stabler, Edward P. 2004. Varieties of crossing dependencies: Structure dependence and mild context sensitivity. *Cognitive Science* 93:699–720.

Stabler, Edward P. 2011. Computational perspectives on minimalism. In *Oxford handbook of linguistic minimalism*, ed. Cedric Boeckx, 617–641. Oxford: Oxford University Press.

Torrence, Harold, and Khady Tamba. 2014. Factive relative clauses in Wolof. University of Kansas, forthcoming.

Vijay-Shanker, K., David Weir, and Aravind K. Joshi. 1987. Characterizing structural descriptions produced by various grammatical formalisms. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, 104–111.

Many thanks to...

- My chair, Ed Stabler
- My committee: Ed Keenan, Carson Schutze, Martin Monti (psychology)
- Chuck Taylor for bringing me onto the birdsong project
- My RA, Shayna Lurya
- Seth Ronquillo for recording Tagalog stimuli
- This wonderful department
- My fellow students, particularly Floris van Vugt, Kathleen O'Flynn, Thomas Graf, Yu Tanaka, and Michael Lefkowitz, for many fruitful discussions
- My family and friends

Thank You!

Birdsong Project

Edward Stabler	Linguistics, UCLA
Meaghan Fowlie	Linguistics, UCLA
Floris van Vugt	Motor Control Lab, McGill University
Martin Cody	Ecology and Evolutionary Biology, UCLA
Charles Taylor	Ecology and Evolutionary Biology, UCLA

NSF grant IIS-1125423

Some of these slides are modified from Ed Stabler's slides *Self-mimicry in the California Thrasher* made for the Understanding Bird Song conference at UCLA, 2014-12-04

California Thrasher

California Thrasher song

Martin Cody recorded single birds: 2618, 2881 phrases, repetoire of 105, 181
California Thrasher song

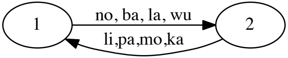
```
aiw aix aiw aix aiw aix aiu ago ait ago ait aiz ##
aja aiz aje aii aiy ajd aiy ajd aiu ago ait #
aiw aix aiu ago ait ##
aiu ago ait ago ait aiz aja aiz aje aiz aja aiz aiy ajd aiy ajd
aiu ##
aiz aja #
aiw aix aiw aix aiw aix ait #
aiu ago aiy aiw aix aiw aix ##
...
```

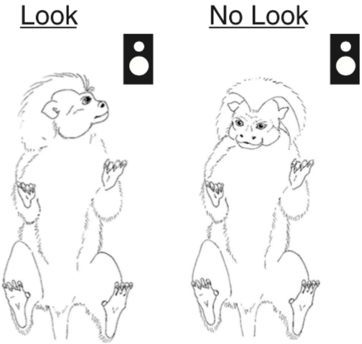
Animal communication

- Most previous research: Animal communication has far less complex syntax than human language
- Namely syntax, when utterances go beyond one call, is thought to be strictly local/n-grams/transitional probabilities/Markov chains
- Eg: Fitch and Hauser (2004): Monkeys can't recognise languages with centre-embedding, but humans can

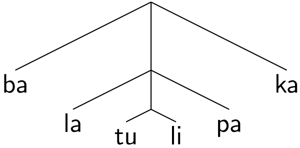
Animal communication

Regular grammar:





Context-free grammar



Our claim: California Thrasher has human-language level syntax

(Heinz, 2010)

. . . constituent structure languages are more natural, easier to cope with, than regular languages. . . The hierarchical structure of strings generated by constituent-structure grammars is characteristic of much other behavior that is sequentially organized . . . . . . easier for people than . . . the left-to-right organization characteristic of strings generated by regular grammars

(Miller, 1967)

Supra-regular distinctiveness hypothesis

- (Fitch&al'12) Summing up [the] data, it currently seems plausible to suppose that most animal species do not display the tendency Miller observed in humans, to find a supra-regular grammar 'more natural and easier to cope with' (p1929)
- (ten Cate & Okanoya'12) ... while the patterns in animals differ between species and may also differ among individuals within species, they all have in common that the structures are no more complex than a probablistic finite state grammar (p.1992)
- (Lipkind&al'13) A bigram Markov model was found to account for the bulk of song sequence structure during the experimental period... (p104)
- (Kershenbaum&al'14) The vocal systems of [many species] are more consistent with a non-Markovian [renewal process] than with the Markovian models traditionally assumed.

