Dynamical Systems

John Goldsmith

May 24, 2017

1 dynamical systems, and force

What do we mean when we talk about *dynamics*, or about a *dynamical system*? The first word that might come to mind is *time*: a dynamical system is one whose change or evolution in time is significant – indeed, is of prime significance. In fact, if we were to ask what a dynamical approach contrasts with, we would undoubtedly use the word ”static” in some way – a dynamical approach is the opposite of a static approach, which is to say, one in which time and change plays little or no role.

But time is not the most important aspect of a dynamical system unless it is used in a particular way. Let's focus for a moment on graphical representations of language. Both in traditional orthographies and in linguistically sophisticated traditions, we use distance on a page (typically from left to right, but that is just a habit) to represent change. (1) *The farmer killed the duckling.* is a sentence of five words that are uttered in the same order that the words occur on the page. Whether we are talking about standard orthography or a phonetic transcription, we can use space as a place-holder (or a metaphorical stand-in) for time. And that is what we do when we read, when we write, and when we transcribe.

But using space to represent the flow of time is not enough to make something a dynamical system. To repeat: writing the names of sounds (or of words) on a piece of paper in a given spatio-temporal order does not place us in the middle of a dynamic system.

What then do we need to be doing in order to be engaged in developing a dynamic system?

The answer was provided by Isaac Newton in the mid 17th century, in perhaps the single greatest feat in scientific revolution we have ever known. Newton created a new mathematics (which we call the calculus) which allowed him to talk of two entirely new concepts: velocity and acceleration.

Newton adopted some of the mathematics that Descartes and others on the Continent had developed in order to talk about location in space, and that allowed him to describe the difference in an object's location between two different times, two different moments. He then took a leap and showed that the ratio of the locations' difference to the difference in time would converge to a specific value as the difference in time was considered smaller and smaller. $\frac{y(t+\Delta t)-y(t)}{\Delta t}$ That ratio, of the difference of location to the difference of time, he called the velocity of the object, and he developed a mathematics that could deal with such notions. Not perfectly, and his mathematics was criticized for centuries until it was set upon firmer foundations – but it was an enormous breakthrough both in physics and in mathematics.

The mathematics incorporated what we call the (first) derivative of the location of an object, which is also called its velocity. It is like the object's speed, but it is not speed: it is speed along with a specific spatial direction.

But it was not the really revolutionary part of what Newton accomplished. Others were working on the notion of velocity and speed and momentum at the same time, influencing each other. What was totally revolutionary was a combination of two things: first, applying in a recursive way the mathematics that he had just invented to describe velocity, and second, to devise a new metaphysics of the physical world in which the velocity at which an object moved required no explanation, not external influence, if the object continues it course both in speed and in direction.

We know that statement as Newton's first law of motion: an object will stay at rest, or stay in uniform motion if it is not at rest, unless acted upon by an outside force.

Which is all well and good, if we have some notion of force – something which did not really exist before Newton.

In fact, while Newton knew that he had to devise a notion of force to make his system work, but he was actually fighting an uphill battle at the time. Virtually all of the smart money at the time – by which I mean the other natural philosophers who were trying to understand the nature of the universe – were convinced that there wre no occult forces, and in particular, no invisible forces that could operate at a distance. No one was more convinced of this than René Descartes, who not too long before Newton proposed a theory of gravitation that did away with any effect-at-a-distance whereby the Sun (for example) held the planets close to it.

The reasons that Newton was convinced that Descartes had over-reached with his theory of no occult forces are complex and fascinating, but we will not go into them here. What is important for us is that Newton told us to focus not on velocity, but on the first derivative of the velocity.

You recall that Newton's first derivative of an object's location is its velocity, and so the derivative of the velocity is the second derivative of the location. To this we have given the name ″acceleration″. And Newton's insight was that all motion in the universe could be modeled and understood if we simply understand velocity and acceleration. If there are no forces, then velocity is very easy to understand: it never changes. And if there *are* forces, then velocity changes. Newton had invented the mathematics of dealing with changes of velocity, or acceleration; he could measure acceleration, and give it a number and a direction.

