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
   • HIs 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
* Causal factors in language structure restricted
* The kind of recursive system is a language universal

Kinds of recursive systems

Fin:
\[
S \rightarrow a_0 \ldots a_n
\]
Reg:
\[
A \rightarrow aB, A \rightarrow \epsilon
\]
CF:
\[
A \rightarrow X
\]
CS:
\[
X \rightarrow Y \text{ with } |X| < |Y|
\]
RE:
\[
X \rightarrow Y
\]

Example:
\[
S \rightarrow \text{aardvarks ate}
S \rightarrow \text{aardvarks ate apples}
S \rightarrow \text{zebras zoom}
\]

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

Fin: \( A \rightarrow aB, A \rightarrow \epsilon \) without cycles
• FSA state A is cyclic if, from A, A can be reached in \( \geq 1 \) steps
• Category A is recursive iff, from A, \( \ldots A \ldots \) derivable in \( \geq 1 \) steps.

Example:
\[
S \rightarrow \text{aardvarks A}
A \rightarrow \text{ate B}
B \rightarrow \text{apples C}
C \rightarrow \epsilon
\]

\[
S \rightarrow \text{zebras D}
D \rightarrow \text{zoom E}
E \rightarrow \epsilon
\]

\[
L(\text{list fsa}) = L(\text{acyclic fsa})\] (‘weakly’ equivalent)
But the machines/grammars are different!

<table>
<thead>
<tr>
<th>sentence</th>
<th>acyclic fsa size</th>
<th>la size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>25</td>
</tr>
<tr>
<td>n</td>
<td>( n+1 )</td>
<td>( n^2 + 1 )</td>
</tr>
</tbody>
</table>

Diffs in size of minimal, weak equivalents can be exponential.
Why English $\not\in$ Fin

- English has infinitely many sentences. For any $s \in$ 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 $\not\in$ 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!)

\begin{align*}
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
\end{align*}

\begin{align*}
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
\end{align*}

Eng $\not\in$ 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 $\not\in$ Reg

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

Neither did John claim that he neither smokes while . . . nor snores, no
did anybody believe it.

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 $\not\in$ 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]…

[Test 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

Why English $\not\in$ 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?

When does a physical computer use a language $\not\in$ 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 $\not\in$ 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.

Why English $\not\in$ Reg

CF: $A \rightarrow X$

(context free grammar)

This grammar $G$ not regular, but $L(G)$ is.

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

Kinds of recursive systems

### grammar rules

- **Fin**: $S \rightarrow a_0 \ldots a_n$
- **Reg**: $A \rightarrow aB, A \rightarrow \epsilon$
- **CF**: $A \rightarrow X$
- **CS**: $X \rightarrow Y$ with $|X| < |Y|$
- **RE**: $X \rightarrow Y$

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.

### Deriving aabaab

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

(H1) $G(\text{English}) \not\in \text{Reg}$

(H2) $L(\text{English}) \not\in \text{Reg}$

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 $L(\text{English}) \not\in \text{Reg}$.

exercise: spell out that argument carefully

Kinds of recursive systems: refining the area of interest
Kinds of recursive systems: refining the area of interest

Reg: \( A \rightarrow aB, A \rightarrow \epsilon \)

CFG: \( A \rightarrow X \)

TAG: tree adjoining grammar

MG: minimalist grammar

MGC: MG with copying

CS: \( X \rightarrow Y \) with \(|X| < |Y|\)

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

‘minimalist’ grammars (MG)

Fin

Reg

CF

CS, MCS

MG

Reg: \( A \rightarrow aB, A \rightarrow \epsilon \)

CFG: \( A \rightarrow X \)

TAG: tree adjoining grammar

MG: minimalist grammar

MGC: MG with copying

CS: \( X \rightarrow Y \) with \(|X| < |Y|\)

The \(<\) ‘points toward’ the head of the phrase. The largest subtree with a given head is a maximal projection.

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

\[ \text{praises} = \text{D} \rightarrow \text{D V} + \text{Pierre} \rightarrow \text{D} \rightarrow \text{V} \rightarrow \text{Pierre} \]

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

example grammar:

0

Pierre::D

who::D -wh

praises::=D =D V

c::=l +wh C

steps 1,2,3

1

Marie::D

2

praises::=D =D V

c::=V +wh C

3

know::=C =D V

(4,5,6)

(2.4)=7

(7.1)=8

(5.8)=9

---

Examples:

Marie::D

who::D -wh

praises::=D =D V

c::=l +wh C

These lexical items combined by merge…

---

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

\[ \text{praises} = \text{D} \rightarrow \text{D V} \rightarrow \text{Pierre} \rightarrow \text{D} \rightarrow \text{V} \rightarrow \text{Pierre} \]

(SMC) ○ applies only when exactly 1 head has -f first feature
Introduction, goals, method, alternatives
From the Chomsky hierarchy to linguistic theory
Summary

**APPENDICES**

A grammar for the non-CF language \( \{xx \mid x \in \{a, b\}^+\} \)

\[

c:: T \rightarrow r \rightarrow l \\
\begin{array}{c}
  a:: T \rightarrow r \quad A \rightarrow r \\
  b:: T \rightarrow B \rightarrow r \\
  a:: A \rightarrow T \rightarrow l \\
  b:: B \rightarrow T \rightarrow l
\end{array}
\]

E. Stabler, EALING, ENS Paris, 2012
Patterns and factors in natural systems

Kinds of recursive systems, areas of interest refined

Reg: \( A \rightarrow aB, A \rightarrow \epsilon \)

CFG: \( A \rightarrow X \)

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
- \( G(HL) \not\subseteq \text{Reg}, \) supporting \( L(English) \not\subseteq \text{Reg} \)
- \( G(HL) \) has distinct factors: derivation, spellout

Next:

- \( G(HL) \not\subseteq (M)CF, \) and more factors
- \((\neg \text{RegDistinctness}) \) rejected: many animals possess supra-regular processing power

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:

```
                                o
                                X  y
```

or by ‘combining’ with something already in the derivation:

```
                                o
                                X
```

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 if the state of the root is $C$ (or whatever counts as an acceptable phrasal category)

Checking the good derivation:

```
leaves⇒states
```

```
V' features

DP features
```

```
V' features

V features
```

Checking the good derivation:

```
continuing...
```

```
V' features

VP features
```

```
C' features

CP feature success!
```
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

(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
- 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
- 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
- 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.