Bigrams

$\Sigma = \{a, b\}$, boundary markers $\rtimes$, $\ltimes$

Bigrams: $\rtimes a$, ab, bb, $b\ltimes$, $a\ltimes$

- a
- ab
- abb
- abbb

Hypothesis language: ab^*

'b-i factive relative clauses' in Wolof (Torrence and Tamba, 2014)

(11)	[Sàcc	ñaari	tééré]	b-i	ñu	[sàcc	ñaari	tééré]
	steal	two	book	CL-C_{Rel}	3PL	steal	two	book
	'the fact that they stole two books'							
(12)	[daw	bu.gaaw]	b-i	mu	[daw	bu.gaaw]		
	run	quickly	CL-C_{Rel}	3SG	run	quickly		
	'the fact that he ran quickly'							

- Yoruba, Russian, Spanish, Mandarin Chinese, Nweh, Ngumbe,...
- Other copying (Kobele'06;Stabler'04;Michalis&Kracht'96)
- This kind of copying is allowed in PMCF grammars (Seki&al'91)

H2. The Thrasher uses recursive copying too.

In the corpus, 6 sequences of length 13 phrases are repeated exactly. E.g.,

```
cv ct ct cx cu cx cu cx cs cu cs cs cu
cv ct ct cx cu cx cu cx cs cu cs cs cu
```

Note the copies in these copies. Some sequences of length 6 occur more than 40 times.

. . . Is this actual copying or just the kind of repetition that is expected even in simple unigram and bigram models?

Method

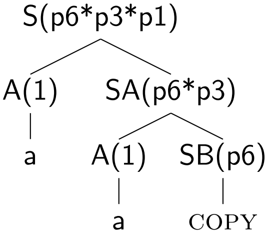
- CKY parser for probabilistic context-free grammars
- Implemented bigrams as PCFG rules
- Bigram Grammar: just these rules
- Copy Grammar:
  - 2-level grammar
  - a COPY node that must then be turned into a copy of some amount of preceding material

Method

Sample Grammar

- S→A SA p1
- S→a p2
- SA→A SB p3
- SA→COPY p4
- SB→a p5
- SB→COPY p6
- A→a p=1

Method



p(a a COPY) = p6*p3*p1

COPY →a (p7) or aa (p8)

p(aaa) = p6*p3*p1*p7

p(aaaa) = p6*p3*p1*p8

Method

To figure out the rule probabilities:

- Divide corpus into 2 parts: training items and testing items
- Use Expectation Maximisation algorithm to learn the probabilities of the rules, using just the training items
- Calculate how likely each grammar makes the testing items

Method

Expectation Maximisation:

- Start with random probabilities for the rules
- Parse each sentence (give it a tree)
- The parser also gives the probability of the sentence, given the current grammar
- Count how many times each rule was used, versus how many times you would expect it to be used given the rule probabilities
- Adjust the probabilities accordingly and try again with the new grammar

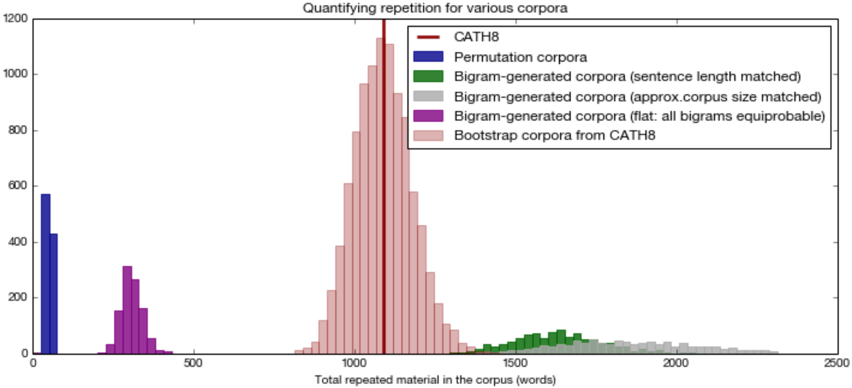
$$P(\text{Data}|\text{bigram FS} + \text{copy}) \; > \; P(\text{Data}|\text{bigram FS})$$

