What Can Formal Languages Tell Us About Cognition?

Jon Rawski

Department of Linguistics
Institute for Advanced Computational Science
Stony Brook University

Adaptive specialization of mechanism is so ubiquitous and so obvious in biology, at every level of analysis, and for every kind of function, that no one thinks it necessary to call attention to it as a general principle about biological mechanisms...

From a biological perspective, the idea of a general-learning mechanism is equivalent to assuming that there is a general-purpose sensory organ, which solves the problem of sensing.

Gallistel and King 2009

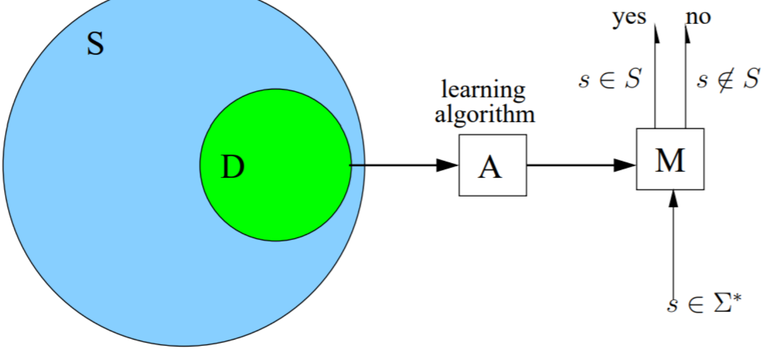
> *The trouble is that an observer who notices everything can learn nothing, for there is no end of categories known and constructable to describe a situation*

Lila Gleitman (1990)

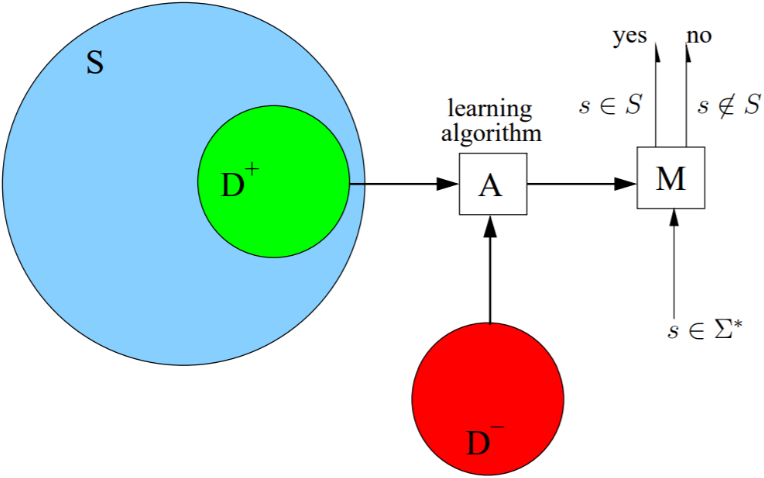
> *There is no possible way in which we could think of a device, natural or artificial, that is freed from all structural information.*
>
> *Within the limits set, however, there are infinitely many variations possible. Thus the outer form of languages may vary with relatively great freedom, whereas the underlying type remains constant*

Eric Lenneberg, *Biological Foundations of Language*

What does Learning Mean?



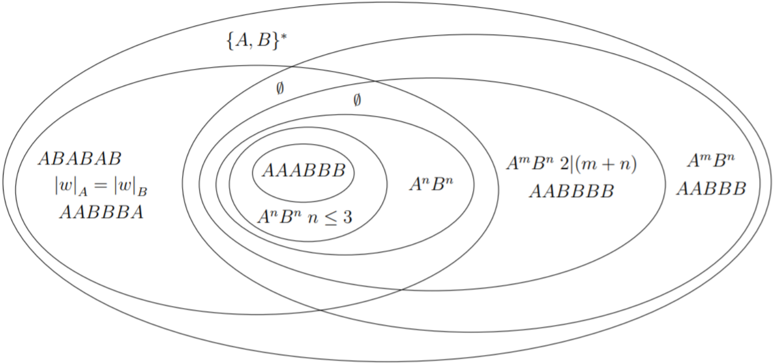
What does Learning Mean?



Some Patterns

$(AB)^n$ vs A^nB^n in English
$\{(\text{ding dong})^n\}$
$\{\text{people}^n\ \text{left}^n\}$
$\{\text{people (who were left by people)}^n\ \text{left}\}$
$\{\text{people (who were left by people)}^2 n\ \text{left}\}$

Stringset Inference (Jager & Rogers 2012)



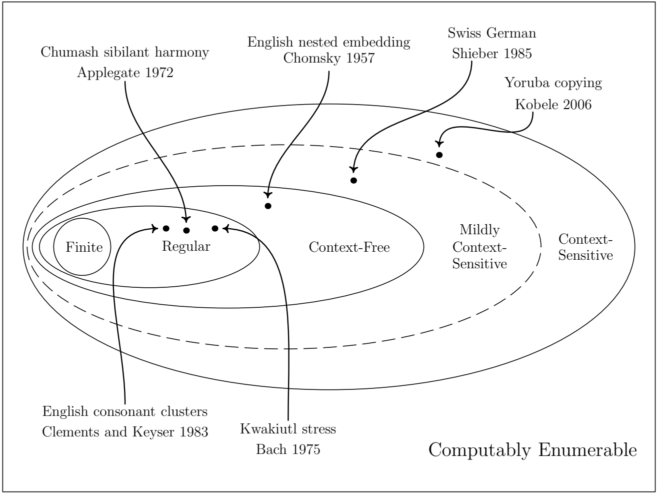
Cognitive Complexity from First Principles

What kinds of distinctions does a cognitive mechanism need to be sensitive to in order to classify an event with respect to a pattern?

