Generalized Syntactic Recurrence: Growth, Dimensionality, and Factorization

David P Medeiros

April 21, 2012

ASU/UA Linguistics Symposium

Discrete infinity

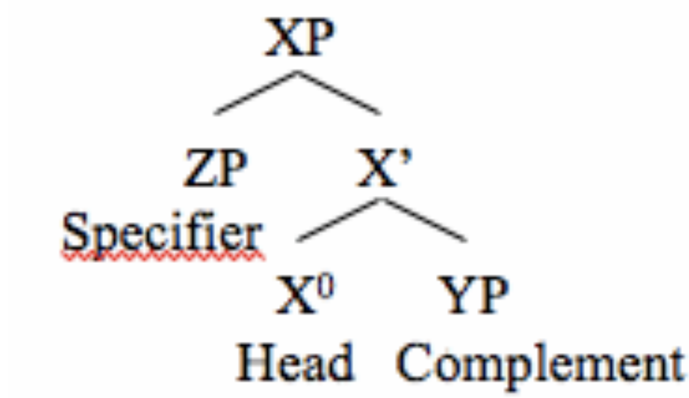
- It is an old insight that natural language is a system of "discrete infinity"
- von Humboldt's "infinite use of finite means"
- The *finite means* being discrete atomic elements: words, morphemes, features.
- And *infinite use* implicating recursion.

Studying discrete infinity

- In what follows, I report some results obtained from a study of generalized discrete infinity.
- I examine the distinct binary-branching (Kayne 1984, 1994) recurrence patterns that could form the basis for discrete infinity.
  - i.e., self-similar arrangements of terminals and non-terminals
  - Terminals being the "discrete" part, non-terminals the "infinite" part.
- Goal: describe and classify the possibilities and their properties.

A none-too-innocently-chosen example

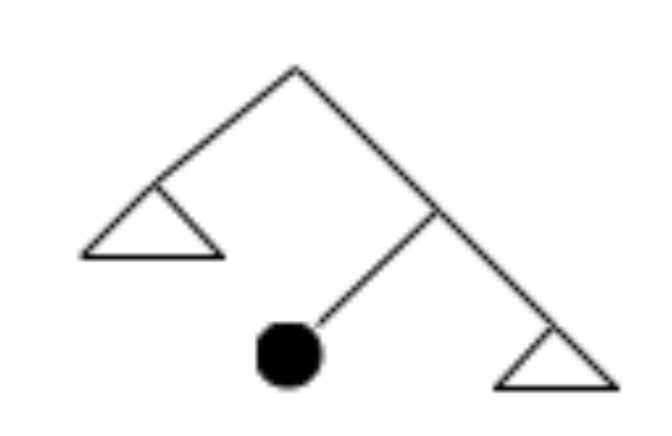
- As an illustration of one kind of pattern of recurrence, consider the familiar "X-bar schema" (Chomsky 1970, Jackendoff 1977).



- This is a "recipe" for structure building: to build a phrase (XP), combine a terminal (X^0) with a phrase (YP), then combine the result (X') with another phrase (ZP).

Bare recursion

- Of course, the X-bar schema is more than just a structure-building pattern; it also incorporates the further notion of *labeling* or *headedness*.
- In what follows, I ignore this aspect, considering only the recurrence pattern.
- On this view, the X-bar schema resolves as a simpler object that can be depicted as below:

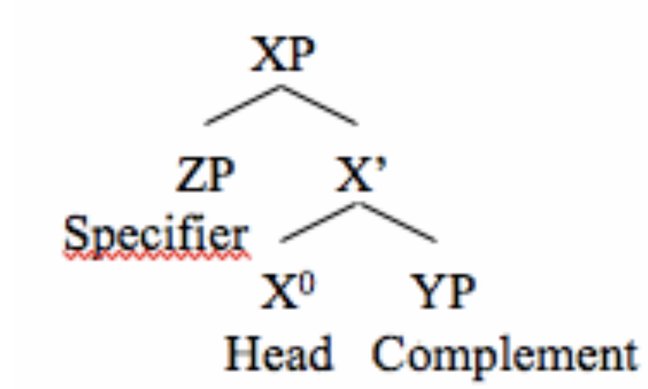


X-bar form without labels

- A traditional way to describe this particular pattern is with phrase structure rules (PSRs), as below:

  XP → ZP X'

  X' → X^0 YP



- Ignoring labels, we can write this as:

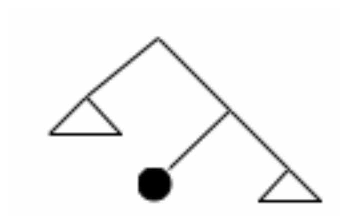
  XP → XP X'

  X' → X^0 XP

- Or, even more simply and abstractly, using 0s to represent terminals, and higher numbers (1, 2) to represent distinct kinds of non-terminals:

