# Anything Goes

Alan Prince
Department of Linguistics & Center for Cognitive Science
Rutgers University, New Brunswick
Draft of 8/14/02


Abstract

Representing OT hierarchies in terms of weighted sums of violations requires in the general case exponential growth of weights; and any such weight system will only work for finite fragments of grammars in which the quantity of violations incurred by optima does not grow without bound (typically, those fragments in which input size is bounded). This conclusion is based on a worst-case analysis, and the worst is not always at hand. Noting that greater dominance can be minimally respected as greater weight without conditions on rate of weight growth, Paul Smolensky poses the following question: what OT systems are such that *any* dominance-respecting weighting whatever will recover the OT optima? These are systems in which 'anything goes' in the weighting systems (within the overarching restriction of weight-dominance concord). This note provides an answer to that question, and explores some systems in which anything goes and others in which it does not.


## 0. Introduction: OT and Weighting

OT works by strict domination: constraints are ranked into a total domination order; candidate *q* is better than candidate *z* iff *q* is better than *z* on the highest-ranking constraint that distinguishes them. (On a constraint, the candidate with fewer violations is the better one.)

*Strictness* comes in because no matter how poorly the better candidate fares on constraints ranked lower than the decisive constraint — no matter how massively it violates them — its failed competitor cannot be redeemed: the contest is over. From the OT point of view, there is no distinction between these comparisons:

(1) **Different Violations, Same Relations**

a.

| $q \succ z$ | $C_1$ | $C_2$ |
|---|---|---|
| q | 0 | 1 |
| z | 1 | 0 |

b.

| $q' \succ z'$ | $C_1$ | $C_2$ |
|---|---|---|
| q′ | 0 | 77 |
| z′ | 1 | 23 |

**Acknowledgment**. Thanks to Paul Smolensky for raising and discussing the issues treated here.

All that OT knows about any comparison *x* ~ *y* can therefore be represented in reduced form, using 'W' to mark constraints on which *x* betters *y*, 'L' to mark those on which *y* betters *x*.

(2) **Candidate Comparison**: $x \succ y$

| | $C_1$ | $C_2$ |
|---|---|---|
| x ~ y | W | L |

The logic of OT proceeds from this point (Prince & Smolensky 1993, Prince 2002).

Other ways of thinking about optimality treat such comparisons quantitatively, as in the Harmonic Grammar of Legendre, Miyata, and Smolensky 1990abc. From this point of view, the relationship between candidates is evaluated in terms of a weighted sum of constraint violations, where each constraint's weight represents the seriousness of violating it. We then calculate the numerical 'score' that each candidate earns, summing over the products (weight of constraint *k*) × (number of violations of constraint *k*). The candidate that earns the least score is the better one.

Suppose we assign a weight $w_1 = 2$ to $C_1$ and a weight $w_2=1$ to $C_2$. (Throughout we will use the notational convention that $w_k$ is the weight associated with constraint $C_k$.) The scores in example (1a) come out as follows:

$w_1=2,\ w_2=0$:

score = $(w_1\times v_{1)}) + (w_2\times v_{2)})$ , with $v_i$ = number of violations of $C_i$.

score(q) = $(2\times\mathbf{0}) + (1\times\mathbf{1}) = 1$

score(z) = $(2\times\mathbf{1}) + (1\times\mathbf{0}) = 2$

score(q) < score (z) $\quad\therefore\quad q\succ z$

The weighting system $(w_1, w_2) = (2,1)$ adequately reproduces the OT relation in example (1a). But things fall apart in (1b):

score(q′) = $(2\times\mathbf{0}) + (1\times\mathbf{77}) = 77$

score(z′) = $(2\times\mathbf{1}) + (1\times\mathbf{23}) = 25$

score(z′) < score(q′) $\quad\therefore\quad z' \succ q'$ (!)

Here the low scorer — the numerical winner – loses the OT competition. The performance of q′ on $C_2$ is bad enough to overshadow its success on the weightier but insufficiently dominant $C_1$.

For any comparison, or any finite set of comparisons, a weighting scheme can be found which yields the same results as strict domination. For example, in the case at hand, let $w_1=45$, $w_2=1$. Case (1b) now falls into line:

score(q′) = $(45\times\mathbf{0}) + (1\times\mathbf{77}) = 77$

score(z′) = $(45\times\mathbf{1}) + (1\times\mathbf{23}) = 78$

score(q′) < score(z′) $\quad\therefore\quad q' \succ z'$

The weight $w_1=45$ is chosen to overcome the violation deficit of 44 by which the desired optimum loses to its competitor on $C_2$.

Let us call a weighting scheme 'successful' if it accurately recapitulates the results of strict domination. Prince & Smolensky 1993 observe that a successful weighting scheme can be arrived at by selecting weights that increase *exponentially*. This result is obtained through a worst-case analysis which we sketch here.

Suppose we want *q* to beat *z* over a ranking $C_1$>> ... >>$C_n$ . Suppose further that V is the maximal number of violations accrued by *q* on any constraint. Then the worst possible situation faced by *q* at the bottom of the hierarchy is the following:

(3)

| | ... | $C_{n-1}$ | $C_n$ |
|---|---|---|---|
| q | ... | 0 | V |
| z | ... | 1 | 0 |

Assume that q and z fare equally on all constraints ranked above $C_{n-1}$. For convenience of calculation, let $w_n$=1. We now have:

score(q) < score(z) iff $w_{n-1} > V$

The smallest integer weight that will satisfy the inequality is $w_{n-1}$= V+1. (The same result will be reproduced in slightly messier form if we use noninteger weights.)

Now consider the worst situation that can happen one constraint up the hierarchy.

(4)

| | ... | $C_{n-2}$ | $C_{n-1}$ | $C_n$ |
|---|---|---|---|---|
| q | ... | 0 | V | V |
| z | ... | 1 | 0 | 0 |

From this, we may calculate the following:

$$\begin{aligned} w_{n-2} > (w_{n-1} \times V) + (1\times V) &= (V+1)\times V + V \\ &= V^2 + 2V \\ &= (V+1)^2 - 1 \end{aligned}$$

The smallest satisfactory integer weight will be $w_{n-2} = (V+1)^2$.

Continuing in this fashion, we will find that setting $w_{n-k} = (V+1)^k$ is guaranteed to reproduce strict domination. (Within the assumed restriction! — namely that the violations accrued by the desired winner stay under the V limit.)