Average log likihoods of testing corpus:

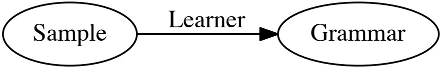
BiCopy	BiNoCopy	difference
-2724.4	-2782.8	58.4

That's 25 orders of magnitude. (10^{25})

- Wait, Markov models also predict copying!
- True. In fact, a Markovian grammar built from the transitional probabilities in one of our corpora predicts *more* repetition than is really in the corpus
- But, this means a Markov model **can't predict** the amount of repetition present in the real corpus



Summary



1. **Grammars**: What would a grammar look like that modelled adjuncts as adjuncts, yet accounted for ordering?
2. **Models**: How do formal models of language learning learn properties of adjuncts?
3. **People**: How do people learn properties of adjuncts?
4. **Birdsong**: What do birds generate?

Cartography: default adjunct orders

- Using the same architecture, how can we capture ordering?

(13) a. Wear the enormous ugly green hat
Wear the hat that is enormous, ugly, and green
b. #Wear the ugly enormous green hat
Of your enormous green hats, wear the ugly one.

Cartography (Cinque, 1999)

(14)

a. The small ancient triangular green Irish pagan metal artifact was lost.

b. *The metal green small artifact was lost. **Adjectives**

c. Frankly, John probably once usually arrived early.

d. *Usually, John early frankly once arrived probably. **Adverbs**

e. [Il premio Nobel]$_{\text{top}}$, [a chi]$_{\text{wh}}$ lo daranno?
[the prize Nobel]$_{\text{top}}$, [to whom]$_{\text{wh}}$ it give.fut
'The Nobel Prize, to whom will they give it?' **Left periphery**

f. DP zhe [NumP yi [ClP zhi [NP bi]]]
DP this [NumP one [ClP Cl [NP pen]]]
'this pen' **Functional DP projections**

Adjoin: formal definition

Definition (Adjoin)

Let $s, t \in \Sigma$ be strings, $\mathtt{Y,X} \in \textbf{sel}$ be categories, $i, j, n, m \in \mathbb{N}$, $mvrs \in (\Sigma^* \times F)^*$ be a mover list, and $\alpha, \beta \in F^*$.

$$\textbf{Adjoin}(\langle s, [\mathtt{X}, \mathtt{i}, \mathtt{j}]\alpha :: mvrs\rangle, \langle t, [\mathtt{Y}, \mathtt{n}, \mathtt{m}]\beta\rangle) = \begin{cases} \langle ts, [\mathtt{X}, \mathtt{i}, \mathtt{n}]\alpha\rangle :: mvrs & \text{if } \mathtt{n} \geq \mathtt{j} \;\&\; \mathtt{Y} \in \textbf{Ad}(\mathtt{X}) \;\&\; \beta = \epsilon \\ \langle s, [\mathtt{X}, \mathtt{i}, \mathtt{n}]\alpha\rangle :: \langle t, \beta\rangle :: mvrs & \text{if } \mathtt{n} \geq \mathtt{j} \;\&\; \mathtt{Y} \in \textbf{Ad}(\mathtt{X}) \;\&\; \beta \neq \epsilon \end{cases}$$

Merge: new formal definition

Definition (Merge)

For $\alpha, \beta \in F^*$; s, t strings:

$\textbf{Merge}(\langle s, \texttt{=X}\alpha\rangle :: \text{mvrs}_s,\ \langle t, [\texttt{X}, \texttt{i}, \texttt{j}]\beta\rangle :: \text{mvrs}_t) =$

$$\begin{cases} \langle st, \alpha\rangle :: \text{mvrs}_s \cdot \text{mvrs}_t & \text{if } \beta = \epsilon \\ \langle s, \alpha\rangle :: \langle t, \beta\rangle :: \text{mvrs}_s \cdot \text{mvrs}_t & \text{if } \beta \neq \epsilon \end{cases}$$