  2 → 2 1

  1 → 0 2



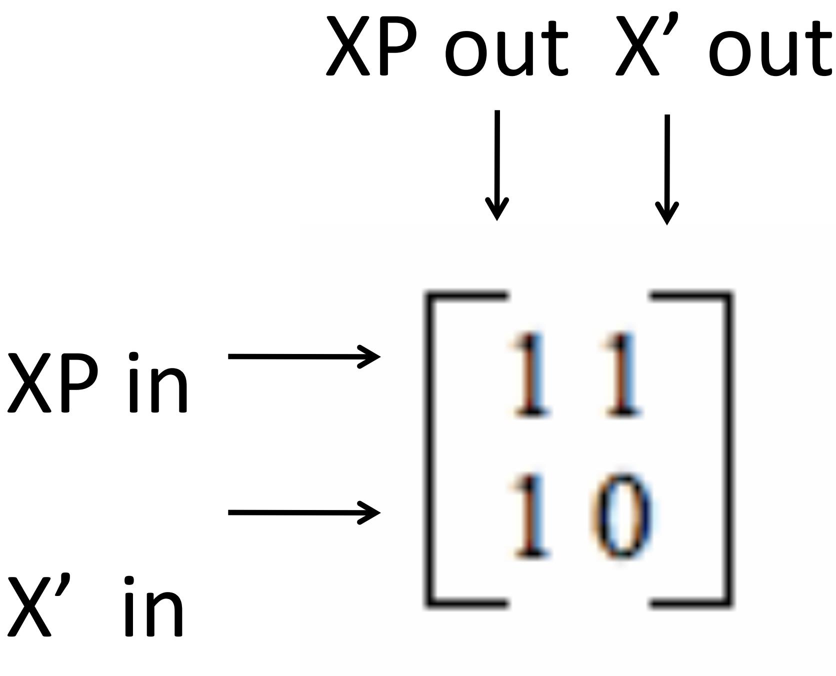
Even more abstract

- A further abstraction will help in understanding this (and other) pattern(s).
- Namely, we can represent the recurrence relations by means of a ***matrix***:

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

- The rows and columns are associated with the distinct kinds of non-terminal.
- The rows correspond to non-terminal inputs to the PSRs; the columns represent non-terminals in the output of each PSR.

Let’s make that clearer:



This representation doesn’t record linear order.

Terminals are expressed only indirectly here.

A non-terminal introduces a terminal if its associated row sums to less than 2.

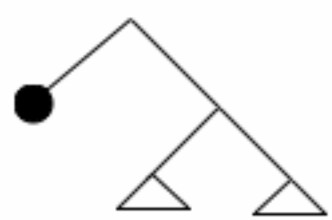
E.g. the second type of non-terminal (X’) introduces a single terminal because its row adds up to 1; the first (XP) row sums to 2, indicating it immediately dominates no terminals.

An alternative

- Let’s compare the X-bar pattern with a superficially very similar pattern,
- Which also constructs each phrase from a terminal and two further phrases.
  - X-bar: Phrase = [Phrase [terminal Phrase]]
  - **HH D-bar: Phrase = [terminal [Phrase Phrase]]**

$XP \rightarrow X^0 X'$

$X' \rightarrow XP\ XP$



$$\begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix}$$

Same ingredients, different recipe: different result.

- Consider what happens as we "inflate" these patterns, expanding them maximally:

2) 9 generations of X-bar (Phrase → [Phrase [terminal Phrase]])

3) 9 generations of HH D-Bar (Phrase → [terminal [Phrase Phrase]])

A notion of syntactic "growth"

- As we saw on the last slide, the X-bar pattern "grows" more nodes per line than the alternative (HH D-bar) as it is expanded.
- I've explored elsewhere some reasons to think that faster growth in this sense is a desirable property (all else equal); I won't review that here.
- For present purposes, let's just assume that growth in this sense is something worth investigating.
- How can we quantify this notion of growth, and what are the growth properties of the conceivable discrete infinite recurrence patterns?

Growth factor

- Intuitively, we want to find a “growth factor” G for each pattern.
- This number describes how the number of nodes on one line of the tree relates to the number of nodes on the previous line.
- We take G to be (basically) the limit of the ratio of the number of nodes on line *n*, to the number of nodes on line *n*-1, as *n* gets large.
  - Thus, nodes (n) = G * nodes (n-1)

G is the **dominant eigenvalue** of the phrasal recurrence matrix.

- Here, expressing phrasal recurrence patterns as matrices brings its first rewards.
- Matrices can be interpreted in several different ways; a natural and important interpretation is as a linear mapping.
- Under this interpretation, the *n* x *n* (square) matrices we'll be considering (expressing how *n* kinds of non-terminals link to each other), transform a point in *n*-dimensional space into another point in *n*-space.

Phrasal growth ~ iterated matrix multiplication

- Take **A** to be the relevant phrase structure matrix
- Take $\mathbf{x}_i$ to be a column vector expressing the number of each kind of non-terminal on the *i*th line of the tree.
- (we identify the non-terminals with the coordinate axes: the number of non-terminals of a given type is expressed as distance along the associated axis).
- Then maximal expansion of the pattern is simply iterated matrix multiplication.

$$\mathbf{A}\,\mathbf{x}_i = \mathbf{x}_{i+1}$$

An example

- Growth of the X-bar pattern in these terms:
- At the root, there is a single XP-type non-terminal; thus the initial vector x_0 is $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

$$\mathbf{A\, x_0 = x_1} \qquad \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

- That is, the next line (x_1) contains one XP-type non-terminal, and one X'-type.
- Continuing, we get the following sequence of vectors, representing the number of non-terminals on successive lines of the tree:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \end{bmatrix} \begin{bmatrix} 8 \\ 5 \end{bmatrix} \begin{bmatrix} 13 \\ 8 \end{bmatrix} \begin{bmatrix} 21 \\ 13 \end{bmatrix} \begin{bmatrix} 34 \\ 21 \end{bmatrix}$$

