«Natural» Grammars and Natural Language

Juan Uriagereka
UMD

Abstract

Lucas structures can be abstracted from within Fibonacci structures expressed as Lindenmayer-trees by «atomizing» certain chunks of structure (which yields the Padovan series) and by pruning the immediate context of said atomizations. Such conditions may define a space with arguable syntactic significance. Fibonacci and Lucas structures are common in nature, and have been most studied in botanics. Padovan structures have been independently studied in aesthetics.

0. **Introduction***

Fibonacci patterns are amply attested in nature (Jean 1994). A simple way to generate a Fibonacci pattern is in terms of an extension of Chomsky-style rewrite rules proposed by Aristid Lindenmayer, the L-system (Lindenmayer & Prusinkiewicz 1990). In L-systems, which may be seen as "natural" (or growth) grammars, several rewrite rules —*all* applicable rules— apply simultaneously to a given derivational line. Moreover, in these devices no significant distinction is made between terminal and non-terminal nodes, so rule application iterates indefinitely. Consider the example in (1):

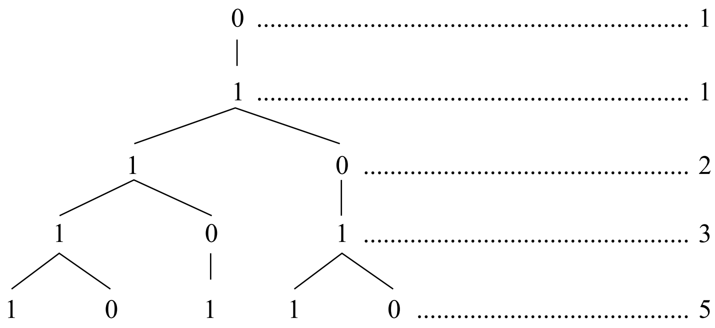
(1) $0 \rightarrow 1$ ("rewrite '0' as '1' "), $1 \rightarrow 1\ 0$ ("rewrite '1' as '1 0' ")

This system, applied Lindenmayer-style, generates a graph like (2), which involves 1, 1, 2, 3, 5, … number of symbols in each derivational line. That is, of course, the Fibonacci sequence, obtained by adding successive numbers starting with 1:

---

* It is a great honor for me to present this paper in celebration of Ibon Sarasola's career, spanning from engineering to philology. I appreciate useful feedback from Bill Idsardi, Terje Lohndal, Guillermo Lorenzo, Roger Martin, David Medeiros, Massimo Piattelli-Palmarini and Doug Saddy. All errors, mischaracterizations, or involuntary lacks of reference stem from my own, vast, ignorance.

(2)



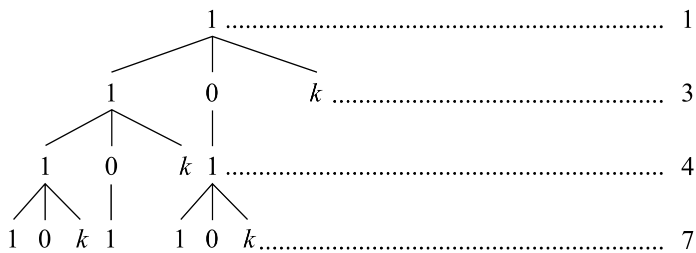
Note that a "semantic" way to generate the Fibonacci series (starting with 0) is by adding the actual arithmetical value of the numbers in each line (that is: 0, 1, 1, 2, 3…).[1]

The Lucas series, too, can be derived in L-systems by making two assumptions: a) the grammar allows for a trifurcation, but b) one of the trifurcated symbols is a "stump" —that is, it is terminal (technically, the symbol can be treated as a constant in the system):

(3) $0 \rightarrow 1$, $1 \rightarrow 1, 0, k$, where k is a constant.