Reasoning about patterns

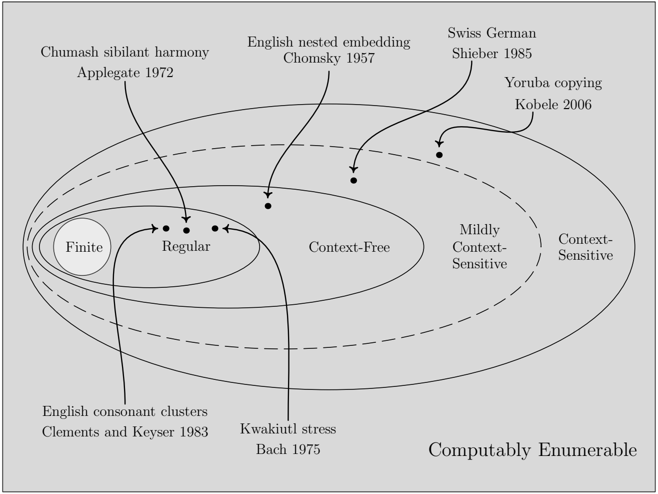
- What objects/entities/things are we reasoning about?
- What relationships between them are we reasoning with?

Computational Theories of Language



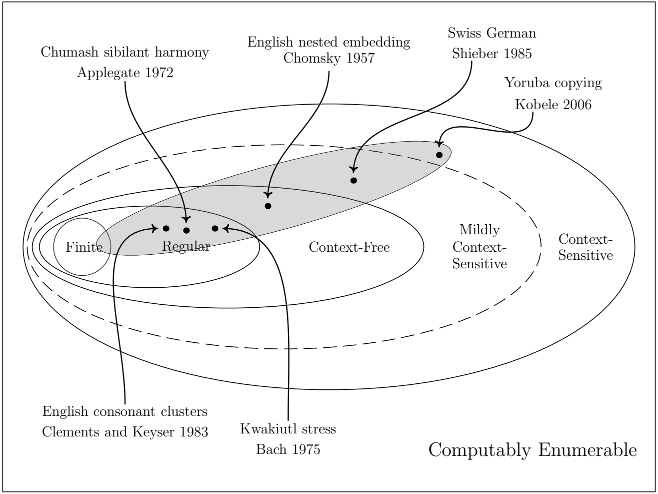
(Garcia et al. 1991, Heinz 2010)

Computational Theories of Language



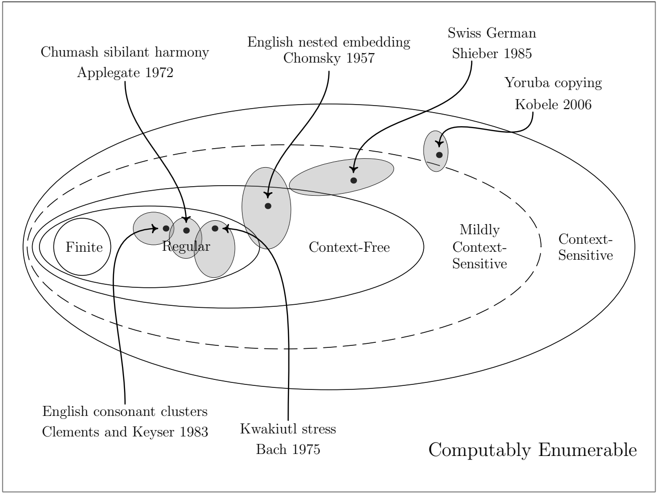
(Garcia et al. 1991, Heinz 2010)

Computational Theories of Language



(Garcia et al. 1991, Heinz 2010)

Computational Theories of Language



(Garcia et al. 1991, Heinz 2010)

Dual characterizations of complexity classes

Computational classes

- Characterized by abstract computational mechanisms
- Equivalence between mechanisms
- Means to determine structural properties of stringsets

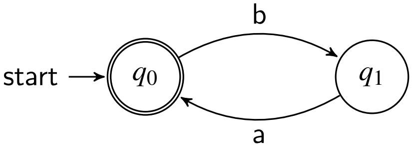
Descriptive classes

- Characterized by the nature of information about the properties of strings that determine membership
- Independent of mechanisms for recognition
- Subsume wide range of types of patterns

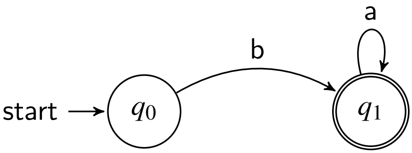
Regular Languages & Finite-State Automata

