Patterns and factors in natural systems

E. Stabler, EALING, ENS Paris, 2012

1. Human languages: what they are, how we can study them
   - from the Chomsky hierarchy to linguistic theory
2. Factored grammars and models of language recognition
   - derivation, spellout, agreement
3. A factored model of birdsong
   - HLs and birdsong are non-FS, non-CF, non-MCS

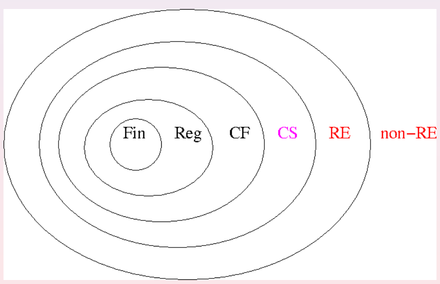
(Chomsky, 1965)

- Identify factors underlying behavior
- Universal regularities hide in irregular details
  (universals restricting the possible languages)
- Linguistic structure not a mirror of thought
- Grammar as some kind of recursive system
- 'Implementation' of grammar may be nontrivial

* Causal factors in language structure restricted

* The *kind* of recursive system is a language universal

Kinds of recursive systems



grammar rules

Fin: $S \to a_0 \ldots a_n$

Reg: $A \to aB, A \to \epsilon$

CF: $A \to X$

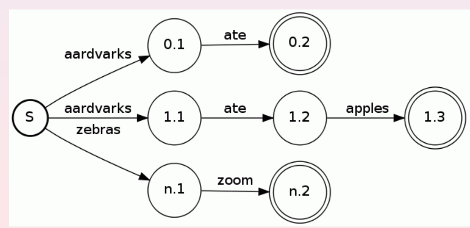
CS: $X \to Y$ with $|X| < |Y$

RE: $X \to Y$

Fin: $S \to a_0 \ldots a_n$, a list

Example: S → aarvarks ate
S → aarvarks ate apples
S → zebras zoom

Each such grammar G corresponds to a 'list automaton'



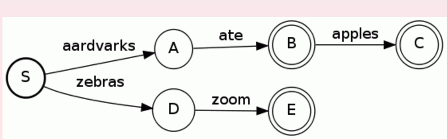
This is a simple kind of 'finite state automaton', FSA.

Fin: $A \to aB, A \to \epsilon$ without cycles

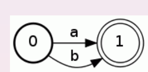
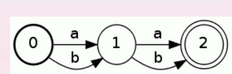
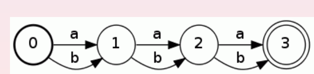
- FSA state A is cyclic if, from A, A can be reached in ≥ 1 steps
- Category A is recursive iff, from A, . . . A. . . derivable in ≥ 1 steps.

(cf Aho&Ullman'72,p153)

Example: S → aardvarks A　S → zebras D
A → ate B　D → zoom E
B → apples C　E → ϵ
B → ϵ
C → ϵ



L(list fsa) = L(acyclic fsa)　('weakly' equivalent)
But the machines/grammars are different!

|sentence|	acyclic fsa	size	la size
1		2	3
2		3	9
3		4	25
n		n+1	$n2^n + 1$

Diffs in size of minimal, weak equivalents can be exponential.

Why English $\notin$ Fin

- English has infinitely many sentences.
  For any s ∈English, I can create a longer one.
  Worry: This arg only works for precise concepts like *prime number*, not for concepts with indefinite boundaries like *heap* or *bald person*
- '[T]he assumption that language is infinite is made for the purpose of simplifying the description' (Chomsky, 1956).
  Of course, *any* model of human cognition will make simplifications, and thus be inadequate in certain ways. . . (Fitch and Frederici, 2012)
  Worry: Apparently contradicts first idea, raises questions about what 'simplifications' are allowed.