Given such a system, again applied in the Lindenmayer fashion, we obtain:

(4)



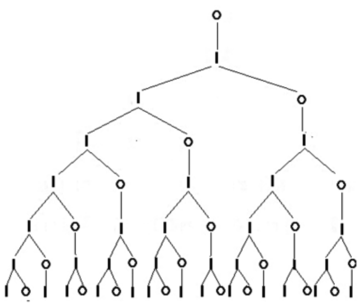
The next derivational lines yield 11, 18, 29, ... symbols, etc. In a direct sense (2) and (4) resemble one another, as "aggregative" systems emerging by adding two successive numbers in the earlier generation of the series, or through two rewrite symbols: a binary skeleton, since the one element

[1] Readers can check that the next derivational line in (2) will yield 8 symbols, the next 13, etc. In turn adding the values of numbers in the next line will yield 8, 13 in the next and so on. The ratio between successive numbers in the series approximates the golden section – the irrational 1.6180339… or 0.6180555…, solutions to the equation $x^2 - x - 1 = 0$.

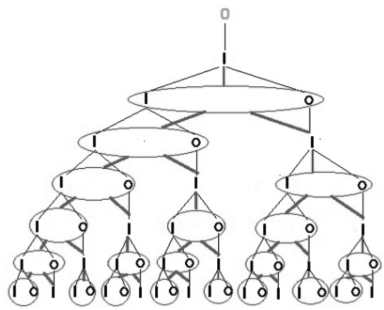
that makes (4) ternary dies out from inception. Note that, by the "semantic method", adding the arithmetical value of the symbols in (4) we obtain the Fibonacci sequence if the constant's value is either ignored or assumed to be of value zero in arithmetical terms — which shows how related the Fibonacci and Lucas series are.[2]

Moreover, there are "higher level" regularities in both (2) and (4) that fall into a "Fibonacci character":

(5) a.



b.



In a basic Fibonacci tree '0' always presents a '1' to its left, and immediately below every '0' there is a '1'. If the grammar in use was a standard Chomsky-style rewrite grammar these regularities need not iterate. But in L-systems the requirement that *all* rewritable symbols be rewritten ends up creating, in effect, a "higher level" object. For (5a) this is represented as in (5b). The "higher" representation could also be generated by the type of L-grammar seen in (1) (ignoring the "start" symbol):

(6) $1 \rightarrow 1\text{–}0,\ 1\text{–}0 \rightarrow 1\text{–}0\ 1$

Structurally, the difference between (6) and (1) is merely semantic. In (1) there is a "weak" symbol (namely, '0') that rewrites as a "strong" symbol; and a "strong" symbol that bifurcates into each of those symbols. The same is true for (6), except this time the weak symbol is '1' and the strong symbol is a compilation into a single rewrite atom of the string 1–0. Let's call such compilations "atomizations":

[2] Another way to show the relationship between these two series is through the rules in (i):

(i) $1 \rightarrow 3,\ 3 \rightarrow 3\ 1$

Readers can check that the syntactic result of a system of this sort still yields the Fibonacci series (counting symbols), while the semantic result yields the Lucas series (adding values).

(7) *Atomization*
Any string of sister symbols can be atomized into a single constituent symbol.

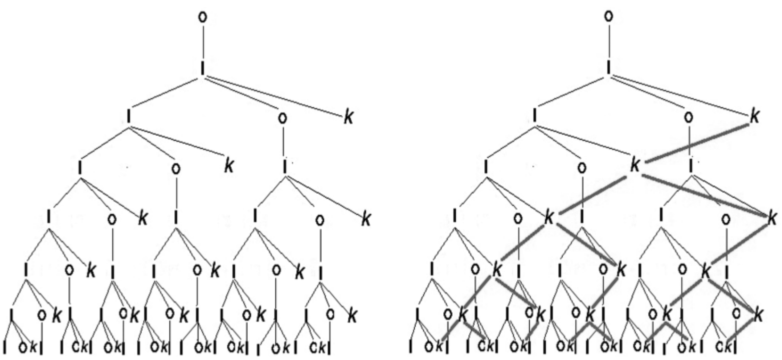
The objects in (5b) and (5a) are structurally identical, the latter also falling into a Fibonacci pattern — so the fundamental structure of objects generated by the L-grammar in (1) is "self-similar".