Eigenvalues and eigenvectors

- An important property of a square matrix is its set of eigenvectors and eigenvalues.
- In general, the transformation of *n*-space induced by matrix multiplication is quite complicated.
- The eigenvectors represent points of stability amidst the complexity of the mapping:
- They are the vectors that, under the transformation, retain their direction.
- i.e., for eigenvector $v = ax + by + cz\ldots$

  $\mathbf{Av} = \lambda\mathbf{v}$ $(= \lambda ax + \lambda by + \lambda cz\ldots)$

- The scaling factor λ is the eigenvalue associated with that eigenvector.

Syntactic growth is iterated mapping

- The syntactic problem we have been considering (how do phrasal patterns grow?)
- Now resolves as a geometric one:
- Given some input vector – a point in *n*-space,
- Where does that vector go as the mapping iterates?
- Thinking of things this way lets us see why G is the dominant eigenvalue.

Why G is the dominant eigenvalue

- Suppose the starting vector is *ax + by* …
- We can rewrite it in terms of the eigenvectors (a standard and powerful technique):

  $ax + by \ldots = cv_1 + dv_2 \ldots$

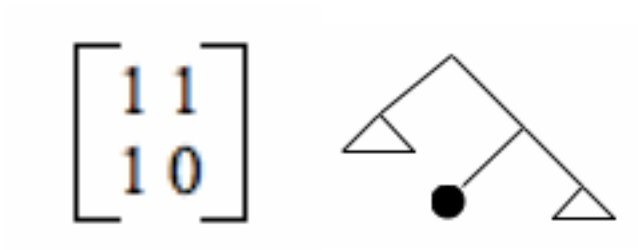
  v_i an eigenvector with eigenvalue λ_i
- Multiplication by the matrix *n* times has a particularly nice expression in terms of the eigenvectors:

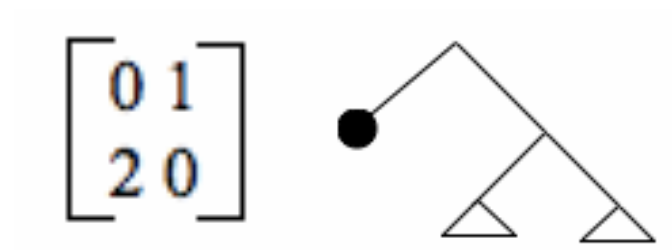
  $\boldsymbol{\lambda_1^n c v_1 + \lambda_2^n d v_2 \ldots}$
- Suppose λ_1 is the largest (i.e. dominant) eigenvalue; as *n* increases, the sum of component vectors converges on $\boldsymbol{\lambda_1^n c v_1}$ (for non-zero c).
- Thus, $x_n \sim \boldsymbol{\lambda_1} x_{\mathbf{n-1}}$; λ_1 is the desired quantity G.

Back to comparing X-bar and HH D-bar

- This insight lets us capture the difference in growth between X-bar, and the HH D-bar alternative, very simply and directly.

X-bar: HH D-bar:





G: φ ~ 1.618 √2 ~ 1.414

(the golden mean)

One more bit of math

- Associated with each square matrix is a **characteristic polynomial**.
- Among other important properties, the characteristic polynomial has as its **roots** (solutions when it's set equal to 0) the **eigenvalues** of the matrix.
- The X-bar pattern has characteristic polynomial $x^2 - x - 1$; for HH D-bar, it's $x^2 - 2$.

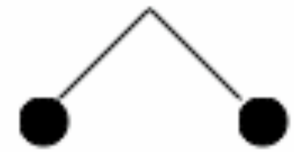
Next: cataloguing the possibilities

- With this in hand, let's turn to cataloguing the various possibilities for discrete infinite patterning.
- The possibilities are naturally partitioned by the number of non-terminal types they are defined over.
- The simplest class has a single non-terminal.

One non-terminal

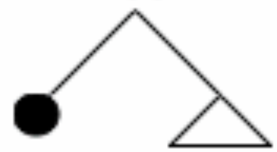
- There is really only one discrete infinite pattern with one non-terminal type: the **Spine**, below.
- The **Pair** (above right) is discrete but not infinite; the **Bush** (below right) is infinite but not discrete.

The Pair

PSR	Matrix	Tree
1 → 0 0	[0]	

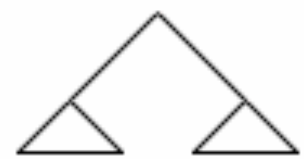
Characteristic polynomial: x – 0
Growth factor: 0

The Spine

PSR	Matrix	Tree
1 → 0 1	[1]	

Characteristic polynomial: x – 1
Growth factor: 1

The Bush

PSR	Matrix	Tree
1 → 1 1	[2]	

Characteristic polynomial: x – 2
Growth factor: 2

Two non-terminals

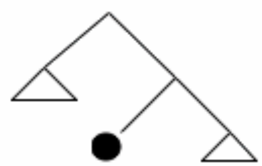
- D-bar and X-bar have "high-headed" variants -- really the same recurrence pattern, oriented differently with respect to the root.
- Again, the matrix formulation allows us to express this nicely; the related patterns have **similar** matrices, in the algebraic sense.