Why English $\notin$ Fin

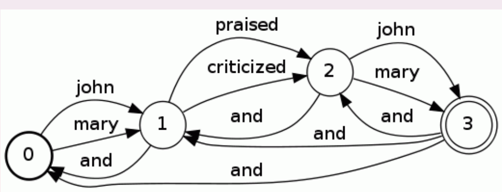
(Chomsky'65,p3) Linguistic theory is concerned primarily with an ideal speaker-hearer . . . unaffected by such grammatically irrelevant conditions as memory limitations. . .

A two-factor account:

- English language recognizers have recursive, cyclic categories, so the language is infinite.
- English language recognizers have finite lifetimes, attention, but the factors influencing this are irrelevant to the factors determining linguistic structure.

Yes, our account of English is incomplete and incorrect in various ways, but these claims are not approximate, but true, and commonplace.
Cf. claims about a pendulum, or a bodily organ.

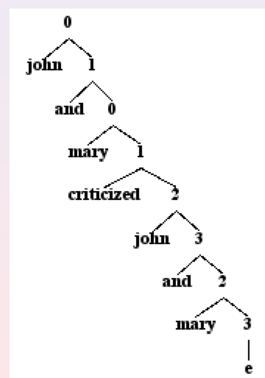
Reg: $A \rightarrow aB, A \rightarrow \epsilon$ (mem finite, but recursion allowed!)



size(A)=4 but $|L(A)|=\infty$

$0 \rightarrow$ john 1
$0 \rightarrow$ mary 1
$1 \rightarrow$ and 0
$1 \rightarrow$ praised 1
$1 \rightarrow$ criticized 1
$2 \rightarrow$ and 1
$2 \rightarrow$ john 3
$2 \rightarrow$ mary 3
$3 \rightarrow \epsilon$
$3 \rightarrow$ and 2
$3 \rightarrow$ and 1
$3 \rightarrow$ and 0

(compare CF)



$0 \rightarrow$ john 1
$0 \rightarrow$ mary 1
$1 \rightarrow$ and 0
$1 \rightarrow$ praised 1
$1 \rightarrow$ criticized 1
$2 \rightarrow$ and 1
$2 \rightarrow$ john 3
$2 \rightarrow$ mary 3
$3 \rightarrow \epsilon$
$3 \rightarrow$ and 2
$3 \rightarrow$ and 1
$3 \rightarrow$ and 0

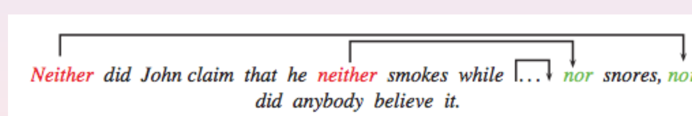
Eng $\notin$ Reg

(Fitch&al'12) Among linguists, psycholinguists and computer scientists today, the supra-regular hypothesis is nearly universally accepted. . . It has become a truism that natural language requires supra-regular resources, which are thus presumed to be present in some form in the human mind and implemented by human brains: this is not an issue debated in the recent literature. It is thus a peculiar historical fact that, until very recently, neither neuroscientists nor experimental and animal psychologists have shown any interest in this issue. (p1927)

(¬RegDistinctness) Supra-regular distinctiveness hypothesis: . . . humans are unusual (or perhaps unique) in possessing supra-regular processing power. (p1929)

Why English $\notin$ Reg

(Jäger&Rogers'12) The crucial insight here is that English has *center embedding* constructions. . .



As far as the grammar of English goes, there is no fixed upper bound on the number of levels of embedding. Consequently, English grammar allows for a potentially *unlimited number* of nested dependencies of *unlimited size*.

Shaky evidence! Why not assume a fuzzy boundary around 2 embeddings, with 3 marginal, but 10 out?

Why AGL language $\notin$ Reg

By testing the learner on examples of an artificial language (A^nB^n) with 2 embeddings, can we tell if the learner has acquired a grammar allowing more than 2?