Then his crowing insight was that there are occult forces in the world, such as gravity, and that the effect of gravity or any force on an object was an effect on the object's acceleration. The object would suffer an acceleration linearly proportional to the force that it was subject to, and in the direction of the force as well.

That law holds true whether the object is at rest or in motion; if there is a force, then it accelerates. If it starts at rest, then it starts moving, slowly at first and then faster as the acceleration continues its effect. If it is not at rest, its motion (its velocity) changes in the direction of the force it undergoes.

Thus on Newton's new world view, there are hidden forces such as gravity, and magnetism as well. The key to understanding the metaphysics and the dynamics of the world is to build mathematical models in which forces and the second derivative of location are related to each other.

This brings us now to the heart of a dynamic system, because it was the dynamical system that Newton had just invented. A dynamical system is one in which we are interested in the location of objects in space (which may be one dimensional, two- or three-dimensional, or 10,000-dimensional) and in the first and second derivatives of their locations.

The rest is history, of course, but there are exactly two parts of history that we care about now. If we ask what kinds of systems emerge most naturally out of dynamical systems, the answer is two. There are systems that go through cyclic repetitions, and there are systems that enter into a condition of static equilibrium. And this brings us to linguistics.

Before we get to the linguistics, though, a few more words from math and physics about why these two systems arise so naturally out of Newtonian dynamics. There is a feature that I have not exactly glossed over, but I have not emphasized it, and it is this: in the real world, dynamical systems do not need to use third, fourth, or higher degrees of differentiation – and we have already seen that velocities (which are first derivatives) remain constant except for the effect of external forces.

Springs have a natural treatment in Newton's terms, and the force on an object which is held by a string to a center is inversely proportional to the distance that separates them. Springs are a great example of forces that are not occult and not at a distance (two differences separating it from gravity). The force is not constant, but is rather linearly inverse with respect to the distance separating the objects. Such a system finds itself in motion that repeats over time, and in theory at least it will continue to move back and forth forever if no other forces enter into the picture.

Gravity works in a similar fashion, but with a force that decreases not with the distance separating the sun and a planet, but with the square of the distance. This, plus the fact that we live in a three dimensional world, leads to a very different set of possibilities. If you were to throw a planet at the Sun from a distance, it would almost certainly not go into orbit around the Sun; it would obey Newton's laws, but it would swing round the Sun and then swing out and go off on a distant arc.

But, as you know since you are here, it is possible for a lucky planet to be in an orbit around the sun, in a way that can be described as an ellipse in which the Sun is at one of the foci.

That is the first interesting outcome for a dynamic system: it can settle into a cycle of motion that it stays in forever unless additional outside forces enter into the picture.

The second interesting outcome of a dynamic system is that it can enter into a state of dynamic equilibrium, one in which there are foces acting on all of the objects, but the forces cancel out, and the system does not move (or moves with constant velocity, which we can ignore by making sure that the observer is moving appropriately).

2 cyclic motion

It has long been noted by observant people that speech consists of sequences of mouth openings and closings, which in some fashion is related to what we linguists call syllables.

How do we treat syllables in phonology?

One tradition in phonology treats syllables as an effect rather than a cause.

3 computational equilibrium

The oldest notion of the algorithm involves a process that can be repeated until some index (some related integer) eventually reaches 0, at which point the computation stops. The grandfather of all such algorithms is the algorithm for finding the largest common divisor of two numbers.

Later, with the invention of the calculus, people began to look at computations that would get indefinitely close to a limit value. If the limit value is the value that we need, then we can stop as soon as we get within some margin of distance to the point.

In the 20th century it was found that some computations are much easier to make if we treat them in this way. This was eventually used as the underlying model for recurrent neural networks in the 1970s.

4 Time and evolution in the history of linguistic thought

1. First two generations of linguistics in 19th century dominated by Hegelian thinking: history means written history, and the changes we have seen in historical times have been degeneration.