Regular Language: Memory required is finite w.r.t. input

(ba)*: {ba, baba, bababa,...}

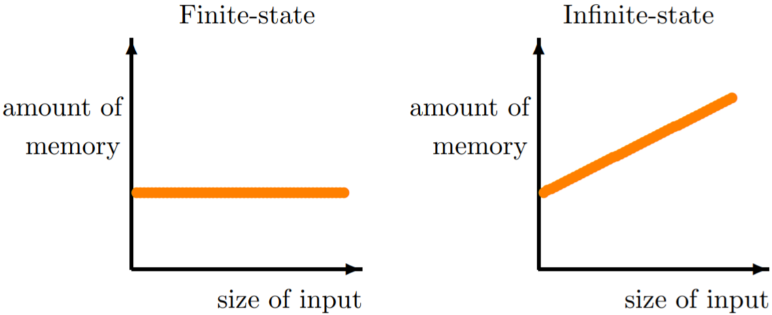


b(a*): {b, ba, baaaaaa,....}



What "Being Regular" Means

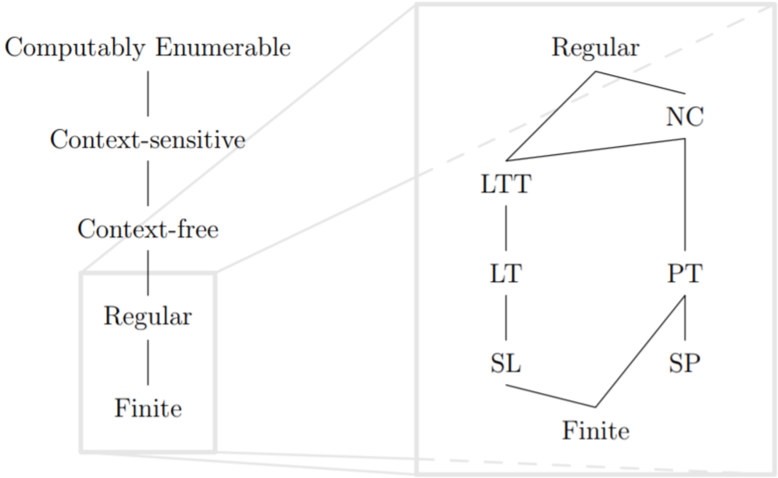
A generalization is finite-state provided the memory required is bounded by a constant, regardless of the size of the input.



Cognitive Interpretation of Finite-State (Rogers et al 2013)

- Any cognitive mechanism that can distinguish member strings from non-members of a finite-state stringset must be capable of classifying the events in the input into a finite set of abstract categories and are sensitive to the sequence of those categories.
- Subsumes any recognition mechanism in which the amount of information inferred or retained is limited by a fixed finite bound.
- Any cognitive mechanism that has a fixed finite bound on the amount of information inferred or retained in processing sequences of events will be able to recognize only finite-state stringsets.

The Subregular Hierarchy



Finite Word Models

Some Assumptions about Linguistic Behaviors

- Perceive/process/generate linear sequence of (sub)events
- Can model as strings—linear sequence of abstract symbols
  - Discrete linear order (initial segment of $\mathbb{N}$).
  - Labeled with alphabet of events
  - Partitioned into subsets, each the set of positions at which some event occurs.

Finite Word Models

'word' is synonymous with 'structure.'

- A model of a word is a representation of it.
- A (Relational) Model contains two kinds of elements.
  - A domain: a finite set of elements.
  - Relations over domain elements.
- Every word has a model.
- Different words have different models.

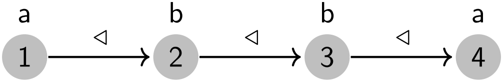
Finite Word Models

$$
\begin{tikzcd}[column sep=small]
 & \mathbb{W}^{\triangleleft,\triangleleft^+} = \langle \mathcal{D}^{\mathbb{W}}, \triangleleft^{\mathbb{W}}, {\triangleleft^+}^{\mathbb{W}}, P_\sigma^{\mathbb{W}} \rangle_{\sigma\in\Sigma} \arrow[dl] \arrow[dr] & \\
\mathbb{W}^{\triangleleft} = \langle \mathcal{D}^{\mathbb{W}}, \triangleleft^{\mathbb{W}}, P_\sigma^{\mathbb{W}} \rangle_{\sigma\in\Sigma} & & \mathbb{W}^{\triangleleft^+} = \langle \mathcal{D}^{\mathbb{W}}, {\triangleleft^+}^{\mathbb{W}}, P_\sigma^{\mathbb{W}} \rangle_{\sigma\in\Sigma}
\end{tikzcd}
$$

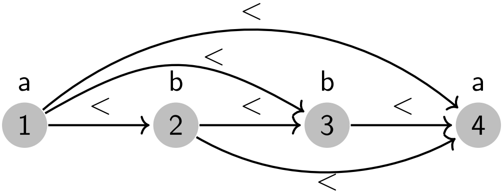
$$
\begin{array}{rcl}
\mathcal{D}^{\mathbb{W}} & — & \text{Finite set of elements (positions)} \\
\triangleleft^{\mathbb{W}} & — & \text{immediate linear precedence on } \mathcal{D} \\
{\triangleleft^+}^{\mathbb{W}} & — & \text{(arbitrary) linear precedence on } \mathcal{D} \\
P_\sigma^{\mathbb{W}} & — & \text{Subset of } \mathcal{D} \text{ at which } \sigma \text{ occurs}
\end{array}
$$

Finite Word Models

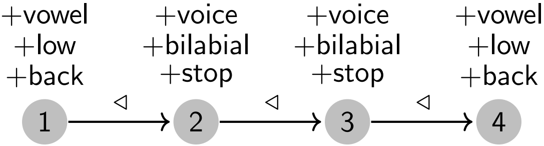
1. Successor (Immediate Precedence)