(Jäger&Rogers'12) To get evidence that the learner has done this, one needs to include strings [allowing more than 2 embeddings]. . . [Tests of between 1 and 3 embeddings] seems very near the boundary of practicality for most experiments involving living organisms.

- 3 embeddings of "neither. . . nor" very marginal
- Independently, the bounds are obviously finite

When does a physical computer use a language $\notin$ Reg?

The answer is not in your formal language theory text that I have ever seen!

(Hopcroft&Ullman'79,p14) The computer itself can be viewed as a finite state system, although doing so turns out to be not as useful as one would like.

(Gurevich'88,p412) . . . the classical theory of finite-state machines is not adequate to deal with real computers. . . there are too many states.

. . . assume infinite when it's useful? When is that?

Why English $\notin$ Reg

(Fitch&Friederici'12,p1940) Let us assume that human language use *could* be modeled by a FSA, augmented with transition probabilities. . . such a Markov grammar would require an enormous number of parameters.

- not necessarily enormous unless each is independent of others
- why assume probabilistic grammars with infinitely many require fewer parameters?

Why English $\notin$ Reg

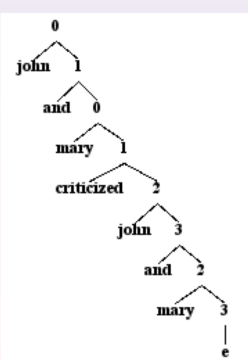
(Chomsky'65,p3) Linguistic theory is concerned primarily with an ideal speaker-hearer . . . unaffected by such grammatically irrelevant conditions as memory limitations. . .

- influences on analysis $\neq$ influences on memory limitations

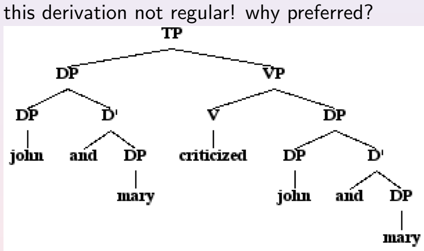
(Miller'67) constituent structure languages are more natural, easier to cope with, than regular languages. . . The hierarchical structure of strings generated by constituent-structure grammars is characteristic of much other behavior that is sequentially organized; it seems plausible that it would be easier for people than would the left-to-right organization characteristic of strings generated by regular grammars

- Assumption that memory is structured and accessed in certain ways makes predictions unavailable in models that ignore that structure.

Miller quote = Fitch&al's 'supra-regular' hypothesis (p1927). These better ideas require change in question.

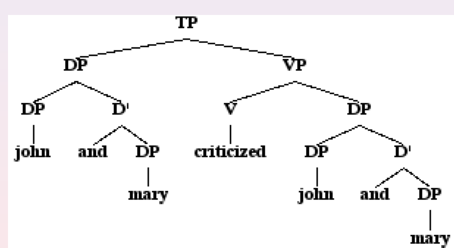


this derivation not regular! why preferred?



- same DPs in diff positions
- new name $\Rightarrow$ both positions
- boundary effects (e.g. click)
- semantic compositionality
- binding theory, etc. . .

CF: $A \to X$ (context free grammar)



TP $\to$ DP VP
VP $\to$ V DP
V $\to$ criticized
DP $\to$ DP D'
D' $\to$ and DP
DP $\to$ john
DP $\to$ mary

This grammar G not regular, but L(G) is. (simpler than Reg)

Not simpler in 'false but useful approximation' sense. Simpler in a way that makes it more believable.

(Fitch&al'12) Among linguists, psycholinguists and computer scientists today, the supra-regular hypothesis is nearly universally accepted... It has become a truism that natural language requires supra-regular resources, which are thus presumed to be present in some form in the human mind and implemented by human brains: this is not an issue debated in the recent literature. It is thus a peculiar historical fact that, until very recently, neither neuroscientists nor experimental and animal psychologists have shown any interest in this issue. (p1927)

(H1) G(English) $\notin$ Reg $\Leftarrow$ clear!