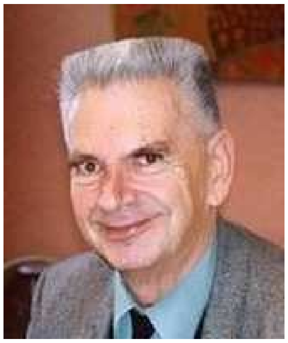
2. Third generation, self-consciously grasped by the Neogrammarians, caught up in the new Darwinian and positivistic view of history: all change is small in the moment and *largely in random directions*, and history in the large is the gradual summation of such small changes.

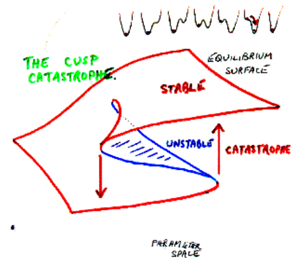
3. European reaction against this, seen both in Germany (as in Gestalt psychology) and in Russia (Trubetzkoy and Jakobson, Eurasianism). All change is toward a goal.

4. Generative grammar (phonology), harmonic phonology, optimality theory, and beyond.

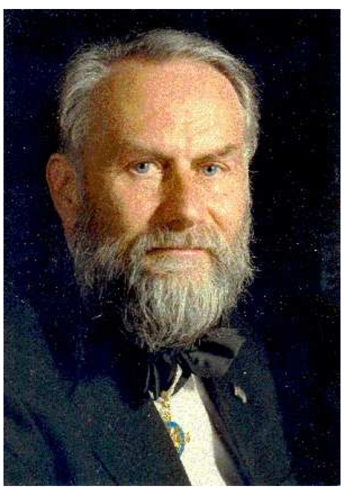
5. Two preceding periods of interest in dynamical systems in the last few decades:

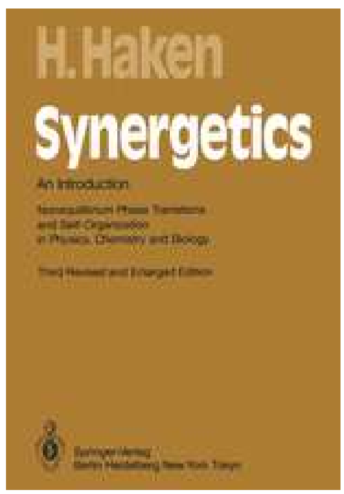
   i. Work inspired by René Thom, a French mathematician who invented what he called catastrophe theory, popularized in the 1970s by Christopher Zeeman.



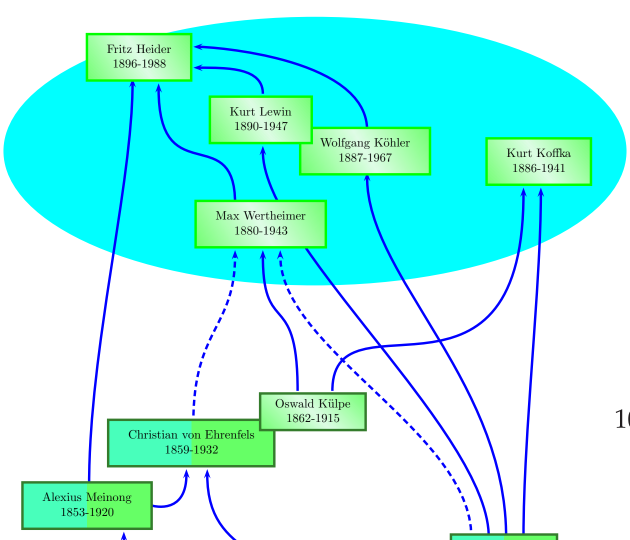


ii. Work inspired by Hermann Haken, German mathematician; this includes a lot of work in the 1990s, some of it reported in the 1995 MIT Press book edited by Robert Port and Tim van Gelder *Mind as Motion : Explorations in the dynamics of cognition*.