This argument is developed from worst-case assumptions that are by no means guaranteed to hold in every candidate set and every hierarchy. Just because the maximum number of violations is V, we cannot conclude that *every* subordinated constraint can assign that many violations to a desired winner; nor that the desired winner faces a maximally-threatening competitor that entirely satisfies all lower-ranking constraints while just barely losing on the decisive constraint.

Consider the following competition within a version of the Basic Syllable Theory of Prince & Smolensky 1993. Assume ONS>>MAX>>DEP. All consonants "c" in this example are epenthetic; the sign ■ is inserted for convenience to indicate deletion.

(5) ONS>>MAX>>DEP, input /VV/

| | /VV/ | ONS | MAX | DEP |
|---|---|---|---|---|
| a. ☞ | cV.cV. | 0 | 0 | 2 |
| b. | cV.■ | 0 | 1 | 1 |
| c. | ■■ | 0 | 2 | 0 |
| d. | cV.V. | 1 | 0 | 1 |
| e. | V.■ | 1 | 1 | 0 |
| f. | .V.V. | 2 | 0 | 0 |

The maximum number of violations assessed by any constraint is 2; from the above argument, we know that an exponential weighting scheme based on 3 will work: $(w_1, w_2, w_3) = (9,3,1)$.

But the fact is, as Paul Smolensky has observed (p.c., Smolensky & Legendre, to appear), that *any weighting scheme at all* will select the OT optimum here, so long as the scheme respects the ranking relations, with higher rank corresponding to larger weight. (Specifically, $C_i >> C_j$ implies $w_i > w_j$, all weights positive). For example, the scheme $(w_1, w_2, w_3) = (1.2, 1.1, 1)$ will pick the correct optimum, CV.CV., assigning it the unique minimal score of 2, while the other scores range from 2.1 to 2.4.[1] This shows that there are OT systems that can be interpreted numerically in a manner that is as far as possible from the exigencies of exponential growth.

Smolensky asks for the conditions that will identify these systems. What patterns of violation will guarantee that the OT optimum is correctly selected by any weighting scheme that respects ranking? The aim of this note is to answer the question, and then to explore the character of some hierarchies which have the property, those for which *anything goes* (AG). We also take a look at the character of systems in which every ranking is AG, which we can call AG-factorial (AGF). The Basic Syllable Structure Theory (Prince & Smolensky 1993:§6) comes close, as the onset-based example (5) suggests, but we will find that when coda behavior is taken into account, the entire basic system falls outside the AGF domain.

To preview the main result, let us admit some terminology. Identify a candidate with its 'violation profile' (Samek-Lodovici & Prince 2002), a vector or n-tuple $\mathbf{q} = (q_1, \ldots, q_n)$ over n constraints ranked in the order $C_1 >> \cdots >> C_n$, where the 'coordinate' $q_i \geq 0$ records the number of violations of $C_i$. We now define the *n* 'accumulations' of a violation vector of length *n* to be the following coordinate sums, where $\Sigma^k(\mathbf{q})$ is the 'the kth accumulation of q':

$$\begin{aligned}
\Sigma^1(q) &= q_1 \\
\Sigma^2(q) &= q_1+q_2 \\
\Sigma^3(q) &= q_1+q_2+q_3 \\
\ldots & \quad \ldots \quad \ldots \\
\Sigma^n(q) &= q_1+q_2+ \ldots +q_n
\end{aligned}$$

[1] In this case, since the ranking between the top two constraints is not crucial, we could also get away with (1.1, 1.1, 1).

We are interested in the relative behavior of candidates under different weighting schemes, where a 'weighting scheme' is a vector of strictly descending positive weights $\mathbf{w} = (w_1,\dots, w_n)$, and the score of a candidate q is simply $\Sigma w_i q_i$, $1 \le i \le n$.

Given two profiles (violation vectors) **q,z**, with the OT relation $q \succ z$ over some hierarchy H of depth *n*, we want to know what it is about the structure of q and z that allows this relation to be captured by any positive descending weighting scheme whatsoever. The result runs like this:

(6) 'Anything Goes'. Assume $q \succ z$ in the OT sense over H. Then score(q)<score(z) for any positive descending weighting scheme iff $\Sigma^k(q) \le \Sigma^k(z)$ for all $i \le n$.

In short, any positive descending weighting scheme will capture the OT relation between q and z just in case no accumulation of q ever exceeds the corresponding accumulation of z.

Eyeballing the onset example (4), it easy to see that the candidate set has the AG property. All profiles accumulate to 2; the optimum (a) has 0 everywhere but in the last constraint. The accumulations of candidate (a) are 0 except in the last coordinate, at which point its accumulation is maximal. In fact, the other possible optima (c) and (f) have the same character, so that we may conclude that for this input at least, the onset mini-system is AG-factorial.

## 1. Anything Goes

Consider a generic 'constraint hierarchy' $H = C_1 \gg \cdots \gg C_n$, a totally-ordered sequence of constraints. For any candidate q, each constraint $C_i$ assigns a certain number of violations $q_i$ to q. The hierarchy H, via its order on the constraint set, associates with each candidate q a violation vector $\mathbf{q} = (q_1, \dots, q_n)$, where the $q_i$ are nonnegative integers.

Define a *weighting* of the hierarchy to be a vector $\mathbf{w} = (w_1, \dots, w_n)$, for $w_i$ real. We are only interested in weightings where $w_1 > \dots > w_n > 0$, that is where all weights are positive and descend strictly from the first coordinate to the last. When we want to emphasize this limitation, we will speak of a 'positive descending weighting' or PDW; but the term 'weighting' in this paper will refer only to PDW's. The descent in magnitude of the weights reflects the descent in dominance of the constraints in the associated hierarchy.

The *score* of vector **q** with respect to a weighting **w** is $\mathbf{w}\cdot\mathbf{q} = \Sigma w_i q_i$, $1 \le i \le n$. For this, we will write score(**q**,**w**) or simply score(**q**), suppressing the **w** argument when it is clear from context. This method of scoring allows us to compare all candidates with each other, and to define a notion of relative goodness based on that comparison.