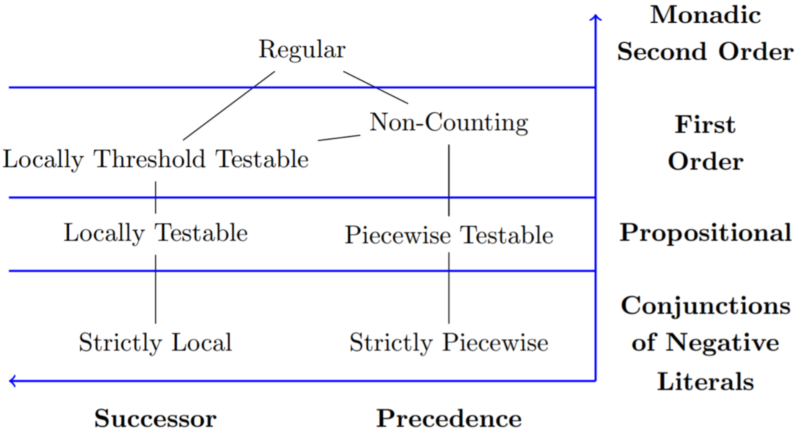
2. General precedence



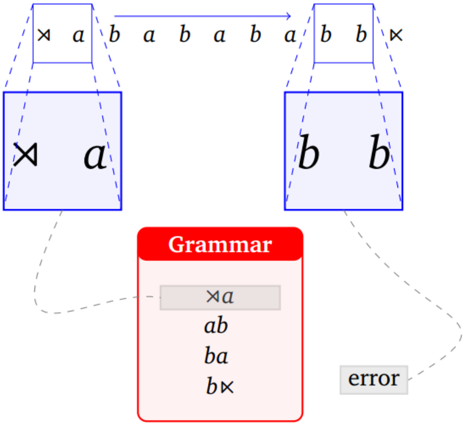
Relational Structures



Subregular Hierarchy (Rogers et al 2013)



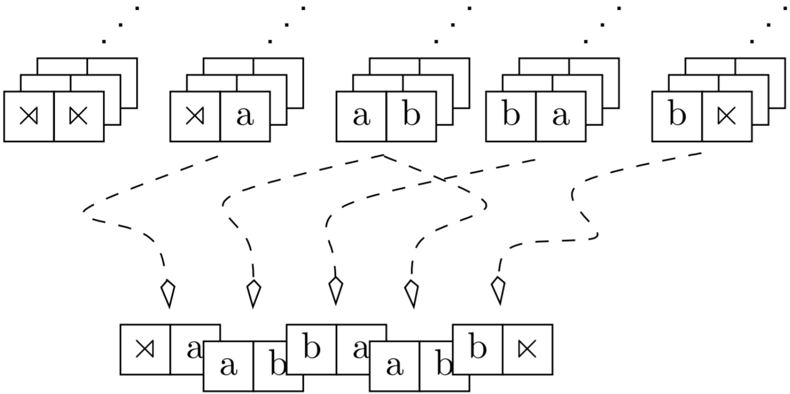
Strictly Local Grammars



- 1 window of size k
- Closed under suffix substitution

(Garcia et al. 1991, Heinz 2010), p.c. Thomas Graf

Local Factors



Pics courtesy of Heinz and Rogers 2014 ESSLLI course.

Phonology: local assimilations

Assimilation and word-final devoicing in Russian

- *Anticipatory obstruent voicing assimilation:*
  ot dveri '*from the door*' → o[ddv]eri
  iz korobki '*out of the box*' → i[sk]oro[pk]i
- *Obstruent word final devoicing:*
  moroz '*frost*' → moro[s]
  morozy '*frosts*' → moro[z]y

- $\Sigma = \{$b, s, z, …, k$\}$ $G = \langle$*td, *zk, *bk, …, *z⋉$\rangle$ $n = 2$
- mozg '*brain*' → mo[sk]
  *⋊mozg⋉ *⋊mosg⋉ *⋊mozk⋉ ^{ok}⋊mosk⋉

Phonology: local assimilations

Assimilation and word-final devoicing in Russian

- *Anticipatory obstruent voicing assimilation:*
  ot dveri '*from the door*' → o[ddv]eri
  iz korobki '*out of the box*' → i[sk]oro[pk]i
- *Obstruent word final devoicing:*
  moroz '*frost*' → moro[s]
  morozy '*frosts*' → moro[z]y

- $\Sigma = \{$b, s, z, …, k$\}$ $G = \langle$*td, *zk, *bk, …, *z⋉$\rangle$ $n = 2$
- mozg '*brain*' → mo[sk]
  *⋊mozg⋉ *⋊mosg⋉ *⋊mozk⋉ *ok* ⋊mosk⋉

Phonology: local assimilations

Assimilation and word-final devoicing in Russian

- *Anticipatory obstruent voicing assimilation:*
  ot dveri '*from the door*' → o[ddv]eri
  iz korobki '*out of the box*' → i[sk]oro[pk]i
- *Obstruent word final devoicing:*
  moroz '*frost*' → moro[s]
  morozy '*frosts*' → moro[z]y

- $\Sigma = \{$b, s, z, $\ldots$, k$\}$ $\quad G = \langle$*td, *zk, *bk, $\ldots$, *z$\ltimes\rangle$ $\quad n = 2$
- mozg '*brain*' → mo[sk]
  *$\rtimes$mozg$\ltimes$ *$\rtimes$mosg$\ltimes$ *$\rtimes$mozk$\ltimes$ ^{ok}$\rtimes$mosk$\ltimes$

Morphotactics: prefixes and suffixes

English affixes

- *Prefix un-:*
  unlock, unhash
- *Suffix -able:*
  lockable, hashable

- $\Sigma = \{$un, able, hash, $\ldots$, lock$\}$ $\quad n = 2$
  $G = \langle$*⋊able, *un⋉, *ableun, $\ldots$, *lockun, *ablehash$\rangle$
- ^{ok}⋊lock⋉ $\quad$ *⋊able-lock⋉ $\quad$ ^{ok}⋊un-hash-able⋉

Morphotactics: prefixes and suffixes

English affixes

- *Prefix un-:*
  unlock, unhash
- *Suffix -able:*
  lockable, hashable

- $\Sigma = \{$un, able, hash, ..., lock$\}$ $\quad n = 2$
  $G = \langle$*⋊able, *un⋉, *ableun, ..., *lockun, *ablehash$\rangle$
- ^{ok}⋊lock⋉   *⋊able-lock⋉ ^{ok}⋊un-hash-able⋉