5 Gestalt psychology



Kohler wrote in 1920, given by Ellis in his *A Source Book of Gestalt Psychology*, p18f:

> Let us consider under what conditions a

physical system attains a state which is independent of time (i.e., a state of equilibrium or a so-called stationary state). In general we can say that such a state is reached when a certain condition is satisfied for the system as a whole. The potential energy must have reached a minimum, the entropy a maximum, or the like. The solution of the problem demands not that forces or potentials assume particular values in individual regions, but that their total arrangement relative to one another in the whole system must be of a certain definite type. The state or process at any place therefore depends in principle on the conditions obtaining in all other parts of the system.

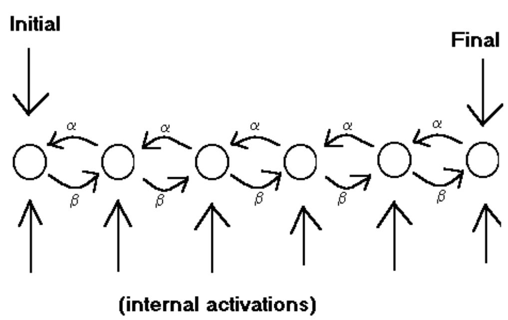
6 phonology

Take a step back from linguistic analysis, and ask: what is the simplest way to perform the computations that are central and important for the data of metrical systems?

From a cognitive point of view: What kind of [...neural...] hardware would be good at performing that kind of computation? Remember: the brain has no paper in it!

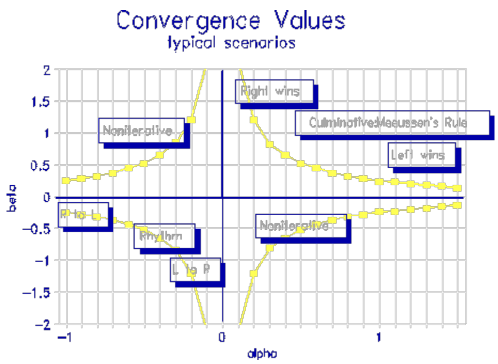
From a more traditional point of view, we may ask the question: **why is there a strict distinction between the mechanism used for *phonological representation* and the mechanism used to *modify* the representations (rules, constraints, etc.)?** Let's build a model in which the two are integrated.

7 Dynamic computational model



5 parameters:

1. α to the left
2. β to the right
3. I Initial positional activation
4. F Final positional activation
5. P Penultimate positional activation



- a_i^t is the activation of the i^{th} unit at time t.
- Inherent activation: Inh(i) = $\delta(1,i) \times I + \delta(-2,i) \times P + \delta(-1,i) \times$

  where[1]

  $\delta(1,i) = 1$ iff i=1;

  $\delta(-2,i) = 1$ iff i is the penultimate position;

  $\delta(-1,i) = 1$ iff i is the ultimate position.

$$a_i^t = Inh(i) + \alpha \times a_{i+1}^{t-1} + \beta \times a_{i-1}^{t-1}$$

In matrix form, we have:

$$\vec{A(t)} = \vec{Inherent} + M \times \vec{A}(t-1)$$

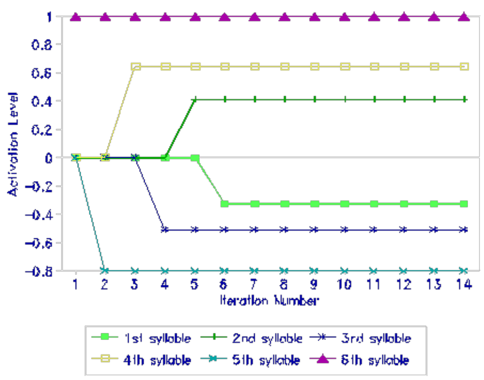
[1] This is a pretty commonly used notational trick to simplify the algebra, sometimes written δ^i_j.

where M is a matrix of the form:

$$\begin{bmatrix} 0 & \alpha & 0 & \dots & 0 & 0 & 0 \\ \beta & 0 & \alpha & \dots & 0 & 0 & 0 \\ 0 & \beta & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \beta & \dots & 0 & 0 & 0 \\ & & \dots & & & & \\ 0 & 0 & 0 & \dots & \beta & 0 & \alpha \\ 0 & 0 & 0 & \dots & 0 & \beta & 0 \end{bmatrix}$$

M(i,i+1) = α

M(i, i-1) = β, for $1 \leq i < n$.