(7) **Def**. '**Better than**' with respect to a weighting. Given candidates q, z, and a constraint hierarchy H, candidate q is 'better than' candidate z with respect to a weighting **w** on H iff score(**q**) < score(**z**), i.e. iff $\mathbf{w}\cdot\mathbf{q} < \mathbf{w}\cdot\mathbf{z}$, i.e. iff $\mathbf{w}\cdot(\mathbf{z}-\mathbf{q}) > 0$. In this case, we write $q \succ_{\mathbf{w}} z$.

With this definition, we have constructed a subspecies of Harmonic Grammar (Legendre, Miyata, Smolensky 1990abc, Smolensky & Legendre, to appear). The species is strictly sub, because Harmonic grammar allows for both positive and negative weights, and does not require that constraints be totally ordered according to weight. What we have called the ‘score’ of a violation vector corresponds inversely to its ‘Harmony’: high Harmony ≈ low score. This is a mere notational difference, which circumvents some arithmetic bother in the current context.

This weight-based notion of ‘better than’ may be contrasted with that of OT.

(8) **Def**. ‘**Better than**’ with respect to strict domination. Given candidates q, z, and a constraint hierarchy H, candidate q is ‘better than’ candidate z with respect to the notion of strict domination iff there is a constraint $C_k$ in H such that $q_k<z_k$ and $q_i=z_i$ for all i<k. In this case, we write $q \succ_{OT} z$.

This definition imposes a ‘lexicographic order’ on the candidate set: order is assessed first in the first coordinate (**a**pricot < **z**ebra); ties there are referred to the next (*a***m**ple < *a***p**ple); ties there to the next (*ap***p**le < *ap***r**icot); and so on.

A candidate is *optimal* in its candidate set just in case it is *best*, unbettered, unworsted, maximal in the ‘better than’ order, however that order is defined.

The two different notions converge dramatically when the optima in the OT sense are exactly the same as the optima in the weighting sense, no matter what weighting is chosen, so long as it is positive-descending. Let us call this situation ‘anything goes’ (AG), and define it as a property of a inter-candidate relations with respect to a ranking of constraints.

(9) **Def. Anything Goes (AG).**. Let K be a candidate set, and ω the OT optimum under a given ranking R of a set of constraints.

• If $q \succ_{OT} z$ under R, for q,z∈ K, the relation between q and z is ‘AG’ with respect to R iff for every positive descending weighting **w** of R, $q \succ_w z$, i.e. score(q)<score(z).

• If ω is the OT optimum of K under R, then K is ‘AG’ with respect to R iff for every positive descending weighting of R, score(ω) < score(z) for all z∈K with **z≠ω**.

The two clauses of the definition are really the same: when we have q≻z, the set {q,z} may be taken to be a candidate set in which q is optimal (Samek-Lodovici & Prince 1999:38). When $q \succ_{OT} z$ is AG, we will write $q \succ_{AG} z$, where (as with $q \succ_{OT} z$) the ranking R is to be understood from context.

To see how the AG property comes about, we need one further notion: the sum of the first *k* entries of a vector, which we will call its ‘$k^{th}$ accumulation’.

(10) **Def. Accumulation**. For any violation vector **v** = $(v_1,\ldots, v_n)$, an *accumulation* is the sum of some initial sequence of coordinates. The $k^{th}$ accumulation is the sum of the first *k* coordinates, which we will abbreviate as $\Sigma^k(\mathbf{v})$ for a violation vector **v**.

$$k^{th} \text{ accumulation of } \mathbf{v},\ \Sigma^k(\mathbf{v}) := \sum_{i=1}^{k} v_i$$

The sum of all coordinates of **v** will be referred to as its ‘total accumulation’ and for this we use the indexless notation $\Sigma(\mathbf{v})$.

To approach the matter of identity of optima, let us examine the preservation of candidate order. We want a condition on the violation structure of any q and any z guaranteeing that $q \succ_{OT} z$ entails $q \succ_{\mathbf{w}} z$ for all positive-descending **w**, i.e. that $q \succ_{AG} z$. This condition will hold generally of pairs of candidates, whether the optimum is involved or not. Extension of this result to the reasoning about the optima is straightforward: if q is OT-optimal, then $q \succ_{OT} z$ *for all* $\mathbf{z} \neq \mathbf{q}$ in the candidate set. To demonstrate preservation of the optimum in any given case, we can simply check that the condition guaranteeing $q \succ_{AG} z$ holds *for all competitors z*.

The determining condition concerns the relationship between corresponding accumulations of the two violation vectors. To guarantee that anything goes, preserving $q \succ_{OT} z$, we must have it that the $k^{th}$ accumulation of q never exceeds the $k^{th}$ accumulation of z, for all $k \leq n$, where n is number of constraints.

(11) **Theorem**. **Anything Goes**. Let q,z be competing candidates evaluated by a constraint hierarchy $H = C_1 >> .... >> C_n$. Suppose $q \succ_{OT} z$. Then $q \succ_{\mathbf{w}} z$ for all positive-descending **w** iff $\Sigma^k(\mathbf{q}) \leq \Sigma^k(\mathbf{z})$.

Before we prove this, it is useful to observe two general properties of positive-descending weightings.

(12) **Remark 1. Zeroing Out**. Let **x** be an arbitrary vector with real coordinates (which may be positive, negative, or zero). We can choose a PDW **w** that makes score(**x**) as close to 0 as we want.

*Pf.* Note that score($\mathbf{x}$)= $\mathbf{w} \cdot \mathbf{x}$ is bounded above by by $\mathbf{w} \cdot (|x_1|, \ldots, |x_n|)$ and bounded below by $\mathbf{w} \cdot (-|x_1|, \ldots, -|x_n|)$. Either of these can be squeezed as close to 0 as we want by a PDW. To see this, let $x_{max}$ be the largest coordinate in $(|x_1|, \ldots, |x_n|)$. If we want score($\mathbf{x}$)<ε for some ε>0, choose $w_i < \varepsilon/(n \cdot x_{max})$ for all i. □

(13) **Remark 2**. **Near Miss**. Let **x** be an arbitrary vector with real coordinates (which may be positive, negative, or zero). We can choose a PDW **w** so as to make score(**x**) as close as we want to the total accumulation of **x**, $\Sigma x_i$, $i \leq n$.