Morphotactics: prefixes and suffixes

English affixes

- *Prefix un-:*
  unlock, unhash
- *Suffix -able:*
  lockable, hashable

- $\Sigma = \{$un, able, hash, ..., lock$\}$ $\quad n = 2$
  $G = \langle$*⋊able, *un⋉, *ableun, ..., *lockun, *ablehash$\rangle$
- ^{ok}⋊lock⋉ *⋊able-lock⋉ ^{ok}⋊un-hash-able⋉

Cognitive interpretation of SL (Rogers et al 2013)

- Any cognitive mechanism that can distinguish member strings from non-members of an SL_k stringset must be sensitive, at least, to the length k blocks of events that occur in the presentation of the string.
- If the strings are presented as sequences of events in time, then this corresponds to being sensitive, at each point in the string, to the immediately prior sequence of $k-1$ events.
- Any cognitive mechanism that is sensitive only to the length k blocks of events in the presentation of a string will be able to recognize only SL_k stringsets.

Unbounded Dependencies

- **Samala Sibilant Harmony**
  Sibilants must not disagree in anteriority.
  (Applegate 1972)

(1)
a. * hasxintilawaʃ
b. * haʃxintilawas
c. haʃxintilawaʃ

Example: Samala

*$ h a s x i n t i l a w a ʃ $

$ h a ʃ x i n t i l a w a ʃ $

Unbounded Dependencies

- **Samala Sibilant Harmony**
  Sibilants must not disagree in anteriority.
  (Applegate 1972)

(1) a. * hasxintilawaʃ
    b. * haʃxintilawas
    c. haʃxintilawaʃ

Example: Samala

* $ h a s x i n t i l a w a ʃ $

$ h a ʃ x i n t i l a w a ʃ $

Unbounded Dependencies

- **Samala Sibilant Harmony**
  Sibilants must not disagree in anteriority.
  (Applegate 1972)

(1) a. * hasxintilawaʃ
b. * haʃxintilawas
c. haʃxintilawaʃ

Example: Samala

* \$ h a $\boxed{\text{s x i n t i l a w a ʃ}}$ \$

\$ h a ʃ x i n t i l a w a ʃ \$

Unbounded Dependencies

- **Samala Sibilant Harmony**
  Sibilants must not disagree in anteriority.
  (Applegate 1972)

(1) a. * hasxintilawaʃ
  b. * haʃxintilawas
  c. haʃxintilawaʃ

Example: Samala

* $ h a s x i n t i l a w a ʃ $

$ h a ʃ x i n t i l a w a ʃ $

Unbounded Dependencies

- **Samala Sibilant Harmony**
  Sibilants must not disagree in anteriority.
  (Applegate 1972)

(1) a. * hasxintilawaʃ
b. * haʃxintilawas
c. haʃxintilawaʃ

Example: Samala

$$^{*}\$\ \mathrm{h\ a}\ \boxed{\mathrm{s\ x\ i\ n\ t\ i\ l\ a\ w\ a\ ʃ}}\ \$$$

$$\$\ \mathrm{h\ a}\ \boxed{\mathrm{ʃ\ x\ i\ n\ t\ i\ l\ a\ w\ a\ ʃ}}\ \$$$

- **But:** Sibilants can be arbitrarily far away from each other!

* $ s t a j a n o w o n w a ʃ $

Unbounded Dependencies

- **Samala Sibilant Harmony**
  Sibilants must not disagree in anteriority.
  (Applegate 1972)

(1) a. * hasxintilawaʃ

b. * haʃxintilawas

c. haʃxintilawaʃ

Example: Samala

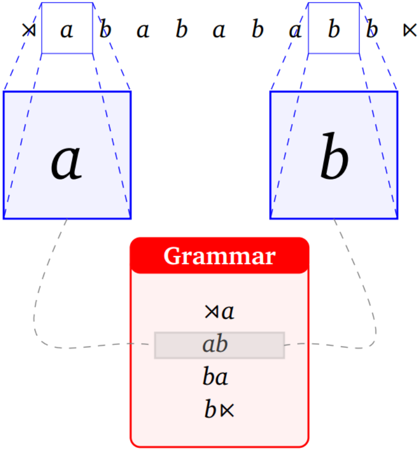
$* \$ h a \boxed{s x i n t i l a w a ʃ} \$$

$\$ h a \boxed{ʃ x i n t i l a w a ʃ} \$$

- **But:** Sibilants can be arbitrarily far away from each other!

$* \$ \boxed{s t a j a n o w o n w a ʃ} \$$

Strictly Piecewise Grammars



- 1 window of size k
- Closed under subsequence

(Heinz 2010), p.c. Thomas Graf

Cognitive interpretation of SP (Rogers et al 2013)

- Any cognitive mechanism that can distinguish member strings from non-members of an SP_k stringset must be sensitive, at least, to the length k (not necessarily consecutive) sequences of events that occur in the presentation of the string.
- If the strings are presented as sequences of events in time, then this corresponds to being sensitive, at each point in the string, to up to $k-1$ events distributed arbitrarily among the prior events.
- Any cognitive mechanism that is sensitive only to the length k sequences of events in the presentation of a string will be able to recognize only SP_k stringsets..