8 Sonority and syllabification

8.1 Tashlhit Berber

Dell and Elmedlaoui

8	7	5	4	3	2	1	0
a	i u	liquids	nasals	voiced fric	voiceless fric	voiced stops	voiceless stops

2nd p sg perfective	3rd p fem sg w 3rd m sg obj	
tRgL-t	tRgl-As	lock
tSkR-t	tSkr-As	do
tZdM-t	tZdm-As	gather wood

t	l	w	a	t	(α, β)
0	5	7	**8**	0	inherent activation
-1.1	3.6	4.6	**8.0**	0	(-0.3,0)
-1.8	**3.5**	3.0	**8.0**	0	(-0.5,0)
-2.1	**4.2**	2.2	**8.35**	1.3	(-0.5,.15)
-2.6	**4.3**	1.5	**8.4**	-0.9	(-0.6,.1)

t	i	z	r	w	a	l	i	n	(α, β)
0	**7**	3	5	7	**8**	5	**7**	4	inherent activation
-4.4	**7.3**	0.2	**3.4**	2.6	**6.8**	1.6	**4.6**	3.8	(-0.6,-0.1)

This next word is especially interesting, because it illustrates how a high sonority segment can fail to be a syllable nucleus, because its sonority is dampened by its right-hand neighbor:

i	h	a	u	l	t	n	(α, β)
7	2	8	7	5	0	4	inherent activation
8.85	-3.1	**7.0**	2.2	**6.9**	-3.5	**4.6**	(-0.6,-0.1)

9 Stress and cyclicity: Indonesian

Based on material in Cohen 1989.

bicára	speak
bìjiksána	wise
xàtulistíwa	equator
òtobìográfi	autobiography
àmerikànisási	Americanization

Cohen's analysis:

1. Final syllable is extrametrical.
2. End rule: Final ("penultimate stress").
3. End Rule: Initial (but blocked if clash would ensue).
4. Perfect Grid (Right to Left, blocked if clash would ensue)

rule	o o o	o o o o	o o o o o	o o o o o o
1	o o (o)	o o o (o)	o o o o (o)	o o o o o (o)
2	o ó (o)	o o ó (o)	o o o ó (o)	o o o o ó (o)
3	clash	ó o ó (o)	ó o o ó (o)	ó o o o ó (o)
3	o ó (o)	ó o ó (o)	ó o o ó (o)	ó o ó o ó (o)
$\alpha = -0.5$	$\beta = 0.0$	I = 0.7	P = 1.0	
	0.7 1 0	0.7 0 1 0	0.7 0 0 1 0	0.7 0 0 0 1 0
	0.2 **1.0** 0.0	**0.95** -0.5 **1.0** 0.0	**0.58** 0.25 -0.5 **1.0** 0.0	**0.76** -0.13 **0.25** -0.5 **1.0**

Morphologically complex cases:	
1st cycle	o o o o o
1	o o o o (o)
2	o o o ó (o)
3	ó o o ó (o)
4	DNA (clash avoidance)
2nd cycle	ó o o ó o o
1	ó o o ó o (o)
2	ó o o ó ó (o)
3	ó o o ó ó (o)
4	DNA (clash avoidance)
output	ó o o ó ó (o)
clash resolution	ó o o o ó (o)
output	ó o o o ó (o)

$\alpha = -0.5$	$\beta = 0.0$	I = 0.7	P = 1.0			
	0.7	0	0	1	1	0
	0.64	0.13	-0.25	-0.5	**1.0**	0.0

10 Stress systems: typology

Examples, from Hayes

10.1 Pintupi

(Hansen and Hansen 1969, 1978; Australia):