*Pf.* Let $\mathbf{w} = (w_1, \ldots w_n)$ be a PDW. Consider $\mathbf{w}' = (1+w_1, 1+w_2, \ldots, 1+w_n)$; this is also a PDW. Now choose the $w_i$ to make $\Sigma w_i x_i$ as close to 0 as we want $\mathbf{w} \cdot \mathbf{x}$ to be to the total accumulation of **x**, i.e. $\Sigma x_i$, $1 \leq i \leq n$. □

Let us now proceed to the main result. The accumulation condition $\Sigma^k(\mathbf{q}) \leq \Sigma^k(\mathbf{z})$ can be conveniently and equivalently re-phrased in terms of the difference vector $\mathbf{v} = \mathbf{z} - \mathbf{q}$, requiring $\Sigma^k(\mathbf{z} - \mathbf{q}) = \Sigma^k(\mathbf{v}) \geq 0$. To wit: all accumulations of **v** must be nonnegative. We use this formulation in the proof.

**Proof of Theorem**. Let q, z be such that $q \succ_{OT} z$ on $H = C_1 >> .... >> C_n$ Let $\mathbf{v} = \mathbf{z} - \mathbf{q}$. We want to show that $\Sigma^k(\mathbf{v}) \geq 0$ for all $k \leq n$ iff $q \succ_{\mathbf{w}} z$ for all positive descending weightings **w**.

• Let us first establish the RL direction of implication through its contrapositive. Assume that $\Sigma^k(\mathbf{v})<0$ for some k. We show that there is a positive descending weighting $\mathbf{w}$ such that $z \succ_{\mathbf{w}} q$, thereby running counter to the assumed OT relation $q \succ_{OT} z$.

Consider the subvector $\mathbf{v}^k=(v_1,...v_k)$. Because $\Sigma^k(\mathbf{v})<0$, we are guaranteed by Remark 2 a weighting $\mathbf{w}^k$ such that $\mathbf{w}^k \cdot \mathbf{v}^k <0$, since we can make this score as close to $\Sigma^k(\mathbf{v})$ as we like. Say $\mathbf{w}^k \cdot \mathbf{v}^k = -d$, for some d>0.

Now consider the rest of $\mathbf{v}$, call it $\mathbf{v}^{kn} = (v_{k+1},\ldots,v_n)$. By Remark 1, we can find a weighting $\mathbf{w}^{kn}$ which will drive the score of $\mathbf{v}^{kn}$ as close to 0 as desired. Choose $\mathbf{w}^{kn}$ so that $\mathbf{w}^{kn} \cdot \mathbf{v}^{kn} < d$.

Now define a weighting $\mathbf{w}$ such that its first k coordinates are just those of $\mathbf{w}^k$. We now derive its remaining coordinates from those of $\mathbf{w}^{kn}$. (We write $\mathbf{x}[k]$ for the $k^{th}$ coordinate of $\mathbf{x}$.) Choose some positive factor $\alpha$ $\alpha \leq 1$, such that $\mathbf{w}^k[k] > \alpha \cdot \mathbf{w}^{kn}[1]$. Now use the coordinates of $\alpha \cdot \mathbf{w}^{kn}$, in order, to supply the coordinates $w_{k+1},\ldots,w_n$ of $\mathbf{w}$. By choice of $\alpha$, $\mathbf{w}$ is positive descending. Observe further that $\alpha \cdot \mathbf{w}^{kn} \cdot \mathbf{v}^{kn} = \alpha d<d$. From this, $\mathbf{w} \cdot \mathbf{v} = -d+\alpha d<0$. This establishes the claim, since $\mathbf{w} \cdot (\mathbf{z}-\mathbf{q}) <0$ means score($\mathbf{z}$)<score($\mathbf{q}$).

• Let us now turn to the (more useful) LR direction of implication. We have $q \succ_{OT} z$ on H. Assume that all accumulations of $\mathbf{v} = \mathbf{z} - \mathbf{q}$ are nonnegative. We wish to show that $q \succ_{\mathbf{w}} z$ for any positive descending $\mathbf{w}$. Specifically, we need to demonstrate that the overall score of $\mathbf{v}$ is positive under any such $\mathbf{w}$: in short, $\mathbf{w} \cdot \mathbf{v} = \Sigma w_i v_i >0$.

For convenience, let us cast away all coordinates $v_i$ for which $v_i=0$ and construct a vector $\mathbf{v}'$ which is exactly like $\mathbf{v}$ except that all 0 coordinates have been removed. Because $q \succ_{OT} z$ *ex hypothesi*, there must be at least one non-zero coordinate in $\mathbf{v}$, and hence at least one coordinate in $\mathbf{v}'$, so $\mathbf{v}'$ is a licit vector. And clearly score($\mathbf{v}'$)=score($\mathbf{v}$) for any weighting, since the 0-coordinates add nothing to the score.

Divide $\mathbf{v}'$ into a sequence of positive and negative blocks, where each block contains a sequence of coordinates of only one sign. Let us designate a positive block as [+m], where m is the sum of the coordinates in the block; and similarly, let us designate negative blocks as [−k]. Each block corresponds to a contiguous sequence of constraints which all prefer the same competitor in the q *vs.* z comparison. The block [+m] corresponds to constraints on which q beats z by a total margin of m marks (i.e., there are m excess marks borne by z); the block [−k] consists of constraints preferring z by a total margin of k marks (the excess by which q loses to z). Given a weighting, we will find that every negative mark can be matched to a positive mark of greater weight, leading to a positive overall score, as desired. The rest of the proof consists of nothing more than cashing in this observation.

Let $\mathbf{w}$ be any positive descending weighting. Let $\mathbf{w}'$ be derived from $\mathbf{w}$ in the same way as $\mathbf{v}'$ is derived from $\mathbf{v}$, by omitting the coordinates $w_i$ for which $v_i=0$. If $\mathbf{v}'$ has only one block, then it must be positive. This is because the first coordinate of $\mathbf{v}'$ comes from the highest-ranked constraint that distinguishes q and z. Since $q \succ_{OT} z$, that first coordinate must be positive, and the block it belongs to will also be positive. For any $\mathbf{w}$, then, we have $\mathbf{w}' \cdot \mathbf{v}' = \mathbf{w} \cdot \mathbf{v} >0$ for such a monoblock $\mathbf{v}'$.