(41) X-bar

PSRs:
$2 \rightarrow 2\ 1$
$1 \rightarrow 0\ 2$

Matrix:
$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

Characteristic polynomial: $x^2 - x - 1$
Growth factor: $\varphi \sim 1.618$

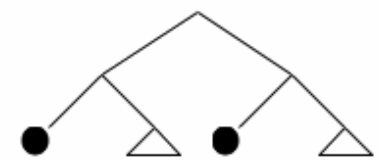


(43) D-bar

PSRs:
$2 \rightarrow 1\ 1$
$1 \rightarrow 0\ 2$

Matrix:
$$\begin{bmatrix} 0 & 2 \\ 1 & 0 \end{bmatrix}$$

Characteristic polynomial: $x^2 - 2$
Growth factor: $\sqrt{2} \sim 1.414$

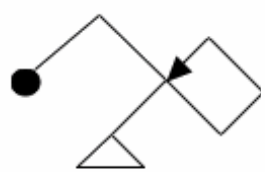


(42) High-headed X-bar

PSRs:
$2 \rightarrow 0\ 1$
$1 \rightarrow 1\ 2$

Matrix:
$$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$$

Characteristic polynomial: $x^2 - x - 1$
Growth factor: $\varphi \sim 1.618$

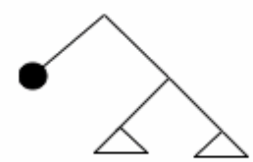


(44) High-headed D-bar

PSRs:
$2 \rightarrow 0\ 1$
$1 \rightarrow 2\ 2$

$$\begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix}$$

Characteristic polynomial: $x^2 - 2$
Growth factor: $\sqrt{2} \sim 1.414$



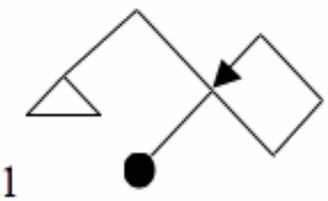
Degenerate systems

- With two non-terminals, we find our first examples of "degenerate" systems:
- Patterns that have "smaller" subpatterns (i.e., subtrees generated with less than the full set of non-terminals)