"syllable trochees": odd-numbered syllables (rightward); extrametrical ultima:

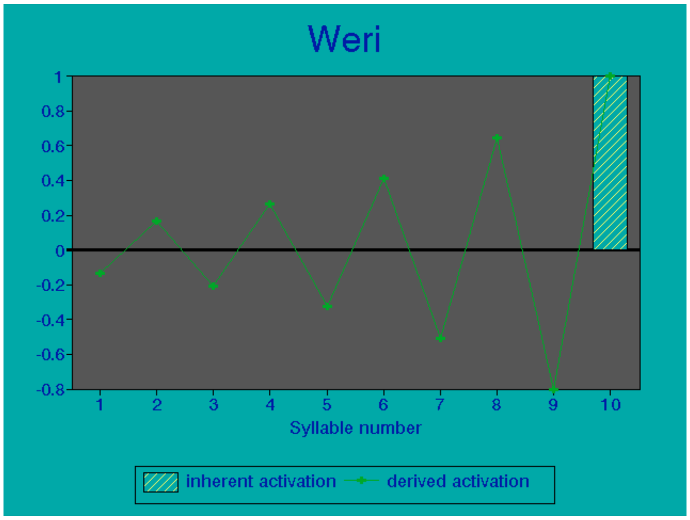
S	s					
S	s	s				
S	s	S	s			
S	s	S	s	s		
S	s	S	s	S	s	
S	s	S	s	S	s	s

10.2 Weri

Boxwell and Boxwell 1966, Hayes 1980, HV 1987

ŋintíp	bee
kùlipú	hair of arm
ulùamát	mist
àkunètepál	times

- Stress the ultima, plus
- Stress all odd numbered syllables, counting from the end of the word.



I = 0.0

F = 1.0

α = -0.8

β = 0.0

10.3 Warao

(Osborn 1966, HV 1987)

I = 0.0

F = -1.0

α = -0.7

β = 0.0

- Stress penult syllable;
- all even-numbered syllables, counting from the end of the word.

jiwàranáe	he finished it
japurùkitànehátse	verily to climb
enàhoròahàkutái	the one who caused him to eat

10.4 Maranungku

[2] Stress first syllable, and All odd-numbered syllables from the beginning of the word.

I = 1.0

F = 0.0

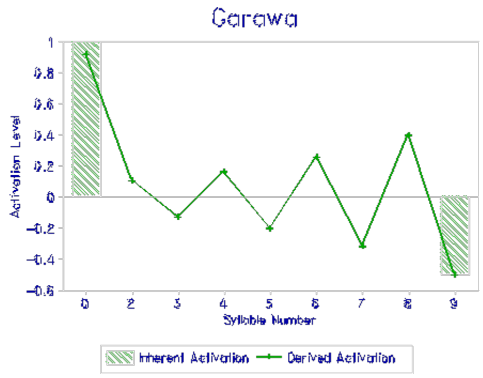
α = 0.0

β = -0.7

10.5 Garawa

[3] ...or Indonesian, ...

[2]Tryon 1970
[3]Furby 1974



Garawa (Furby 1974) illustrates this quite common class: accent falls on both the initial syllable and on the penult, corresponding to a positive setting of I, a negative setting of F, and a negative value of α (in order that the negative value of F should translate into a positive value for the penultimate syllable). In such systems, we typically find either accent iterating from left to right, on odd-numbered syllables counting from the first, or else accent iterating from right to left, on every other syllable to the left of the penult, depending on the relative magnitudes of α and β. Garawa falls into the latter category, and this pattern illustrates the result of a system in which the α-effect is stronger than the β effect: in which, that is, α¡ β (though, more to the point, the absolute value, —α— is greater than —β— , since α is negative)

yámi	eye
púnjala	white
wátjimpàŋu	armpit
náriŋinmùkunjìnamìra	at your own many

- Stress on Initial syllable;
- Stress on all even-numbered syllables, counting leftward from the end;

but "Initial dactyl effect": no stress on the second syllable permitted.