(H2) L(English) $\notin$ Reg

The strong evidence for H2 is not our judgments about *neither-nor* and other center-embeddings, but H1.

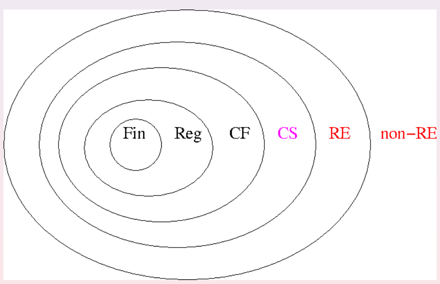
ARGUMENT SKETCH: Notice that the structure of the embedded clause looks very similar to the simple sentence:

the claim that [the claim is funny] is funny
the claim is funny

If we say it is the same, then L(English) $\notin$ Reg.

exercise: spell out that argument carefully

Kinds of recursive systems



grammar rules

Fin: $S \to a_0 \dots a_n$

Reg: $A \to aB, A \to \epsilon$

CF: $A \to X$

CS: $X \to Y$ with $|X| < |Y$

RE: $X \to Y$

- 3 factors: G(HL) + memory limits + lifespan
- G(HL) non-Reg, supporting, for example, L(English) non-Reg

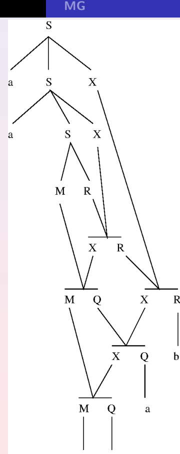
These claims are not approximate, but true, and commonplace. Cf. a computer parses arithmetic expressions.

CS: $X \to Y$ with $|X| < |Y|$

$S \to aSX$	$S \to bSY$	$S \to LQ$	$S \to MR$
$QX \to XQ$	$RX \to XR$	$QY \to YQ$	$RY \to YR$
$LX \to LQ$	$MX \to MQ$	$LY \to LR$	$MY \to MR$
$L \to a$	$M \to b$	$Q \to a$	$R \to b$

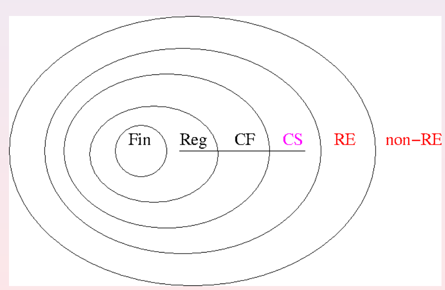
This defines the non-CF language $\{xx|\ x \in \{a,b\}^+\}$ (Mateescu and Salomaa, 1997)

Deriving aabaab:

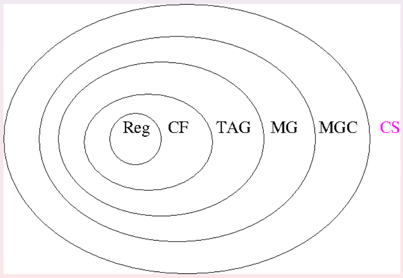


These grammars can define difficult languages, and they are difficult to reason about.

Kinds of recursive systems: refining the area of interest



Kinds of recursive systems: refining the area of interest



Reg: $A \to aB, A \to \epsilon$

CFG: $A \to X$

TAG: tree adjoining grammar

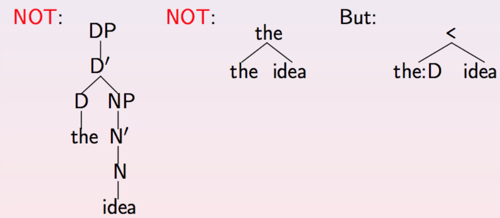
MG: minimalist grammar

MGC:MG with copying

CS: $X \to Y$ with $|X| < |Y|$

Joshi'85 (MCS): HLs are weakly and strongly 'mildly context sensitive'

'minimalist' grammars (MG)