Locality and Projection

Theorem (Medvedev) *A set of strings is Regular iff it is a homomorphic image of a Strictly 2-Local set.*

Theorem (Thatcher) *A set of* Σ*-labeled trees is recognizable by a finite-state tree automaton (i.e. regular) iff it is a projection of a local set of trees.*

Theorem (Thatcher) *A set of strings* L *is the yield of a local set of trees (equivalently, is the yield of a recognizable set of trees) iff it is Context-Free.*

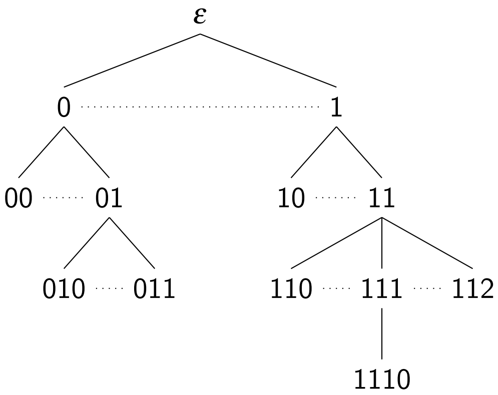
Supra-Regularity in Natural Language

The dog the cat the mouse feared loathed barked.

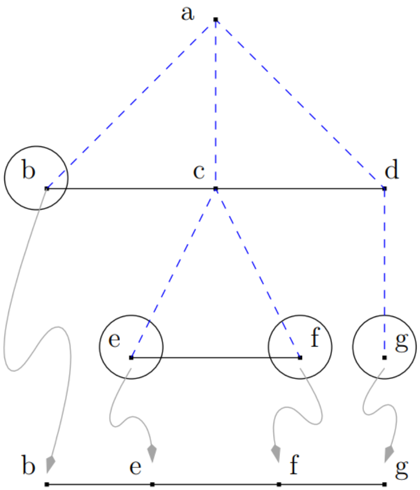
dat	*ik*	*Anna*	*Maria*	*Hans*	*zag*	*laten*	*helpen*	*zwemmen.*
THAT	I	ANNA	MARIA	HANS	SAW	LET	HELP	SWIM

that I saw Anna let Mary help Hans swim

Tree Models (Rogers 2003)

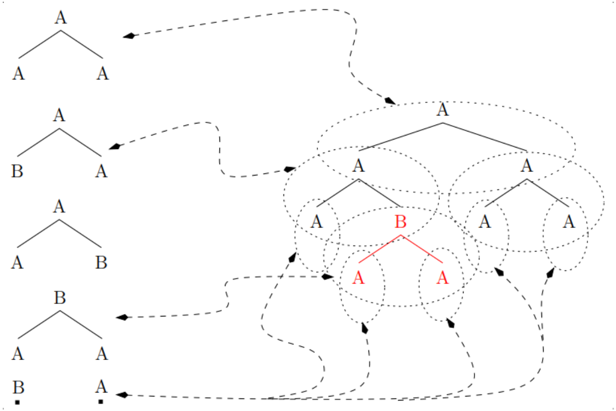


Tree Models (Rogers 2003)



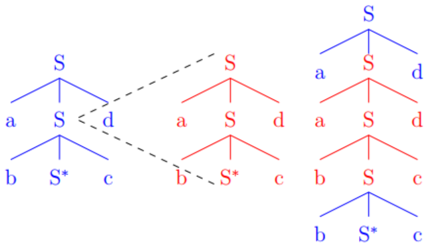
Pic courtesy of Rogers 2014 ESSLLI course.

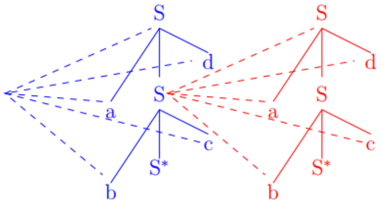
Local Tree Grammars



Intuition: Local Tree grammars are closed under subtree substitution.
p.c. Jim Rogers

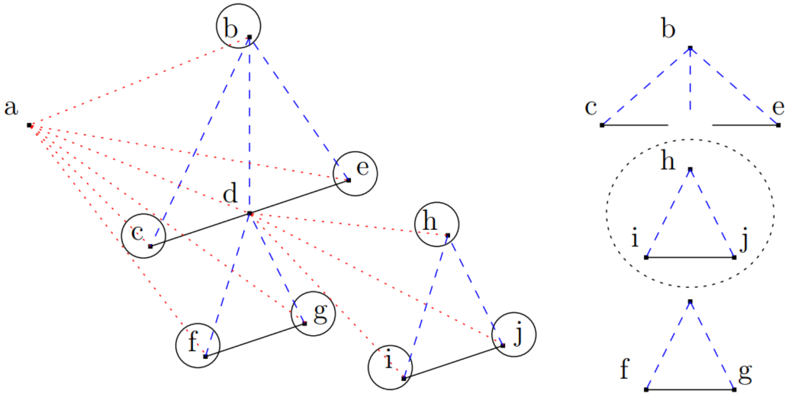
Beyond CFLs: Multi-Dimensional Trees





Pics courtesy of Jim Rogers

Beyond CFLs: Multi-Dimensional Trees



Pics courtesy of Jim Rogers

Syntax: Merge operation

It is possible to have n-gram constraints over trees.

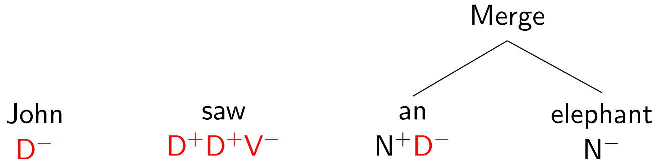
Local dependencies in syntax

- Chomsky-Borer hypothesis: grammar is just a finite list of feature-annotated lexical items.
- *Minimalist grammars (MGs)* are a formalization of Minimalist syntax. (Stabler 2011)
- Operations: **Merge** and **Move**.
- Merge is a *feature-driven* operation:
  category feature N^-, D^-, ...
  selector feature N^+, D^+, ...