I = 1.0

F = -0.5

α = -0.7

β = -0.1

10.6 Lenakel

Lynch 1978; Hayes 1995: 167-78. As is well-known, accent in Lenakel is unusual in that stress is assigned according to principles that appear to be quite different in nouns when compared with the principles operative in verbs and adjectives. Verbs and adjectives (see) are stressed on the penultimate syllable, on the first syllable, and on every alternate (odd numbered) syllable as we count from left to right, starting with the beginning of the word, with the exception that the antepenult is never stressed. Nouns, on the other hand, bear penultimate stress, and show a pattern of accent assignment on alternate syllables counting from the end of the word, alternating leftward from the penultimate syllable.

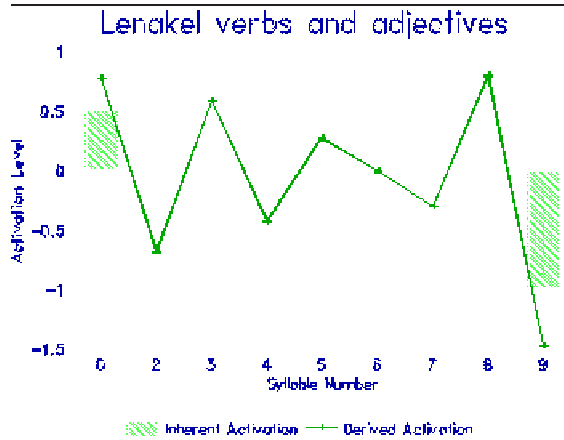
I = 1.0 for verbs and adjectives; 0.0 for nouns

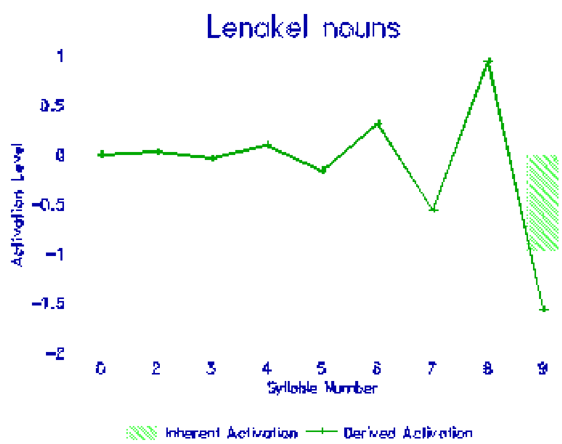
F = -0.5

α = -0.4

β = -0.6

verbs and adjectives:	
r`ɨmɔlgέygɛy	he liked it
n`ɨmarɔlg′ɛgɛy	you p. liked it
n`ɨmamàrɔlgέygɛy	you pl. were liking it
t`ɨnagàmarɔlgέygɛy	you pl. will be liking it
nouns (four or more syllables):	
nɨmwàgəlágəl	beach
tubwàlugálʊkh	lungs



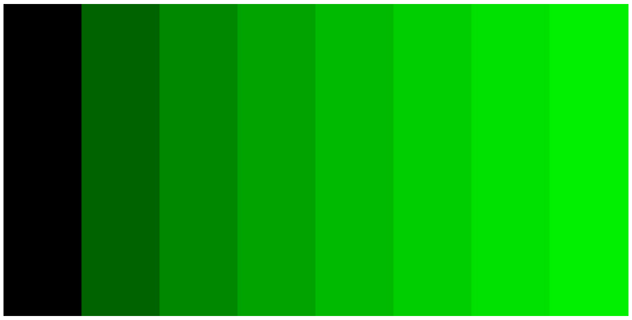


This pattern is a peculiar embarrassment to traditional accounts of Lenakel, accounts which distinguish essentially between rules and representations.