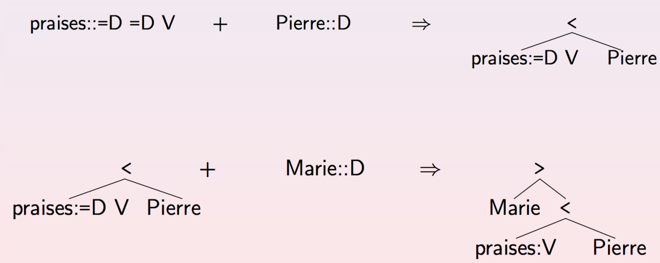
The < "points toward" the **head** of the phrase.
The largest subtree with a given head is a **maximal** projection.

every,some,student,...	(vocabulary)
C, T, D, N, V, P,...	(categories)
=C, =T, =D, =N, =V, =P,...	(selectors)
+wh, +case, +focus,...	(licensors)
-wh, -case, -focus,...	(licensees)

Examples: **Marie::D**
who::D -wh
praises::=D =D V
ϵ::=I +wh C

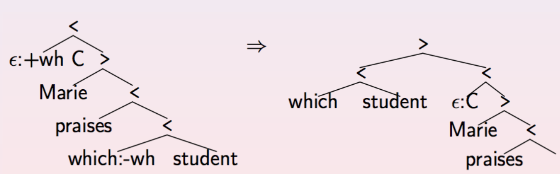
These lexical items combined by *merge*...

External merge (•) first merge on right, then on left



(2 features deleted, and :: in lexical items changes to : in derived structures)

Internal merge (○) in a tree whose head has first feature +f, move maximal -f subtree specifier position:

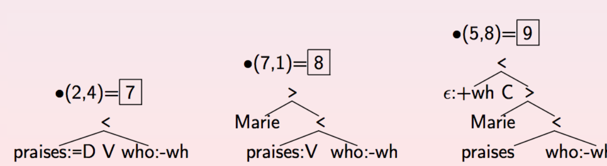


(SMC) ○ applies only when exactly 1 head has -f first feature

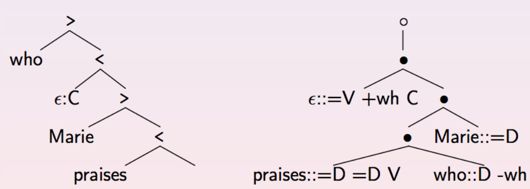
example grammar:

0	`Pierre::D`	`who::D -wh`	4
1	`Marie::D`	`ε::=V +wh C`	5
2	`praises::=D =D V`	`know::=C =D V`	6
3	`ε::=V C`		

steps 1,2,3



step 4 **all 4 steps**

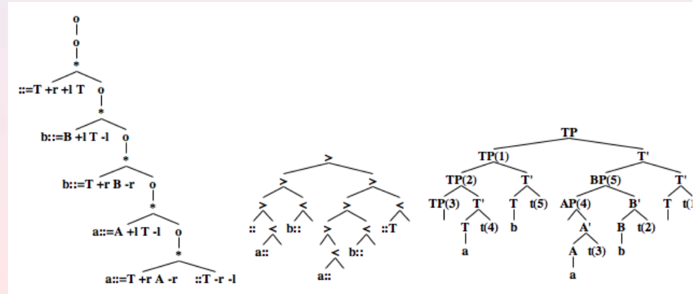


derived tree **derivation tree**

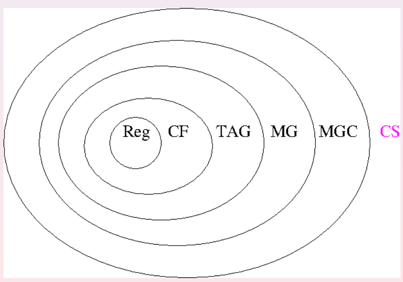
⇒ 2 steps: derivation on right mapped to derived tree (Kobele&al'07)
Cf. Chomsky'12: merge assembles; checked at interfaces (2s)