Now suppose that $\mathbf{v}'$ has more than one block and consider its block sequence, which we may represent as $[+m_1][-k_1]\ldots$ . Let B[ij] represent a block running from the $i^{th}$ to the jt[h] coordinate, inclusive. The score associated with a block B[ij] is $\Sigma w_k v_k$, with k running from i to j. We may expand this score as the simple sum of numbers drawn (sometimes repeatedly) from the weight set $\{w_i', \ldots, w_j'\}$, with $w_k'$ appearing $v_k'$ times. Call this the 'expanded sum'. The accumulation of the block B[ij], namely $\Sigma v_k'$, $i \le k \le j$, gives the total number of terms in the expanded sum.

Let us now examine the first two blocks of $\mathbf{v}'$, $[+m_1][-k_1]$. The score of the block $[+m_1]$ can be written as an expanded sum with exactly $m_1$ components; that of $[-k_1]$ as an expanded sum with exactly $k_1$ components (each negatively signed), where each 'component' is a weight appearing in the appropriate coordinate of $\mathbf{w}'$. We may now estimate the score associated with $[+m_1][-k_1]$ by matching each weight in the expanded sum associated with $[-k_1]$ to a unique weight in the expanded sum of $[+m_1]$. The number of unmatched weights is exactly $m_1 - k_1$. This is just the total accumulation over the blocks $[+m_1][-k_1]$. By hypothesis, this is nonnegative, i.e. we have $m_1 - k_1 \ge 0$.

It follows immediately that score($[+m_1][-k_1]$)>0. We have matched each weight in the expanded sum associated with $[-k_1]$ to a greater weight from $[+m_1]$. (The assumption that the weights are positive descending guarantees the weight size differential.)

We may now simply repeat this reasoning about the weight match-up in the general context. Suppose, as the induction hypothesis, that it is true that we have successfully matched each negative weight to a greater positive weight in the expanded sum associated with the first p pairs of blocks, where $p \ge 1$. (It follows that the score after p block pairs is strictly positive.) If there is no further pair, we are done: either we've reached the end of $\mathbf{v}'$ or the remaining block is of the form $[+m_{p+1}]$, and including it will only raise the already positive score. Now assume that there is a (p+1)[st] block pair. We need only continue the matching procedure, taking into account the (p+1)[st] negative block. By hypothesis, the number of unmatched weights at the end of the p[th] block is $\Sigma(m_i - k_i)$, $1 \le i \le p$, which is the accumulation of the first p blocks, and nonnegative. To this we now add the $m_{p+1}$ positive weights of block $[+m_{p+1}]$. By the assumption that all accumulations are nonnegative, we have it that $\Sigma(m_i - k_i) + m_{p+1} \ge k_{p+1}$. This means that the $k_{p+1}$ negative weights of the (p+1)[st] block may be matched with greater positive weights (greater by the assumption of positive descending weights), and we are done. □

As an immediate corollary, we may deduce that if a candidate set is AG under some ranking, no member of the candidate set may have a total accumulation smaller than that of the OT optimum.

(14) **Corollary. AG Accumulation Limit.** Let $\alpha, \omega \in K$, a candidate set, with $\omega$ the OT optimum, and $\alpha$ OT-suboptimal, over H= $C_1 >> \cdots >> C_n$. Suppose K is AG over H. Then $\Sigma(\boldsymbol{\alpha}) \ge \Sigma(\boldsymbol{\omega})$.

*Pf.* Since $\omega \succ_{OT} \alpha$, the theorem gives us $\Sigma^k(\boldsymbol{\alpha}) \ge \Sigma^k(\boldsymbol{\omega})$ for all k, including n. □

Seen in the context of constraints and violations, the proof is telling us that AG weighting schemes work when all the marks incurred by the desired winner (q) can be discounted in a 1:1 fashion against marks incurred by the desired loser (z), where in each case of mark-matching the loser's mark is assessed by the same constraint as the winner's mark is, or by a constraint ranked higher.

We can observe the AG theorem in action in the onset example (5). Consider the competion between the first two candidates, repeated here for convenience:

(15) **AG**: $a \succ_{OT} b$ *and* $a \succ_{AG} b$

| | /VV/ | ONS | MAX | DEP |
|---|---|---|---|---|
| a. | cV.cV. | 0 | 0 | 2 |
| b. | cV.■ | 0 | 1 | 1 |

The marks incurred by candidate (a), namely **DEP, are discounted against *MAX (higher-ranked) and *DEP (same rank).

The proof also tells us that the $k^{th}$ accumulation of the difference vector (here $\mathbf{b}-\mathbf{a}$) keeps track of the number of loser's marks that are not discounted 1:1 against winner's mark at the level of the $k^{th}$ constraint in the hierarchy. This is a kind of reserve fund of difference which cannot be allowed to go in the red, without loss of AG standing.[2]

The *accumulation* might seem like an odd construction, but it is not entirely foreign to ordinary OT. Consider any competition of candidates over a constraint hierarchy H. If, for each candidate, we replace its marks from $C_k$ by the $k^{th}$ accumulation of its marks, the outcome of the competition will be entirely unaffected. Let us call this replacement the 'accumulated representation' of the contest q~z, and write $\mathbf{q}'$ and $\mathbf{z}'$ for the accumulated violation profiles.

To see the equivalence of ordinary and accumulated representations, suppose we have $q \succ z$. Then there is a sequence, possibly empty, of constraints on which q and z incur identical marks. Their accumulations are also identical on this sequence. Then a decisive constraint $C_k$ is encountered, on which z receives more marks than q. But at this point z's $k^{th}$ accumulation is also boosted to be greater than q's, and so it is precisely at this point in the accumulated representation that q betters z. So $C_k$ is as decisive in the accumulated representation as it is in the ordinary representation of violation structure.

In terms of the accumulated representation of a hierarchy, we may rephrase the theorem as saying that (given $q \succ_{OT} z$) we have $q \succ_{AG} z$ iff $\mathbf{q}'$ *harmonically bounds* $\mathbf{z}'$, in precisely the classic sense of this term: namely, that at each coordinate $q_i' \le z_i'$ and at some coordinate $q_k' < z_k'$.

The equivalence between ordinary and accumulated representations is useful in the study of Paninian or special/general relations between constraints (Prince 1997 et seq.). Suppose we have a hierarchy on n constraints of the form $\langle *a_k \rangle$, with each constraint banning the element $a_k$. Let the hierarchy run $\langle *a_1 \rangle >> \cdots >> \langle *a_n \rangle$. Suppose further that the constraints are disjoint in the sense that no element of the type banned by the constraints is ever of both type $a_i$ and $a_j$ for $i \neq j$. (This ensures that each constraint assigns marks to a different substructure.) An example would be the peak-prominence hierarchy, *í>>*é>>á. No vowel is ever both i and a at the same time.