(45) Spine of Spines

PSRs:
$2 \rightarrow 2\ 1$
$1 \rightarrow 0\ 1$

Matrix:
$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

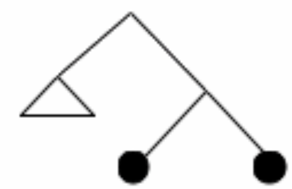


Characteristic polynomial: $x^2 - 2x + 1$
Growth factor: 1

(46) Spine of Pairs

PSRs:
$2 \rightarrow 2\ 1$
$1 \rightarrow 0\ 0$

Matrix:
$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$



Characteristic polynomial: $x^2 - x$
Growth factor: 1

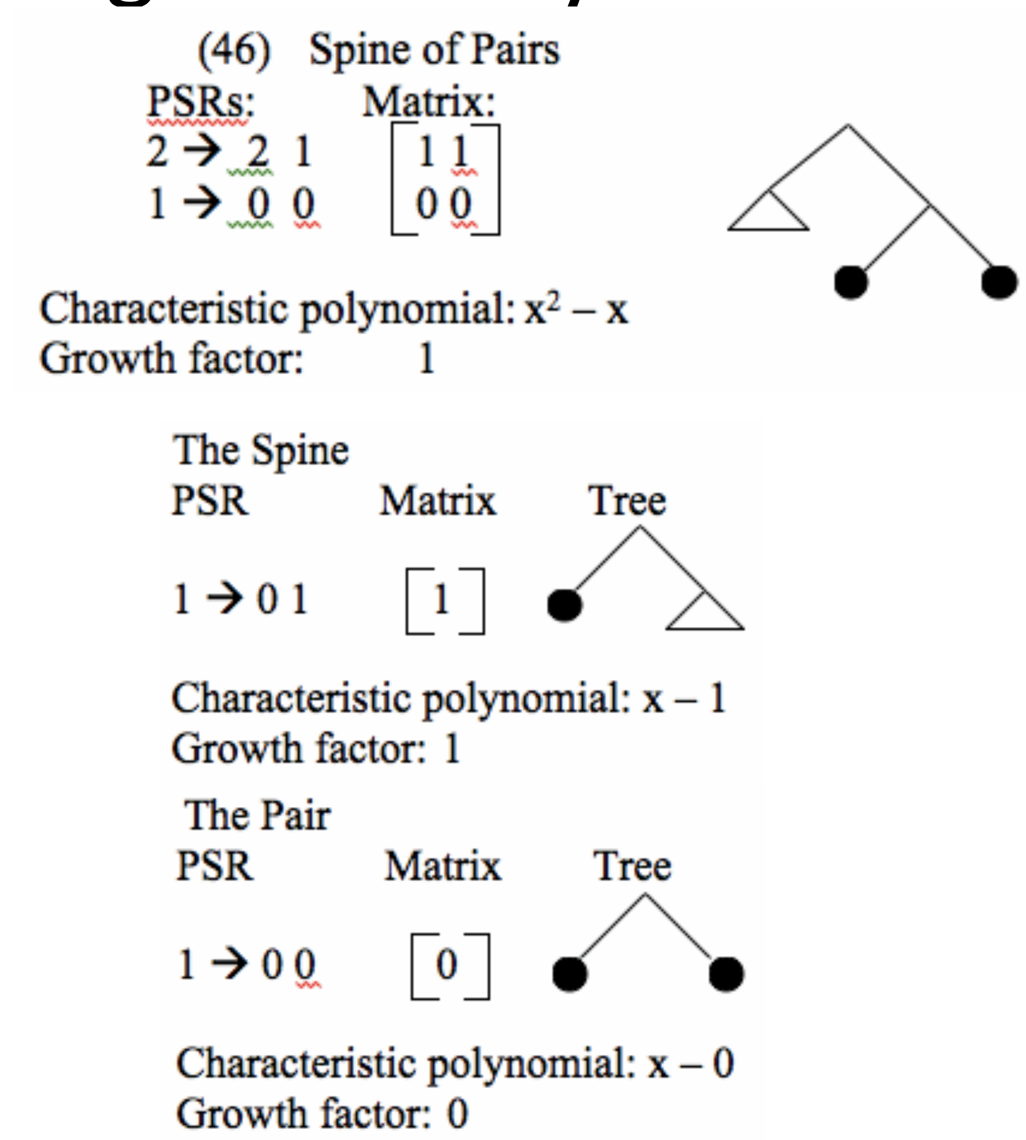
Growth in degenerate systems

- Degenerate systems are composed of simpler patterns, one substituted inside another.
- How does the growth of the larger pattern relate to the growth of its component sub-patterns?
- Here again, the matrix formulation provides the answer:
- **The growth factor of the larger pattern is just the largest of the growth factors among its components.**
- This is so, because the characteristic polynomial of a degenerate system is the product of the polynomials of its component systems.
- When multiplying polynomials, roots are preserved.

Factorization of degenerate systems

- For example, consider the Spine of Pairs (top right).
- This pattern is composed of Pairs (bottom right) substituted within a Spine (center right).
- Its polynomial is the product of the polynomials of its components:

  $x^2 - x = (x - 1)*x$
- Its roots are those of its components; G is the largest of its roots.



Dimensionality

- These are self-similar patterns.
- This invites thinking of them as *fractals*.
- Fractals are self-similar objects of non-whole number dimension.
- Their “size” depends on the scale at which they are measured
- To apply this idea to syntactic objects, we need an appropriate notion of size.
- (treating any one node as being as big as any other will lead to nonsensical results).

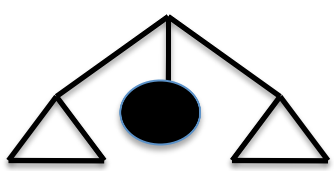
The simplest fractal: the Cantor Set

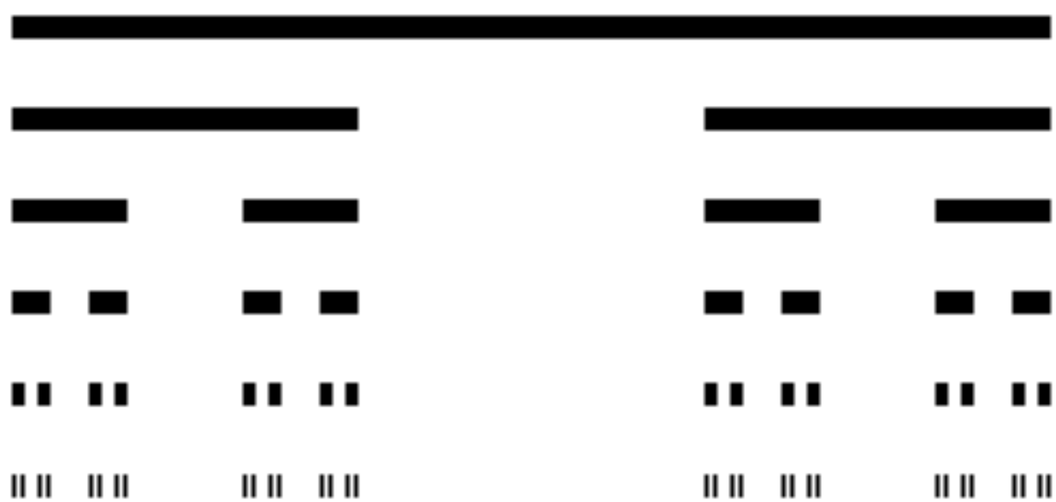
The Cantor set is formed from a line segment, by removing the middle third, then middle thirds of the remainders...

This is the *simplest* fractal:

- Background dimension cannot be lower than a 1-dimensional line.
- Division in thirds is the first scheme yielding a fractal.