A grammar for the non-CF language $\{xx | x \in \{a,b\}^*\}$

ϵ::T −r −l	ϵ::=T +r +l T
a::=T +r A −r	b::=T +r B −r
a::=A +l T −l	b::=B +l T −l



Kinds of recursive systems, areas of interest refined



Reg: $A \to aB, A \to \epsilon$

CFG: $A \to X$

TAG: tree adjoining grammar

MG: ck+spellout

MGC:MG with copying

Claims so far:

- Distinguish G(HL) from L(HL). G(HL) of primary interest.
- 3 distinct factors: G(HL) + memory limits + lifespan
- G(HL) $\notin$ Reg, supporting L(English) $\notin$ Reg
- G(HL) has distinct factors: derivation, spellout

Next:

- G(HL) $\notin$ (M)CF, and more factors
- (¬RegDistinctness) rejected: many animals possess supra-regular processing power

APPENDICES

A. Making the 2 steps explicit

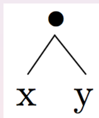
(Kobele, Retoré, and Salvati, 2007)

0. form derivation (trivial!)
1. check derivation (at interfaces?) (FS dbutt)
2. map to PF/LF (FS dmbutt)

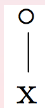
nb: There are many ways to do 0,1,2, and they can be interleaved.
E.g. 0 could proceed until phase boundary reached, then 1 and 2.

0. form derivation: merge

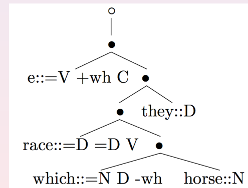
Form derivation by combining any two lexical items or derivations:



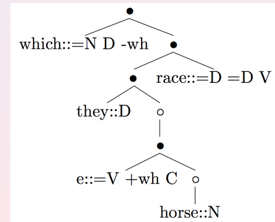
or by 'combining' with something already in the derivation:



good derivation



bad derivation

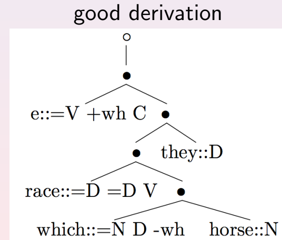


1. check derivation: 'deterministic bottom-up tree transducer'

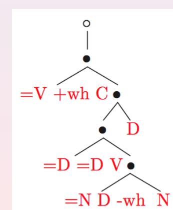
- At the leaves, let the 'state' be the features of the lexical items.
- Compute 'state' of internal nodes by feature checking.
- Derivation is good iff the state of the root is *C* (or whatever counts as an acceptable phrasal category)

Checking the good derivation:

good derivation

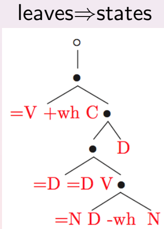


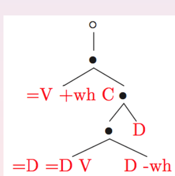
leaves⇒states



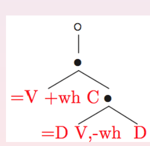
Checking the good derivation:

leaves⇒states now up the tree...





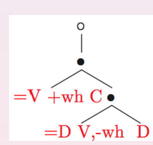
DP features



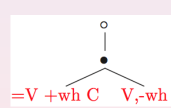
V' features

Checking the good derivation:

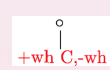
continuing...



V'features



VP features

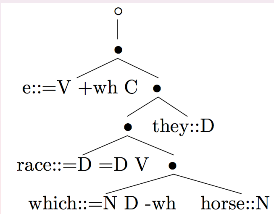


C' features

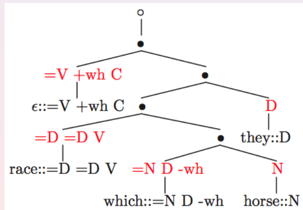
CP feature success!