The accumulated representation of this hierarchy can be understood as deriving from constraints $\langle *a_1 \rangle >> \langle *a_1, *a_2 \rangle >> \langle *a_1, *a_2, *a_3 \rangle >> \cdots >> \langle *a_1, \ldots, *a_n \rangle$. These constraints stand in a

[2] I write in early August, 2002.

special/general or ‘stringency’ relationship: each succeeding constraint bans a strict superset of the elements banned by its predecessors (hence is more ‘stringent’):

$$\{a_1\} \subseteq \{a_1,a_2\} \subseteq \{a_1,a_2,a_3\} \subseteq \cdots \subseteq \{a_1, \ldots, a_n\}$$

Because the ordinary hierarchy and its accumulated version yield the same comparative outcome, we can deduce that a family of constraints that have the stringency relationship, *when ranked in this fashion*, from less to more stringent, is precisely equivalent to a hierarchy of single-element constraints using the constraints $\langle *a_i \rangle$, when ranked so to accumulate to the stringency hierarchy. Of course, when constraints of the stringency form are ranked in different ways, interacting appropriately with conflicting constraints, there is no equivalence and different properties emerge (Prince 1997 et seq., de Lacy 2002). This argument shows that the difference between element-based constraint systems with fixed ranking and their stringency-based relatives is to be found only in the ‘non-paninian’ rankings of the stringency system — those rankings which are distinct from the special-to-general, less-to-more-stringent ranking order just displayed.

## 2. Anything Goes, Factorially

A system of n constraints allows n! different rankings, but quite special conditions are required to achieve anything like n! possible optima, one for each distinct ranking. Here’s an abstract example that works for three constraints.

(16) **P(3)**

| | $C_1$ | $C_2$ | $C_3$ |
|---|---|---|---|
| a | 0 | 1 | 2 |
| b | 0 | 2 | 1 |
| c | 1 | 0 | 2 |
| d | 1 | 2 | 0 |
| e | 2 | 0 | 1 |
| f | 2 | 1 | 0 |

This is an instance of what Prince & Samek-Lodovici (in prep.) call “perfect OT” – every ranking has its own optimum, and this result is achieved with the smallest possible number of violations. It represents P(3) – the unique perfect system on 3 constraints.

It is easily determined that this system is AG. Under the ranking $C_1 \gg C_2 \gg C_3$, candidate (a) is optimal. Evaluating its relationship to the others, we see that its 2nd accumulation is minimal (being 1) and its 3rd or total accumulation is also minimal (as well as being maximal, and the same as every other candidate’s!). Furthermore, due to the extreme symmetry of the system, the outlook for any optimum is exactly the same — every optimum will have the violation profile (0,1,2) under the ranking that yields it — and we may therefore conclude that the entire system, under all ranking permutations, is AG; let’s call this state of affairs ‘AG Factorial’, or AGF.

(17) **Def. AG Factorial (AGF)**. A candidate set K is AG Factorial (AGF) with respect to a set of constraints $\mathbb{C}$ iff it is AG with respect to every ranking of $\mathbb{C}$.

While an AG ranking, and a fortiori an AGF system, produces the same *optima* under any weighting as its OT counterpart, there is no guarantee that the entire structure of 'better than' orderings will be preserved in the depths of the candidate set. Whether this happens depends on whether the individual pairwise relations are also AG. In the system (16), for example, while the optimum is AG with respect to every other candidate, certain other relations within the candidate set are less stable. The relation $b \succ_{OT} c$ under the $C_1 \gg C_2 \gg C_3$ ranking of (16) is not AG, nor is $d \succ_{OT} e$, as may be seen by examining the 2nd accumulations of these candidates in the following accumulated representation:

(18) **Accumulated Representation** of P(3)

| | $C_1$ | $C_1+C_2$ | $C_1+C_2+C_3$ |
|---|---|---|---|
| a′ | 0 | 1 | 3 |
| **b′** | 0 | **2** | 3 |
| **c′** | 1 | **1** | 3 |
| *d′* | 1 | *3* | 3 |
| *e′* | 2 | *2* | 3 |
| f′ | 2 | 3 | 3 |

Whereas the OT order on the candidate set is strict, and runs from a to f, the orders guaranteed to be AG form only a partial order: $\{a\} \succ_{AG} \{b,c\} \succ_{AG} \{d,e\} \succ_{AG} \{f\}$. Observe that '$\succ_{AG}$' can fail to recognize but cannot reverse an OT relation, even deep in the candidate set. If $q \succ_{OT} z$, there is certainly some weighting w on which $q \succ_{w} z$, therefore it cannot be that $z \succ_{AG} q$.

In the general OT situation, which is not AG, much less AGF, the effects of weighting can be more diverse. An unsuccessful weighting can cause possible OT optima to be nonoptimal, and can even allow candidates that *cannot be* OT optima to emerge as optimal. Consider this example.

(19) **Possible Optimum Denied**

| **w** = (2,1) | $C_1$ | $C_2$ | score:$C_1 \gg C_2$ | score: $C_2 \gg C_1$ |
|---|---|---|---|---|
| a | 0 | 3 | 3 | 6 |
| b | 1 | 0 | 2 | 1 |

Candidate (a) is the OT optimum under $C_1 \gg C_2$, but it always earns the greater score under the unsuccessful weighting (2,1), whatever the ranking.

(20) **Nonoptimum Promoted**

| **w** = (2,1) | $C_1$ | $C_2$ | score:$C_1$>>$C_2$ | $C_2$>>$C_1$ |
|---|---|---|---|---|
| a. | 0 | 4 | 8 | 4 |
| b. | 1 | 1 | 3 | 3 |
| c. | 4 | 0 | 4 | 8 |

Candidate (b) can never win under OT: on one ranking it loses to (a); on the other, to (b). It is 'collectively bounded', in the terminology of Samek-Lodovici & Prince 1999. Observe that the total accumulation of (b) is less than that of (a) or (c).