Note that the comparison just discussed works only if we "ignore" the start symbol, '0'. Let us call the process of ignoring such non-branching symbols, under conditions that we need to explore, a derivational "pruning":

(8) *Pruning*
A non-branching symbol can be ignored in certain designated contexts.

Similar issues arise in objects of the sort generated by (3): Not surprisingly, the object in (9b) is identical to the object in (5a) if we ignore the *k* constants —but even the array of constants in (9b) falls into the Fibonacci pattern.

(9) a. b.



1. **Transforming the Fibonacci object into a Lucas object**

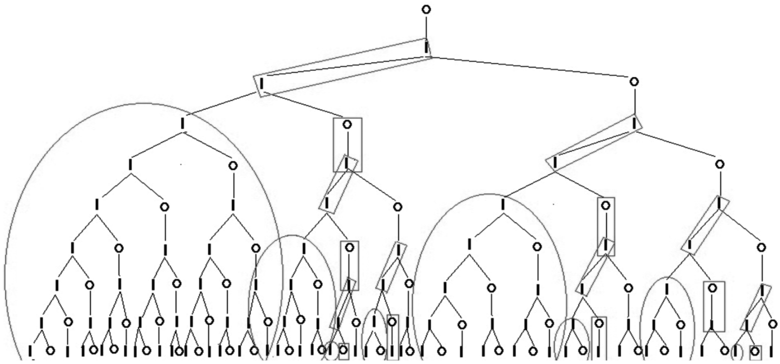
Let's now explore these relations in a more systematic way, making use of the "atomization" (7) and "pruning" (8) tools just discussed, in the following guise:

(10) *Conditions transforming Fibonacci into Lucas L-structures*
a. Atomize all the '1's and their constituent structure in any immediate domination path, except those involving condition (b).

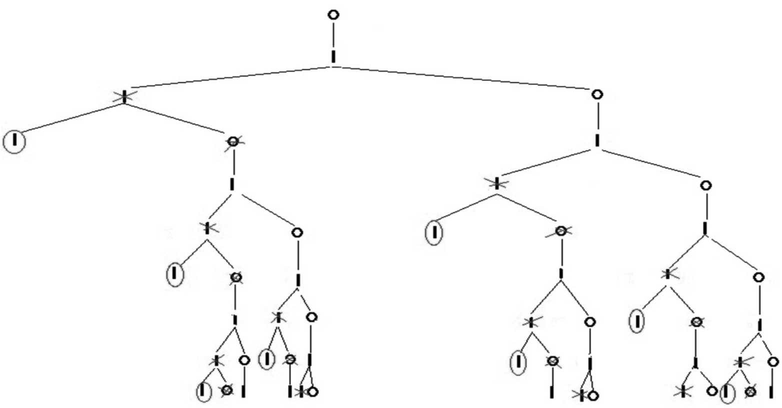
b. Prun any ‘0’ in the immediate context of (i.e. adjacent or immediately dominating) an atomized ‘1’.

The atomization process on the sort of structure in (5a), together with a correlating pruning process, results in the graph in (11). The intended effect of the conditions in (10) is to *atomize as much structure as possible*, while *leaving the syntactic context of the atom “live”*. This syntactic context is the atom’s sister (immediately adjacent) and the atom’s mother (immediately dominating) constituent. Pruning happens in that context. (Note that, by this method, some immediately dominating structures are pruned twice, inasmuch as they happen to be adjacent to an atom *and* immediately dominating another atom.) Graph (10a) identifies atoms in a circle and structures to undergo a pruning collapse in a square. Graph (10b) is a cleaned up version.