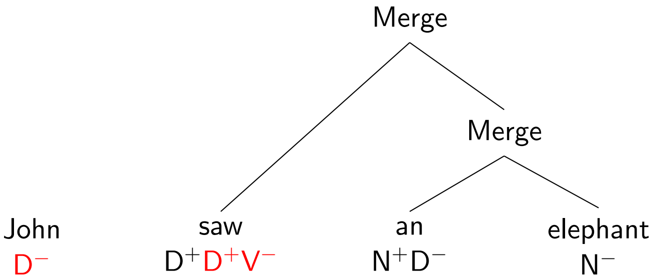
Syntax: Merge operation [cont.]

John	saw	an	elephant
D^-	$D^+D^+V^-$	N^+D^-	N^-

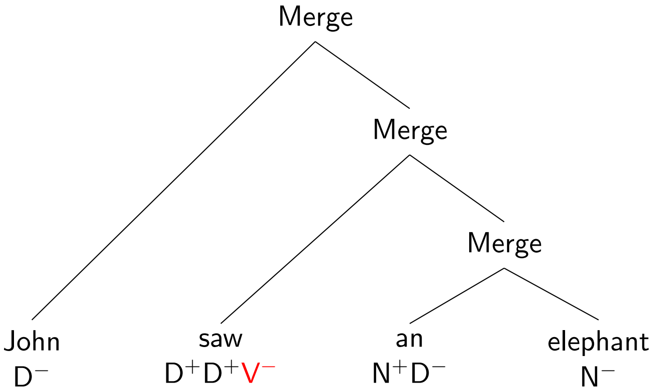
Syntax: Merge operation [cont.]



Syntax: Merge operation [cont.]

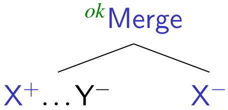


Syntax: Merge operation [cont.]



Syntax: Merge operation [cont.]

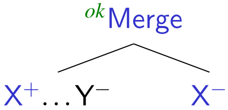
- So what's a tree bigram?



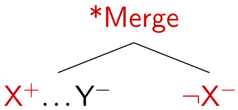
- What is the illicit tree bigram then?

Syntax: Merge operation [cont.]

- So what's a tree bigram?



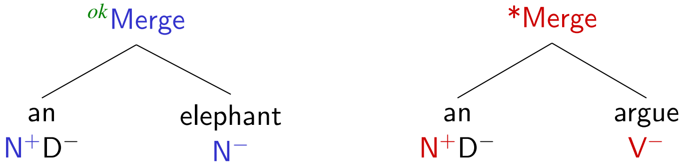
- What is the illicit tree bigram then?



Syntax: Merge operation [cont.]

- Merge operation is *strictly local over trees*. (Graf 2012)
- Move is Tier-Based Strictly Local over trees (Graf 2012)

Syntax: Merge operation [cont.]



- Merge operation is *strictly local over trees*. (Graf 2012)
- Move is Tier-Based Strictly Local over trees (Graf 2012)

The Central Role of Derivation Trees

- Derivation trees are rarely considered in generative syntax.
  (but see Epstein et al. 1998)
- satisfy Chomsky's structural desiderata:
  - no linear order
  - label-free
  - extension condition
  - inclusiveness condition
- contain all information to produce phrase structure trees
  $\Rightarrow$ **central data structure** of Minimalist syntax

Psychological Reality of Derivation Trees

Central role of derivation trees backed up by **processing data**:

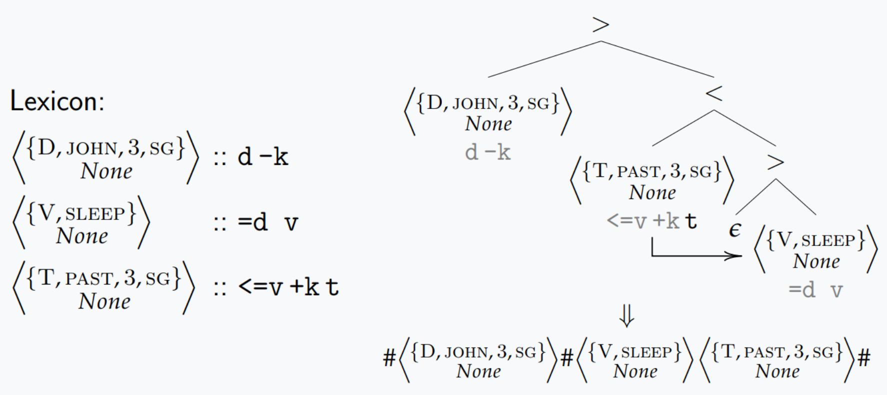
- Derivation trees can be parsed top-down (Stabler 2013)
- Parsing models update Derivational Theory of Complexity, make correct processing predictions for
  - right $<$ center embedding (Kobele et al. 2012)
  - crossing $<$ nested dependencies (Kobele et al. 2012)
  - SC-RC $<$ RC-SC (**?**)
  - SRC $<$ ORC in English (**?**)
  - SRC $<$ ORC in East-Asian (**?**)
  - quantifier scope preferences (Pasternak 2016)

Technical Fertility of Derivation Trees

Derivation trees made it easy for MGs to accommodate the full syntactic toolbox:

- sidewards movement (Stabler 2006; Graf 2013)
- affix hopping (Graf 2012b, 2013)
- clustering movement (Gärtner and Michaelis 2010)
- tucking in (Graf 2013)
- ATB movement (Kobele 2008)
- copy movement (Kobele 2006)
- extraposition (Hunter and Frank 2014)
- Late Merge (Kobele 2010; Graf 2014a)
- Agree (Kobele 2011; Graf 2012a)
- adjunction (Fowlie 2013; Graf 2014b; Hunter 2015)
- TAG-style adjunction (Graf 2012c)

DM is a Regular Relation (Ermolaeva & Edmiston 2018)