Simple harmonic bounding (whereby one candidate is bounded by another rather than by several others) is, however, impervious to weighting. This is an immediate consequence of the AG theorem, for if we have $q_i \le z_i$ for all i, – z bounded by q – we must have $\Sigma^k(q) \le \Sigma^k(z)$ as well, no matter how the coordinates are ordered, and the relationship $q \succ z$ will be preserved under all weightings. This allows us specify rather narrowly the kind of damage that can be done by unsuccessful weightings: OT optima may be lost, but the only OT nonoptima that can be promoted to optimal status are those that are collectively bounded. (And indeed, those where the collective bounding relationship is not AG.) If we can show, in a given case, that there happen to be *no* collectively-bounded candidates in the candidate set, or that the relations giving rise to collective bounding happen to be AG, then we are guaranteed a fairly graceful degradation under weighting – the mere loss of possible optima is the worst that can happen.

We conclude this discussion with an observation about the conditions under which AGF behavior is worth looking for. If a candidate set is AGF under some set of constraints, then the possible optima must all have the same total accumulation. The proof turns on the fact that the total accumulation is invariant over rankings.

(21) **Remark 3**. **Constant Accumulation in AGF Systems**. Let $\alpha$, $\omega \in$K, a candidate set, be optimal under rankings $R_1$,$R_2$, respectively, of a given set of n constraints, and suppose that K is AG under $R_1$ and $R_2$. Then $\Sigma(\boldsymbol{\alpha}) = \Sigma(\boldsymbol{\omega})$, *i.e.*

$$\sum_{i=1}^{n} \alpha_i = \sum_{i=1}^{n} \omega_i$$

*Pf.* Let $\omega_1$, $\omega_2 \in$K be optimal under rankings $R_1$,$R_2$ respectively. Since $\omega_1 \succ_{AG} \omega_2$ under $R_1$, we have from the AG theorem, $\Sigma(\omega_1) \le \Sigma(\omega_2)$. Since $\omega_2 \succ_{AG} \omega_1$ under $R_2$, we have $\Sigma(\omega_2) \le \Sigma(\omega_1)$. Because the total accumulation of any candidate is the same for every ranking, it follows that $\Sigma(\alpha) = \Sigma(\omega)$. ☐

From this, the claimed condition on the whole set of optima immediately follows.

(22) **The Simplex Condition.** If K, a candidate set, is AGF over a given set of constraints, then all the possible optima of K have the same total accumulation.

Pf. Remark 3 shows this holds for every pair of optima. By transitivity of equality, their total accumulations must all be the same. ☐

As a corollary, we may deduce that in an AGF system, no candidate may have a total accumulation *less than* that of the optima.

(23) **Remark 4**. **AGF Accumulation Limit**. Let K be a candidate set that is AGF with respect to some set of constraints. Let $\Sigma_\omega$ be the total accumulation of an optimum. Then for $\alpha \in K$, if $\alpha$ is never optimal under any ranking, $\Sigma(\alpha) \geq \Sigma_\omega$.

*Pf.* From Corollary (14), we know that $\Sigma(\alpha) \geq \Sigma(\omega)$, for any optimum $\omega$. But $\Sigma(\omega) = \Sigma_\omega$. □

It is important to note that we have arrived at necessary rather than sufficient conditions for AGF status. It is entirely possible for the Simplex Condition to be met in a non-AGF system. Consider these two candidates, which have the same total accumulation:

(24)

| | $C_1$ | $C_2$ | $C_3$ |
|---|---|---|---|
| q | 0 | 5 | 1 |
| z | 1 | 2 | 3 |

Candidate q is the OT optimum, yet $q \succ z$ cannot be AG, because $\Sigma^2(\mathbf{q}) > \Sigma^2(\mathbf{z})$. To arrive at a sufficient condition in any given case requires examination of all the accumulations, not just the total accumulation.

The technique for arriving at sufficient conditions generalizes that used in the proof of Remark 3, ex. (21). Suppose that we have two optima $\alpha$ and $\beta$, from rankings $R_1$ and $R_2$, respectively, and suppose further that $R_1$ and $R_2$ are related by a single swap of adjacent constraints.

$\alpha \leftrightarrow C_1 >> \cdots >> \mathbf{C_k} >> \mathbf{C_{k+1}} >> \cdots >> Cn = R_1$

$\beta \leftrightarrow C_1 >> \cdots >> \mathbf{C_{k+1}} >> \mathbf{C_k} >> \cdots >> Cn = R_2$

It is evident that $\alpha$ and $\beta$ must have identical violations on each of $C_1$ through $C_{k-1}$, i.e. $\alpha_i = \beta_i$ for $i<k$. We also have $\alpha_k < \beta_k$ and $\beta_{k+1} < \alpha_{k+1}$. Furthermore, since $\Sigma^{k+1}(\alpha) \leq \Sigma^{k+1}(\beta)$ on $R_1$ and $\Sigma^{k+1}(\beta) \leq \Sigma^{k+1}(\alpha)$ on $R_2$, and since the $(k+1)^{st}$ accumulation of $\alpha$ (resp. $\beta$) is the same under both rankings, we conclude that

$$\sum_{i=1}^{k+1} \alpha_i = \sum_{i=1}^{k+1} \beta_i$$

But since $\alpha_i = \beta_i$ for $i<k$, we have $\alpha_k + \alpha_{k+1} = \beta_k + \beta_{k+1}$.

Continuing in this fashion, we may construct an ever-tightening net of conditions that any array of violation vectors must meet if it is going to be AGF under the ranking conditions that produce distinct optima.

As an example of this kind of approach, let us generalize P(3) from ex. (16) to include all 3 constraint systems that produce 6 distinct optima while retaining the AGF property. Suppose we start with an arbitrary violation profile **α**=(a,b,c), assumed to win on $C_1$>>$C_2$>>$C_3$; we wish to construct a 6-optimum AGF system around it. Consider the optimum **β** of $C_2$>> $C_1$>>$C_3$, which is just one swap away. From the above discussion we have

$$\alpha_1+\alpha_2 = \beta_1+\beta_2$$

and, in addition,. by the Simplex Condition (22),

$$\alpha_1+\alpha_2+\alpha_3 = \beta_1+\beta_2 + \beta_3 .$$

From these, it follows that $\alpha_3=\beta_3 = c$.