2. (Ck tree in course of) mapping to PF/LF

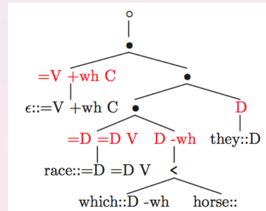
good derivation



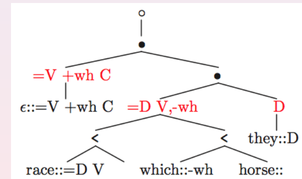
leaves⇒states



now up the tree. . .

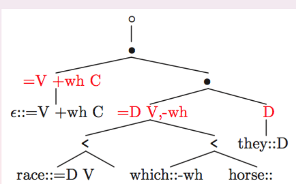


DP features

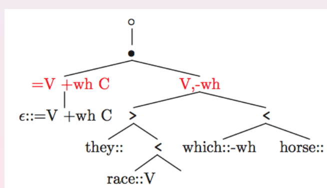


V' features

continuing. . .

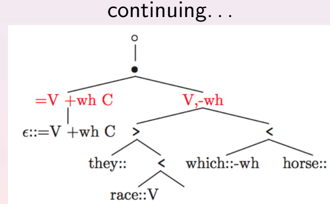


V' features

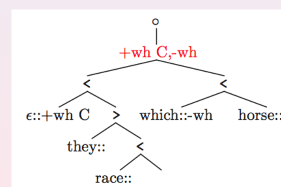


VP features

continuing. . .

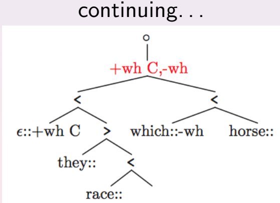


VP features

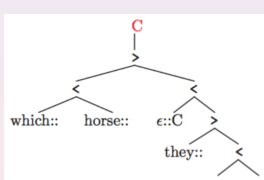


C' features

continuing. . .



C' features



CP features
success!

(back)

B. Chomsky'65 on method

1. Human languages: what they are, how we can study them

(Chomsky, 1965)

- Identify factors underlying linguistic behavior. . .

(p5) The study actual linguistic performance, we must consider the interaction of a variety of factors, of which the underlying competence of the speaker is only one. In this respect, study of language is no different from empirical investigation of other complex phenomena.

(Chomsky, 1965)

- Identify factors underlying behavior
- universal regularities hide in irregular details

(p6) [Although] traditional and structuralist grammars. . . may contain full and explicit lists of exceptions and irregularities, they provide only examples and hints concerning the regular and productive syntactic processes . . . [It] is to be supplemented by a universal grammar that . . . expresses the deep-seated regularities which, being universal, are omitted from the grammar itself.

(Chomsky, 1965)

- Identify factors underlying behavior
- Universal regularities hide in irregular details

(p6) . . . The problem for the linguist, as well as for the child learning the language, is to determine from the data of performance the underlying system of rules that has been mastered by the speaker-hearer and that he puts to use in actual performance.

(Chomsky, 1965)

- Identify factors underlying behavior
- Universal regularities hide in irregular details

  (universals restricting the possible languages)

(1981,p10) In the general case of theory construction, the primitive basis can be selected in any number of ways,. . . But in the case of UG . . . we want the primitives to be concepts that can plausibly be assumed to provide a preliminary, prelinguistic analysis of a reasonable selection of presented data, that is, to provide the data that are mapped by the language faculty to a grammar. . . [Restricting] the class of grammars mad accessible in principle by UG. . . to account for the fact that knowledge of language is acquired on the basis of the evidence available.

(Chomsky, 1965)

- Identify factors underlying behavior
- universal regularities hide in irregular details

  (universals restricting the possible languages)
- Linguistic structure not a mirror of thought

(p6) [A] reason for the failure of traditional grammars, particular or universal, to attempt a precise statement of regular processes of sentence formation and sentence interpretation lay in the widely held belief that there is a "natural order of thoughts" that is mirrored by the order of words. Hence, the rules of sentence formation do not really belong to grammar but to some other subject in which the "order of thoughts is studied."