(11) a.



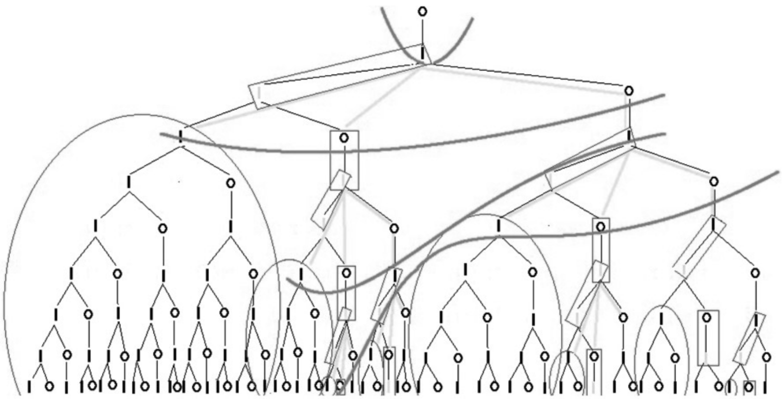
b.



Note that the number of *new* atoms that result per derivational line beyond the initial step is: 0, 0, 1, 0, 1, 1, 1, 2, 2… The number of such atoms in the next line is 3, then 4, then 5, 7, 9, 12… This is the Padovan series (and see fn. 1).[3] The compactness of the object is shown by the fact that the number of "non-atoms" in each derivational line is: 1, 1, 2, 2, 3, 4, 5, 7, 9, 12…, again the Padovan series — albeit five steps into its generation. Also, each derivational line after the root of the object contains 0, 1, 1, 1, 2, 2, 3, 4, 5… pruned elements, again the Padovan sequence, ignoring the seed.[4]

Through the adjustments in the structure in (11), pruned elements and new edges connect the remaining symbols. As we scan the resulting object top-down (which wavy lines emphasize in the clarifying graph in (12)), again ignoring the very top element, the ensuing number of symbols is 1, 3, 4, 7… The Lucas series.

(12)



So the conditions in (10) show a curious structural reduction: computationally, the Fibonacci object *properly contains* the Padovan and Lucas objects. In particular, when imposing certain atomization conditions coupled with related conditions of pruning, the Fibonacci tree reverts to a Lucas structure, as (13) highlights.[5]

---

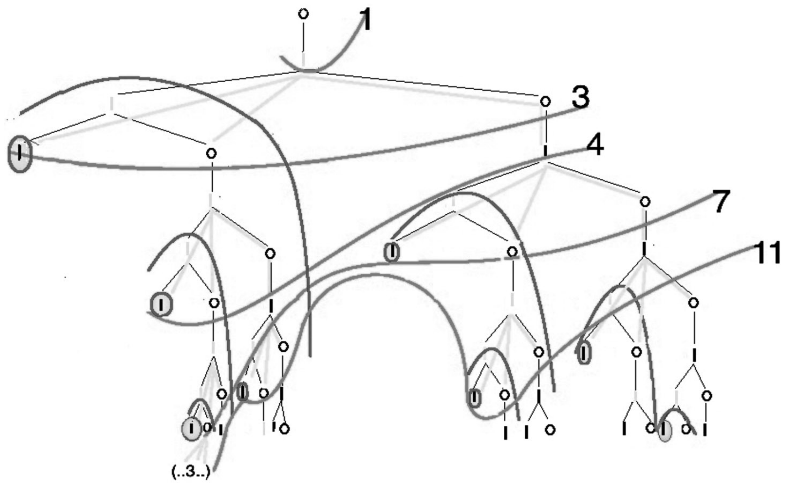
3 Ratios between this sequence's terms approximates the plastic number 1.3247179…, the unique real solution to $x^3 - x - 1 = 0$. Each number in this sequence is obtained by skipping the previous one and adding the two before that, starting at '0, 0, 1', which generates 0, to be skipped while we add 1 to 0, to generate 1, to be skipped, etc. Hans van der Laan used this as a base for the proportion of his architectural constructions, after having performed experiments to discover the limits of humans' ability to perceive relationships between objects.