In nouns, not only is the initial stress of the verbs missing, but the direction of iteration of the rule that creates alternating stress must change depending on lexical category. In the present model, however, nothing of the kind is necessary; not only is this case not an embarrassment, it is precisely the kind of case that is predicted by the theoretical model. We need simply say that in the case of nouns, there is no Initial activation; crucially, however, the values of α and β remain fixed across the entire language. Because there is no Initial activation in the case of nouns, there is no rightward-spreading wave for the β-coefficient to pass on. There is, from a mathematical point of view, both a wave propagated leftward and a wave propagated rightward; the one which is stronger will, by and large, drown out the other from a purely quantitative point of view, but when the rightward moving wave is removed, by the non-occurrence of initial stress in the nominal system, the wave moving sotto voce leftward from the penult becomes entirely audible.

11 Lateral inhibition and Mach bands

Here is one way to see the phenomenon known as *Mach bands*. In the figure below, each of the colored rectangles has a uniform color, and yet we see sharpened edges with a black, or very dark, band separating the lines or bands—but the lines or bands are not there.



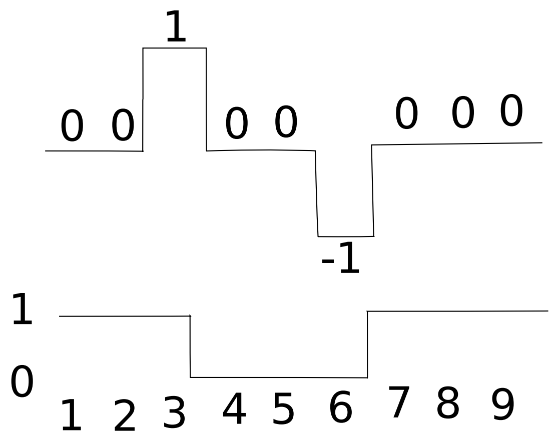
The mathematics of lateral inhibition which creates these bands is the same as the mathematics creating syllables and feet above. It is a system which recalculates until it reaches equilibrium (typically a small number of times).

In a 1- or 2-dimensional array of neurons, neurons:

- excite very close neighbors;
- inhibit neighbors in a wider neighborhood;
- do not affect cells further away

This can serve to create edge detectors.

Consider the following diagram, in which the lower line represents a 1-dimensional retina, and each spot it assigned a number, and in which the upper line represents the difference between the activation of each unit. The lower diagram represents the function f, and the higher diagram represents Δf, where $\Delta f(i) = f(i) - f(i-1)$.



In the next diagram, the function on top is

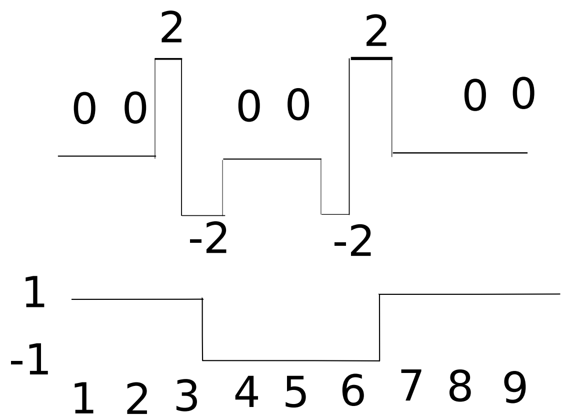
$$\Delta^2 f(i) = 2f(i) - f(i-1) - f(i+1)$$

although it is better to think of it as

$$\Delta^2 f(i) = [f(i) - f(i-1)] - [f(i+1) - f(i)]$$

which is the difference of two successive differences. That makes this the discrete equivalent of the second derivative.[4] A unit on the upper tier is (doubly) activated by its own corresponding element on the lower tier, and inhibited by the neighbor on its left and neighbor on its right. The total effect is to have no activation on this tier in a field of constant activation, and to have an exaggerated response to changes in activation, identifying "edges", areas where activation is changing.

[4]If we were doing this in more than one dimension, we would appeal to the *laplacian*, which is the sum of the second derivatives around a point.


2
2
0 0
0 0
0 0
-2
-2
1
-1
1 2 3 4 5 6 7 8 9