(Chomsky, 1965)

- Identify factors underlying behavior
- Universal regularities hide in irregular details,

  (universals restricting the possible languages)
- Linguistic structure not a mirror of thought
- Grammar as some kind of recursive system

(p6) But the fundamental reason for this inadequacy of traditional grammars is a more technical one. . . the technical devices for expressing a system of recursive processes were simply not available until much more recently.

(Chomsky, 1965)

- Identify factors underlying behavior
- Universal regularities hide in irregular details

  (universals restricting the possible languages)
- Linguistic structure not a mirror of thought
- Grammar as some kind of recursive system
- 'Implementation' in performance model may be nontrivial

(p9) [A] generative grammar is not a model for a speaker or a hearer. It attempts to characterize in the most neutral possible terms the knowledge of the language that provides the basis for actual use of language by a speaker-hearer. . . [It] does not, in itself, prescribe the character or functioning of a perceptual model or a model of speech production.

Aho, Alfred V. and Jeffrey D. Ullman. 1972. *The Theory of Parsing, Translation, and Compiling. Volume 1: Parsing.* Prentice-Hall, Englewood Cliffs, New Jersey.

Chomsky, Noam. 1956. Three models for the description of language. *IRE Transactions on Information Theory*, IT-2:113–124.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax.* MIT Press, Cambridge, Massachusetts.

Chomsky, Noam. 1981. *Lectures on Government and Binding.* Foris, Dordrecht.

Chomsky, Noam. 2012. Problems of projection. *Lingua*, forthcoming.

Fitch, W. Tecumseh and Angela D. Frederici. 2012. Artificial grammar learning meets formal language theory: An overview. *Philosophical Transactions of the Royal Society B*, 367:1933–1955.

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.

Gurevich, Yuri. 1988. Algorithms in the world of bounded resources. In Rolf Herkin, editor, *The Universal Turing Machine: A Half-Century Survey.* Oxford University Press, Oxford, pages 407–416.

Hopcroft, John E. and Jeffrey D. Ullman. 1979. *Introduction to Automata Theory, Languages and Computation.* Addison-Wesley, Reading, Massachusetts.

Jäger, Gerhard and James Rogers. 2012. Formal language theory: Refining the Chomsky hierarchy. *Philosophical Transactions of the Royal Society B*, 367:1956–1970.

Joshi, Aravind K. 1985. How much context-sensitivity is necessary for characterizing structural descriptions. In D. Dowty, L. Karttunen, and A. Zwicky, editors, *Natural Language Processing: Theoretical, Computational and Psychological Perspectives.* Cambridge University Press, NY, pages 206–250.

Kobele, Gregory M. 2006. *Generating Copies: An Investigation into Structural Identity in Language and Grammar.* Ph.D. thesis, UCLA.

Kobele, Gregory M., Christian Retoré, and Sylvain Salvati. 2007. An automata-theoretic approach to minimalism. In James Rogers and Stephan Kepser, editors, *Model Theoretic Syntax at 10. ESSLLI'07 Workshop Proceedings.*

Mateescu, Alexandru and Arto Salomaa. 1997. Aspects of classical language theory. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages, Volume 1: Word, Language, Grammar.* Springer, NY, pages 175–251.

Michaelis, Jens. 2001. *On Formal Properties of Minimalist Grammars.* Ph.D. thesis, Universität Potsdam. *Linguistics in Potsdam 13*, Universitätsbibliothek, Potsdam, Germany.

Stabler, Edward P. 1997. Derivational minimalism. In Christian Retoré, editor, *Logical Aspects of Computational Linguistics.* Springer-Verlag (Lecture Notes in Computer Science 1328), NY, pages 68–95.

Stabler, Edward P. 2010. Computational perspectives on minimalism. In Cedric Boeckx, editor, *Oxford Handbook of Minimalism.* Oxford University Press, Oxford, pages 617–641.