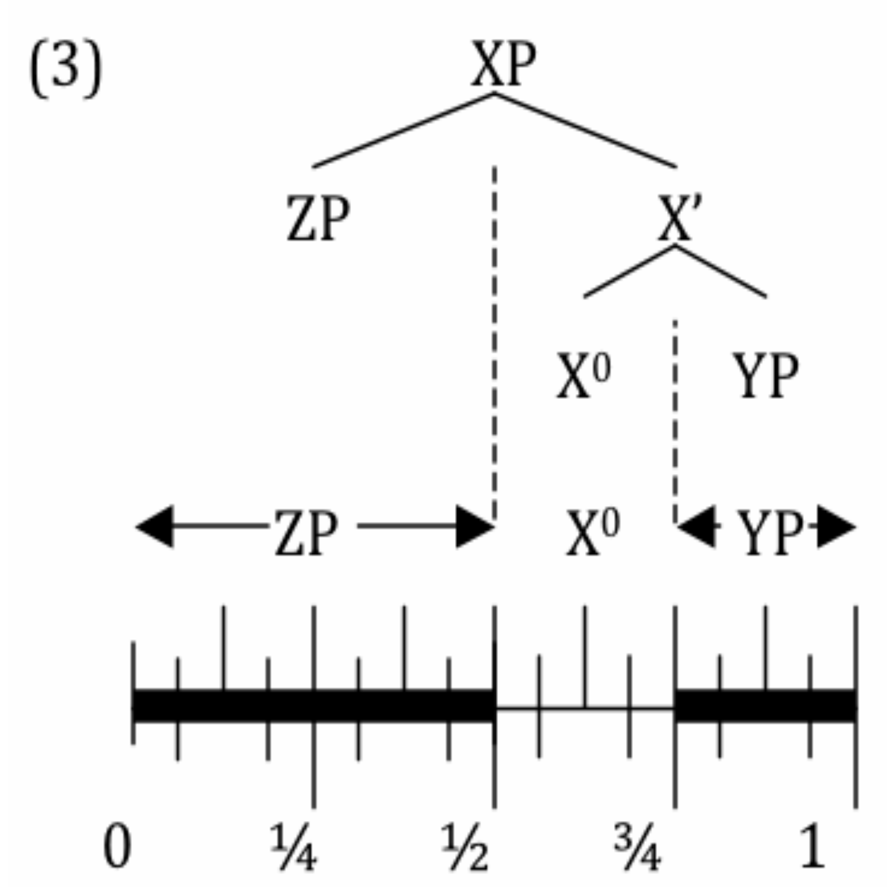
The self-similarity here invites a kind of *phrasal* analysis: within each "generation", there are two copies of the whole, and one "dead end" (deleted segment ~terminal):





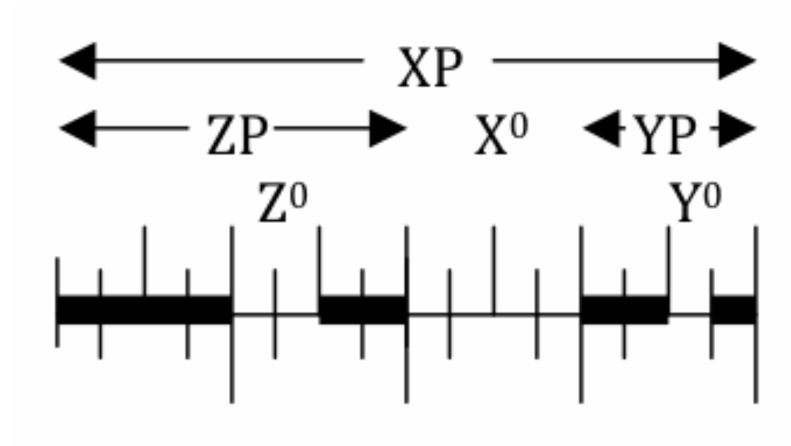
X-bar schema as (multi)fractal

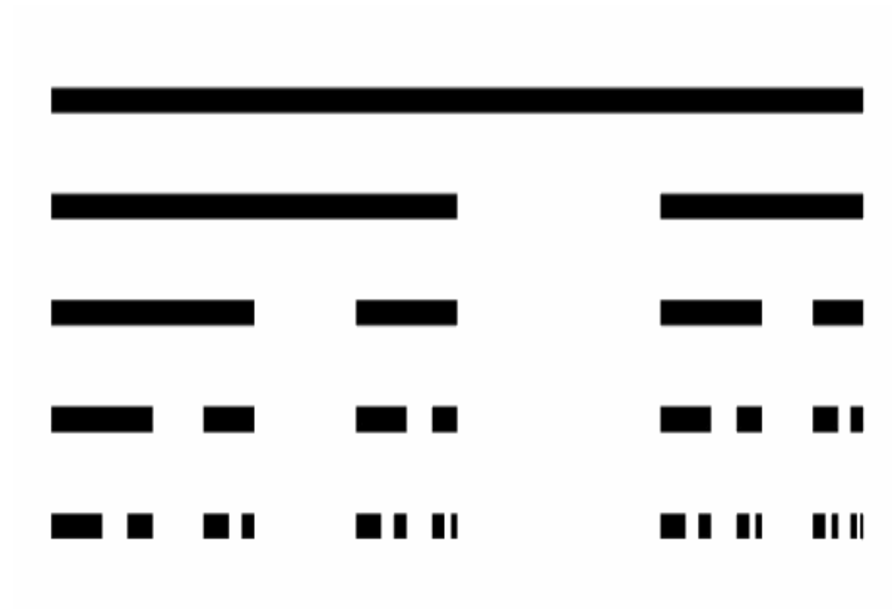
- To construct an appropriate notion of size, we will think of syntactic patterns as line division algorithms.
- Consider mapping the X-bar schema to a line segment,
- Such that binary branching in the syntactic form corresponds to geometric halving,
- And heads/terminals corresponding to deleting a line segment.