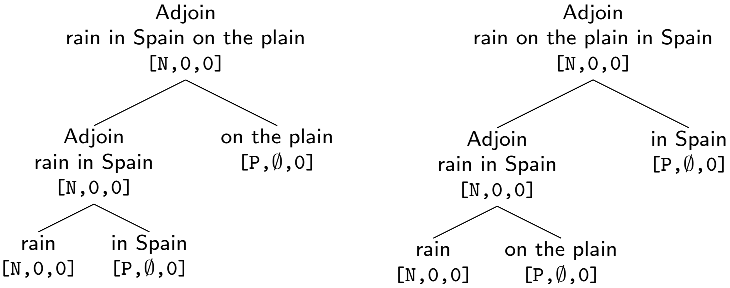
Unordered adjuncts

? Unordered

- Could make them all one level
- Or at every level
- Better: expand index set to include non-number, $\emptyset$
  - When **Adjoin** sees $\emptyset$, *asymmetrically* checks features
  - Hierarchy level of phrase doesn't change

Unordered adjuncts

✓ Unordered



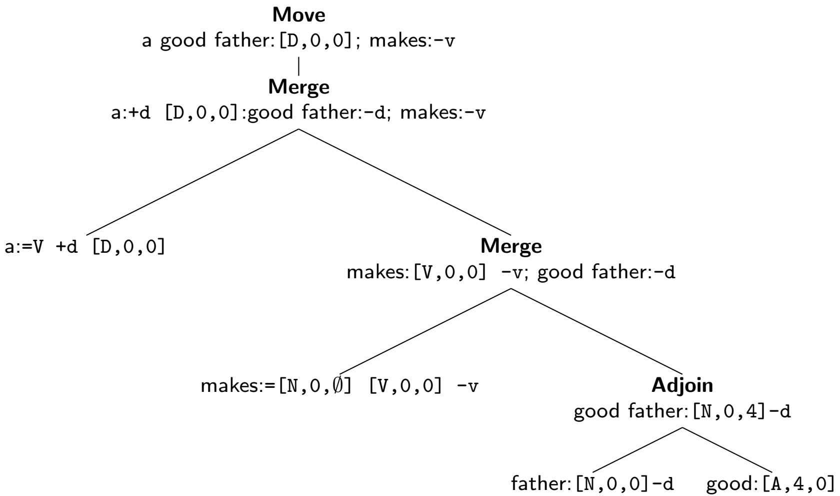
✓ Adjuncts on either side of head

Obligatory adjuncts

(15) a. He makes a **good** father.
b. *He makes a father.

- Noun with no adjuncts: `[N,0,0]`
- Noun with adjunct: `[N,0,3]`
- $\rightarrow$Expand Merge to require last element to be non-zero
- `=[N,0,`$\emptyset$`]` can Merge with `[N,i,j]` for `j>0`

Obligatory adjuncts



Note: Sportiche (2005) proposes that verbs select NPs, and the NPs move to their Ds, which are functional heads on the spine.

Minimalist Grammars with Hierarchies

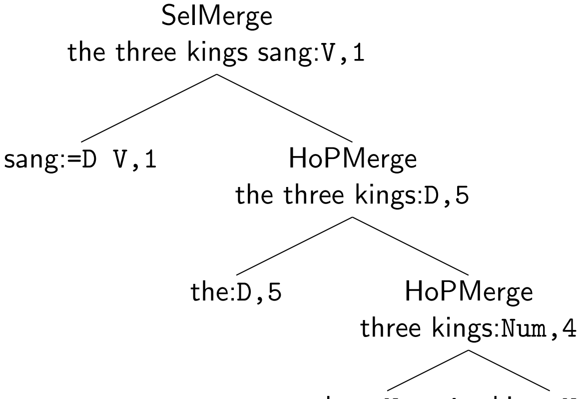
Adger (2008): Hierarchy of Precedence Merge

textbfLex = {the:`=D V,0`, three:`Num,4`, kings:`N,0`, sang:`=D V,0`}

$H_1 = \texttt{D}, 5 > \texttt{Num}, 4 > \texttt{Poss}, 3 > \texttt{n}, 2 > \texttt{N}, 1$

$H_2 = \texttt{C}, 3 > \texttt{T}, 2 > \texttt{V}, 1$

We can derive *the three kings sang* as in Figure **??**.



Minimalist Grammars with Hierarchies