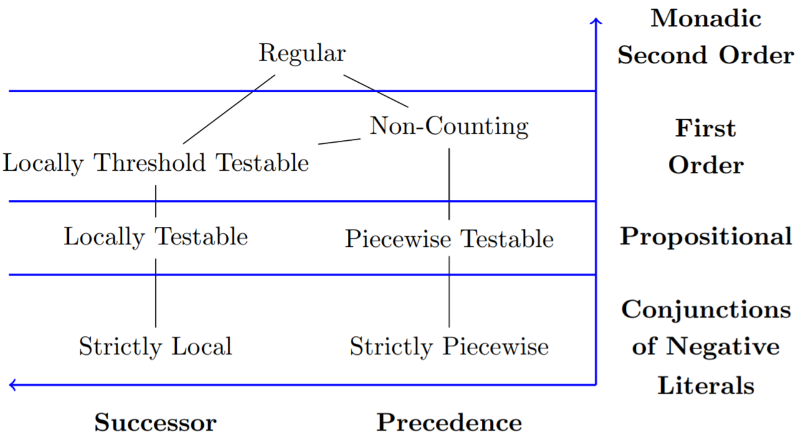


pic from Ermolaeva & Edmiston 2018

Strong Cognitive Parallelism Hypothesis

Phonology, morphology, and syntax have the **same subregular complexity** over their respective **structural representations**.

Subregular Hierarchy



pic: Heinz 2018

References I

Applegate, R.B. 1972. *Ineseno chumash grammar*. Doctoral Dissertation, University of California,Berkeley.

Epstein, Samuel D., Erich M. Groat, Ruriko Kawashima, and Hisatsugu Kitahara. 1998. *A derivational approach to syntactic relations*. Oxford: Oxford University Press.

Fowlie, Meaghan. 2013. Order and optionality: Minimalist grammars with adjunction. In *Proceedings of the 13th Meeting on the Mathematics of Language (MoL 13)*, ed. András Kornai and Marco Kuhlmann, 12–20.

Gärtner, Hans-Martin, and Jens Michaelis. 2010. On the treatment of multiple-wh-interrogatives in Minimalist grammars. In *Language and logos*, ed. Thomas Hanneforth and Gisbert Fanselow, 339–366. Berlin: Akademie Verlag.

Graf, Thomas. 2012a. Locality and the complexity of Minimalist derivation tree languages. In *Formal Grammar 2010/2011*, ed. Philippe de Groot and Mark-Jan Nederhof, volume 7395 of *Lecture Notes in Computer Science*, 208–227. Heidelberg: Springer. URL `http://dx.doi.org/10.1007/978-3-642-32024-8_14`.

Graf, Thomas. 2012b. Movement-generalized Minimalist grammars. In *LACL 2012*, ed. Denis Béchet and Alexander J. Dikovsky, volume 7351 of *Lecture Notes in Computer Science*, 58–73. URL `http://dx.doi.org/10.1007/978-3-642-31262-5_4`.

References II

Graf, Thomas. 2012c. Tree adjunction as Minimalist lowering. In *Proceedings of the 11^{th} International Workshop on Tree Adjoining Grammars and Related Formalisms (TAG+11)*, 19–27.

Graf, Thomas. 2013. *Local and transderivational constraints in syntax and semantics*. Doctoral Dissertation, UCLA. URL `http://thomasgraf.net/doc/papers/PhDThesis_RollingRelease.pdf`.

Graf, Thomas. 2014a. Late merge as lowering movement in Minimalist grammars. In *LACL 2014*, ed. Nicholas Asher and Sergei Soloviev, volume 8535 of *Lecture Notes in Computer Science*, 107–121. Heidelberg: Springer.

Graf, Thomas. 2014b. Models of adjunction in Minimalist grammars. In *Formal Grammar 2014*, ed. Glynn Morrill, Reinhard Muskens, Rainer Osswald, and Frank Richter, volume 8612 of *Lecture Notes in Computer Science*, 52–68. Heidelberg: Springer.

Hunter, Tim. 2015. Deconstructing merge and move to make room for adjunction. *Syntax* 18:266–319.

Hunter, Tim, and Robert Frank. 2014. Eliminating rightward movement: Extraposition as flexible linearization of adjuncts. *Linguistic Inquiry* 45:227–267.

Kobele, Gregory M. 2006. *Generating copies: An investigation into structural identity in language and grammar*. Doctoral Dissertation, UCLA. URL `http://home.uchicago.edu/~gkobele/files/Kobele06GeneratingCopies.pdf`.

References III

Kobele, Gregory M. 2008. Across-the-board extraction and Minimalist grammars. In *Proceedings of the Ninth International Workshop on Tree Adjoining Grammars and Related Frameworks*.

Kobele, Gregory M. 2010. On late adjunction in Minimalist grammars. Slides for a talk given at MCFG+ 2010.

Kobele, Gregory M. 2011. Minimalist tree languages are closed under intersection with recognizable tree languages. In *LACL 2011*, ed. Sylvain Pogodalla and Jean-Philippe Prost, volume 6736 of *Lecture Notes in Artificial Intelligence*, 129–144.

Kobele, Gregory M., Sabrina Gerth, and John T. Hale. 2012. Memory resource allocation in top-down Minimalist parsing. In *Proceedings of Formal Grammar 2012*.

Pasternak, Robert. 2016. Memory usage and scope ambiguity resolution. Qualifying paper, Stony Brook University.

Stabler, Edward P. 2006. Sidewards without copying. In *Formal Grammar '06, Proceedings of the Conference*, ed. Gerald Penn, Giorgio Satta, and Shuly Wintner, 133–146. Stanford: CSLI.

Stabler, Edward P. 2013. Two models of minimalist, incremental syntactic analysis. *Topics in Cognitive Science* 5:611–633.