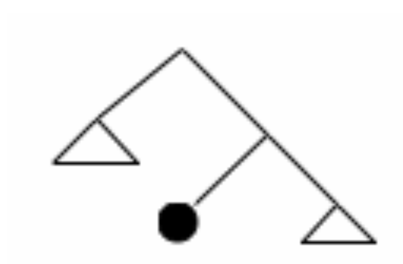


And so on: fractal structure

- Of course, ZP and YP themselves have the same internal structure as XP:
- Continued indefinitely, this produces an asymmetric (or two scale) Cantor set.
- Each generation has one ½ and one ¼ scale copy of the whole.





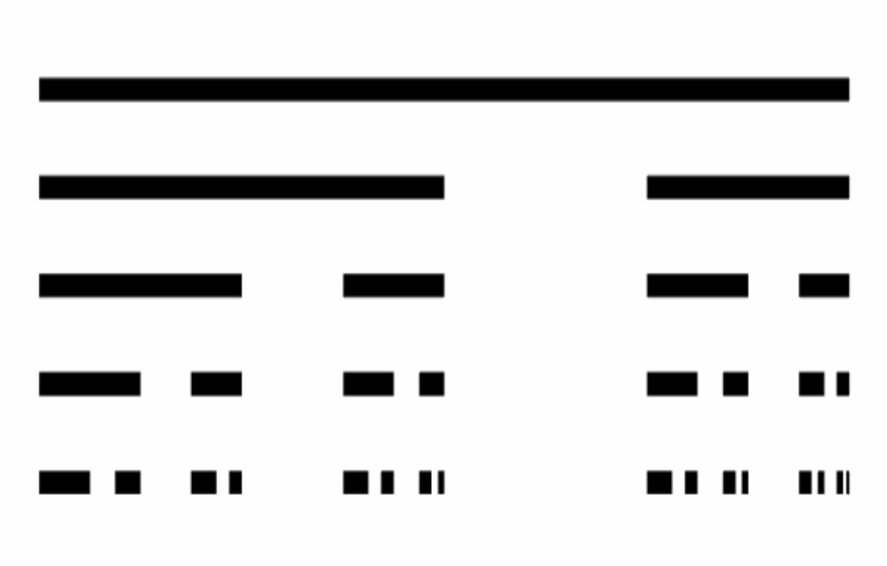


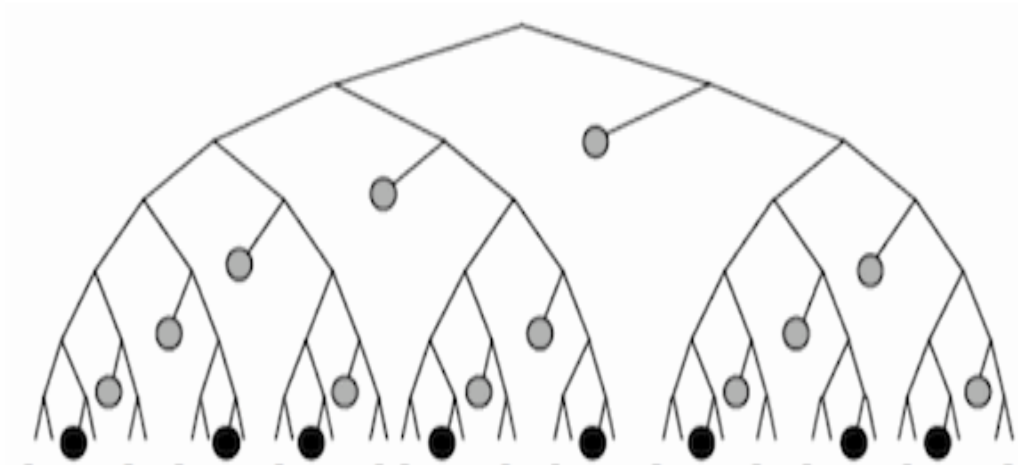
Asymmetric Cantor set ~ X-bar tiling

This is the simplest kind of syntactic fractal;

i.e. it is the smallest kind of self-similar binary-branching object whose non-terminal image on the line is neither the full line, nor a single point.

Its (Hausdorff) dimension is $\log_2(\text{Phi})$ ~.694





Dim = $\log_2(G)$

- Notice that the Hausdorff dimension of the line fractal corresponding to X-bar structure is the logarithm, in base 2, of its growth factor.
- That makes an intuitive kind of sense: this number is, we might say, "what you do to 2 to get Phi".
- The result is general: for non-degenerate systems,

  $$\text{Dim} = \log_2(G)$$

- E.g., D-bar/High-Headed D-bar, with growth factor √2, has Dim = $\log_2(\sqrt{2}) = ½$.
- The Spine, with G = 1, has Dim = 0 (i.e., its non-terminal image converges on a single point, dimension 0).
- (the situation is more complicated for degenerate systems)

Finally, brief survey of 3 non-terminals

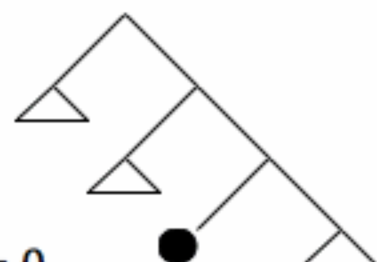
- Moving up the scale of pattern complexity, the next class (built with three kinds of non-terminal) has 57 distinct members.
- (I’ve examined these in detail; I haven’t gone beyond enumerating the 700+ members of the next class up)
- In the next slides, I illustrate just a few of these.