4 The Padovan series can also be expressed in Lindenmayer fashion by applying the rules in (i):

(i) $1 \to 2, 2 \to 3, 3 \to 1\ 2$

5 In (13) we ignore inner elements in the atomized structures and emphasize portions containing doubly pruned structures and associated atomized structures. The (…3…) in the lower part of the graph shows elements from the next derivational line, which the relevant "cut" needs to consider to properly aggregate.

(13)



Needless to say, the formal observations just reviewed are tightly connected to the assumptions made. With a different sort of atomization, different results ensue. Obviously, atomizing the entire structure trivializes it, but even atomizing with room for just a single dependent simplifies matters: the aggregative nature of the structure reduces to a periodic one: 1, 2, 1, 2, 1, 2, 1, 2... Atomizing a smaller structure than what we see in (11), leaving *three* (not two) branches "live" yields the series of atoms 1, 0, 1, 1, 2, 2, 4, 5, 8, 11…, which are sometimes referred to as "poor Phidias numbers". This series converges to 1.4655712 (a cubic Pisot number).[6] It is likely the case that other atomizations yield interesting results too.

2. **Grammatical Considerations**

"Waves" of structure of the sort seen in (13), including an atomic element and two immediately related symbols, are arguably relevant to natural language syntax. Atomic elements as in the previous section are customarily called "heads";[7] a sister to a head is called a "complement",[8] while a further phrasal dependent of a head is called a "specifier". The basic "molecule" of language exhibits the form in (14):

---

[6] A central characteristic of Pisot numbers is that their powers approach integers at an exponential rate. For instance, the number in the text approximates 2. The plastic number is also a Pisot number, the minimal one.

[7] In (12)/(13) always '1's exhaustively dominating other '1's.

[8] In (12)/(13) complements and specifiers stem from a '0'. The difference between them in these graphs is that complements involve double pruning.

(14) a. Neronis incendium Romae
Nero-GEN burning Rome-GEN
"Nero's burning of Rome"

b.
```
                          NP
                      /        \
SPECIFIER   Neronis             N'
                              /      \
                [incendium]N        Romae
                   HEAD           COMPLEMENT
```

Visually this Latin example presents the mirror image of the sorts of structures in (12)/(13), but we are abstracting away from linear ordering considerations in this exercise.[9] The complement/specifier asymmetry can be illustrated as follows:

(15) a. Nero's city burning is legendary. [Intended: *Nero's burning of a city*]
b. *Rome('s) man burning is legendary. [Intended: *a man's burning of Rome*]

(15a) is one among many processes demonstrating the tight connection between a head like *burning* and its complement (in this instance, *city*). An expression like *city burning* ensues from a more basic structure akin to *burning of a city*. *Mutatis mutandis*, we could ask whether something like *a man's burning (of Rome)* couldn't underlie the expression *man burning* presented in (15b). But in such an instance the only meaning available involves the man as the target (not the agent) of the burning.

Head-complement relations in the most general sense define so-called computational *phases* (Chomsky 2001). The most important property of phases is that, for the purposes of the derivational system, they transfer to semantic interpretation at specific points in the derivation. What this means is that, although their component parts are live for interpretation, syntactically they are inert, thereby yielding a variety of specifically linguistic properties.[10] *The scaffolding of computational phases is destroyed* after the transfer. The pruning at the phase level in (12)/(13) (highlighted by the curves) may