Best growth: generalized X-bar

- The highest growth factor in the 2 non-terminal class belongs to X-bar: the golden mean, associated with the Fibonacci numbers.
- The largest growth factor with three non-terminals is the "tribonacci constant", in the generalized X-bar format in this class (an X-bar like pattern with two specifiers per phrase).

(58) 9_3-bar (Generalized X-bar format with two specifiers)

$$\begin{array}{l} 3 \rightarrow 2\ 3 \\ 2 \rightarrow 1\ 3 \\ 1 \rightarrow 0\ 3 \end{array} \quad \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$



Characteristic polynomial: $x^3 - x^2 - x - 1 = 0$

Growth factor: the "Tribonacci" constant, ~1.839...

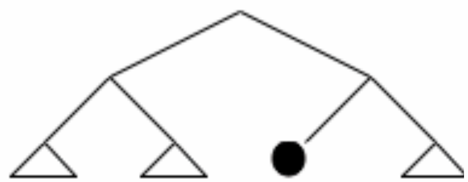
Further 3-type systems

- Here are some more examples from this class:

(57) 2 Power of 3

3 → 1 2
2 → 0 3
1 → 3 3

$$\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix}$$

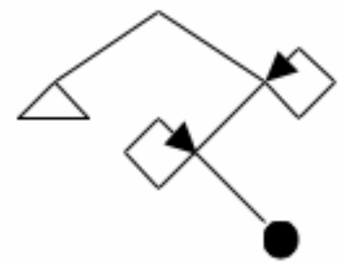


Characteristic polynomial: $x^3 - 3x = 0$
Growth factor: $\sqrt{3} = 1.732...$

(66) K Spine of Spines of Spines

3 --> 2 3
2 --> 1 2
1 --> 0 1

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

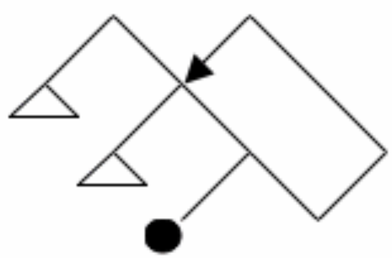


Characteristic polynomial: $x^3 - 3x^2 + 3x - 1$
Growth factor: 1

(63) 39

3 --> 2 3
2 --> 1 3
1 --> 0 2

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

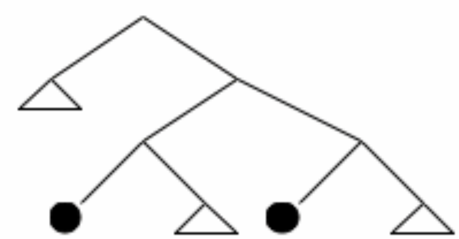


Characteristic polynomial: $x^3 - x^2 - 2x + 1$
Growth Factor: 1.8019…

(62) 29

3 --> 2 3
2 --> 1 1
1 --> 0 3

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 2 \\ 1 & 0 & 0 \end{bmatrix}$$



Characteristic polynomial: $x^3 - x^2 - 2$
Growth Factor: 1.6956...

Growth factors with 3 non-terminals

Systems in family	Polynomial	Growth factor	Special notes
7, 32, 34	$x^3 - 2$	1.2599	$\sqrt[3]{2}$, non-Pisot
13, 25, 35	$x^3 - x - 1$	1.3247	Plastic number *rho*, the smallest Pisot #[23]
22, 28, 33	$x^3 - x^2 - 1$	1.4656	Pisot #
1, 31, 41	$x^3 - x - 2$	1.5214	Non-Pisot
5, 30, 37	$x^3 - 4$	1.5874	
3, 20, 29	$x^3 - x^2 - 2$	1.6956	Non-Pisot
4, 18,27	$x^3 - x^2 - 2$	1.6956	" "[24]
2, 38, 40	$x^3 - 3x$	1.7321	$\sqrt{3}$, Non-Pisot
17, 19, 26	$x^3 - 2x^2 + x - 1$	1.7548	Pisot #; plastic number rho squared
6, 12, 24	$x^3 - 2x - 2$	1.7693	Non-pisot #
11, 16, 39	$x^3 - x^2 - 2x + 1$	1.8019	Non-Pisot, three distinct real roots
9, 10, 21	$x^3 - x^2 - x - 1$	1.8393	Pisot #, the "tribonacci" constant

Towards a conclusion

- This talk has explored some highly abstract mathematical terrain
- That seems likely to be of relevance in understanding the kind of recurrence found in natural language.
- To understand why human language exploits a particular kind of pattern (e.g., the X-bar schema) to achieve discrete infinity
- It may help to know a little more about what other kinds of discrete infinity are possible.

Summary

- I have shown that expressing syntactic recurrence in terms of matrices is especially useful.
- This allows us to directly quantify their "growth" properties: the growth factor G is the dominant eigenvalue of the associated matrix.
- Degenerate patterns, composed of simpler subpatterns, are characterized by the largest G among their components.
- We can think of these patterns as fractal divisions of a line (generalized Cantor sets); on that view, we can calculate their dimensionality (= $\log_2(G)$).