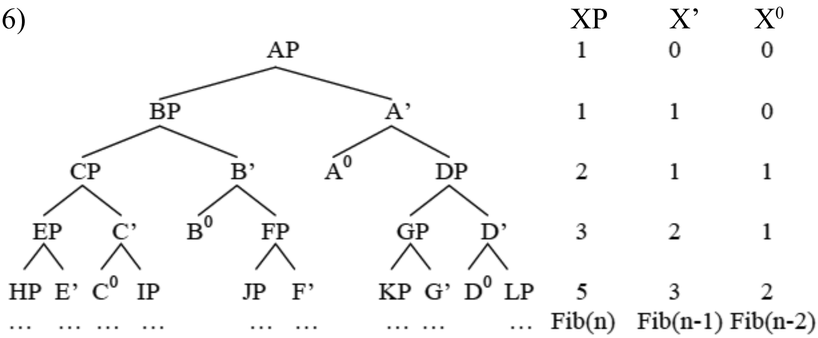
[9] As is well known, Latin allows all other permutations of these particular phrases, although the one in (14) is arguably the default one. In many other languages, of course, other such orderings are the most natural (e.g. the English *Nero's destruction of Rome* or the Spanish *la destrucción de Roma de Nerón*).

[10] For example, binding conditions (licensing reflexives or forcing the obviation of co-referent pronouns) take places within phases, as does the phenomenon of "successive cyclicity" in long-distance Wh-displacement.

be interpreted in just this way: the constituents within the phase attach directly (by the tripartite connections) to the immediately dominating element, higher up in the structure — their previously existing syntactic structure is no longer present after phase-transfer.

It should be emphasized, perhaps obviously, that not all syntactic structures in human language need to have the maximal structure represented in (12)/(13).[11] This is a point emphasized by Medeiros (2008), when studying conditions for syntactic structures that, he argues convincingly, present Fibonacci signatures. They do *at the limit*, when maximally expanding nodes. Observe (16). If we expand every expandable branch in this basic X'-structure, the Fibonacci pattern is clear (for maximal projections XP, intermediate projections X', and heads X^0):

(16)



3. **Conclusions**

This paper has argued that a relatively simple way exists to abstract Lucas structures from within Fibonacci structures. This was done, first, by "atomizing" all the '1's and their constituent structure in any uniform immediate domination path —except the top two '1s'— which yields the Padovan series. In addition the "immediate context" of atomized structures (i.e. any '0' adjacent to an atomized '1' and any '1' immediately dominating the atomized structure) was pruned, which results in the Lucas series. It seems curious that those conditions should define a "topological space" with syntactic significance: what constitutes a "head", what constitutes a "complement", what constitutes a "phase". The question to explore is why conditions on atomization and pruning, at the specific level studied here (i.e. yielding

[11] Although structures with this amount of complexity are not hard to parse: *[[Conditions that provoke [governments to tell lies]] [(normally) entail [[situations triggering mistrust] (at least partly) [affecting [citizens embracing democracy]]]]].*

heads, complements, corresponding phases), bridge two basic mathematical structures, with an overwhelming attestation in nature. Jean (1994), reporting on tens of thousands of observations of hundreds of plant species, shows how Fibonacci patterns emerge in 92 % of instances, while Lucas cases are the next most common (2 % of the observations). Padovan objects ("mediating" Fibonacci and Lucas objects as computationally organized by L-grammars) are also independently studied, at least in aesthetics (see fn. 3). Aside from proving these conjectures and attesting other natural connections beyond botany (well known from a vast literature), the real question to explore is why any of these structural factors should matter to language or cognition more generally.

References

Chomsky, N. 2001. «Derivation by Phase». In Michael Kenstowicz (ed.), *Ken Hale: A Life in Language*. Cambridge (MA): MIT Press. 1-52.

Jean, R. 1994. *Phyllotaxis: A systemic Study of Plant Pattern Morphogenesis.* Cambridge: Cambridge University Press.

Lindenmayer, A. & P. Prusinkiewicz. 1990. *The Algorithmic Beauty of Plants.* Springer-Verlag.

Medeiros, D. 2008. «Optimal Growth in Phrase Structure». *Biolinguistics*, 2, 2-3.