Since $a=\alpha_1<\beta_1$, we can write $\beta_1=a+d$, for some d>0.But then

$$\beta_2 = \alpha_1+\alpha_2-\beta_1 = a + b - (a+d) = b-d .$$

Putting this all together, we find

$$\boldsymbol{\beta} = (a+d,\ b-d,\ c).$$

If we follow this kind of calculation through, we can construct a pattern that all AGF 3-constraint, 6 optimum systems must meet – necessary and sufficient conditions.

In the end, the requirements may be reduced to the following. There is a basic matrix of values that looks like this:

(25) **P(3) extended as AGF**: d,f>0

| | $C_1$ | $C_2$ | $C_3$ |
|---|---|---|---|
| i | 0 | d | d+f |
| ii | 0 | d+f | d |
| iii | d | 0 | d+f |
| iv | d | d+f | 0 |
| v | d+f | 0 | d |
| vi | d+f | d | 0 |

Any 3-constraint, 6-optimum AGF system can be derived by choosing d,f>0 and adding (a,b,c) coordinatewise to each row, for some pick of positive a,b,c≥0.[3] P(3) can be arrived at by setting (a,b,c)=(0,0,0) and d=f=1, the smallest integer choices for each parameter. The extension to general n-constraint, n! optimum systems is straightforward.

[3] What happened to the (a+d, b−d, c) characterization of candidate (iii) from our initial reckoning? If we uniformly add a constant to the violations assessed by a constraint in every candidate, we do not affect the OT outcome. To arrive at matrix (25), continue the calculations in the text, then subtract a, b, c from $C_1$, $C_2$, $C_3$ respectively, and then add d to $C_2$ and d+f to $C_3$.

## 3. Basic Syllable Theory

The basic syllable theory (BST) of P&S 1993:§6 has significant AGF substructures, though in the end it turns out that the whole theory is not AGF. In this section, we take a brief first-look at some of the properties of the BST. To investigate it, we need to be precise about what it is: let us cast the theory in the following slightly simplified form.

- Input: any string from {C,V}+, where C, V are primitive elements.
- Output: fully syllabified, where a syllable is a constituent of the form (C)V(C).
- Constraints:

| | |
|---|---|
| ONS | The initial segment of a syllable is C. |
| NOCODA | The final segment of a syllable is V. |
| MAX | No deletion. |
| DEP | No insertion. |

We gloss over the development of correspondence theory (McCarthy & Prince 1995), which gives an explicit account of 'deletion' and 'insertion'. We also conflate DEP-V and DEP-C, MAX-V and MAX-C, without losing any generality in the discussion.

In the original treatment, the element 'C' is restricted to marginal position and the element 'V' to nuclear position via constraints *P/C, *M/V, which are assumed to be undominated, along with *COMPLEX, which limits margins to less that 2 consonants. For present purposes, we incorporate these into Gen, as part of the basic definition of what linguistic structures the candidate sets will contain. This is not quite arbitrary. Suppose that *P/C, banning C from peaks, were a mere constraint – even a top-ranked one. We can immediately see that the resulting system is going to fail to be AGF. Consider this (quite partial) candidate set:

(26)

| /CC/ | | Ons | NoCoda | Max | Dep | *P/C |
|---|---|---|---|---|---|---|
| a. | $.C[C]_{nuc}.$ | 0 | 0 | 0 | 0 | 1 |
| b. | .Cv.Cv. | 0 | 0 | 0 | 2 | 0 |

Candidate (b) is an OT optimum of the system, with a total accumulation of 2. If there is to be any hope of AGF behavior, the total accumulation of any optimum must be 2. But candidate (a), with a nuclearized C, has a total accumulation of only 1. Therefore, to reach even the necessary Simplex Condition, we must exile candidate (a), and we do so by artificially strengthening Gen to ban it.

Let us first establish that BST, as presented here, actually meets the Simplex Condition. To do this, let us annotate input forms, purely for calculational purposes, in the following way.

(i) If a C occurs anywhere but immediately before a V, insert a ◂ after it.

(ii) If a V occurs anywhere but immediately after a C, insert a ▸ after it.

The input /CC/, for example, will be annotated as 'C◂C ◂', the input /VV/ as '▸V▸V' and so on.

The annotations serve to identify (all) problematic zones in the input, those C or V whose presence leads to constraint violations in the output. The annotated inputs can be related to outputs in the following ways. (For convenience, we write inserted C as ‘c’, inserted V as ‘v’, though it should be remembered that these instances of C and V are not distinct new elements of structure.)

Interpretation of ▸. The right-pointing wedge is placed before problematic V. It may be ‘realized’ in optima in one of three ways:

(i) The ▸ is replaced by a C, resulting in *DEP.
Example: /V/ = ▸V → .cV.

(ii) The vowel following the ▸ may be deleted, resulting in *MAX.
Example: /V/ = ▸V → Ø.

(iii) The ▸V configuration may be realized faithfully as V. *ONS.
Example: /V/ = ▸V → .V.

Interpretation of ◂. The left-looking wedge is placed after C. It may be ‘realized’ in optima in one of three ways:

(i) The ◂ is replaced by a V, resulting in *DEP.
Example: /CVC/ = CVC◂ → .CV.Cv.

(ii) The consonant preceding the ◂ may be deleted, resulting in *MAX.
Example: /CVC/ = CVC◂ → .CV.Ø

(iii) The C◂ configuration may be realized faithfully as C, resulting in *NOCODA.
Example: /CVC/ = VC◂ → .CVC.

Two basic observations may be made about the annotated representation.

[1] Given an input, *no* candidate has fewer total marks than the annotated representation has wedges. Every annotation must be realized as something that earns a mark. It is always possible to garner more marks, by senseless insertion for example, but it is never possible to have fewer.

[2] Given an input, the possible *optima* from that input have precisely the same number of total marks — the same total accumulation — and that is equal to the number of wedges.

These indicate that BST satisfies both the Simplex Condition (22) and the AGF Accumulation Limit (23), the latter because by observation [1] there is simply no candidate with total accumulation less than that carried by the optima.

A third observation is also significant:

[3] Any candidate with marks that do not realize a wedge is harmonically bounded, and indeed bounded by a candidate that is the same except for lacking the marks.

It is then reasonable to ask whether BST, or some fragments of it, are AGF. First the good news: consider the candidate set arising from any input that contains no wedges except possibly pre-V ▸ in its annotated form, for example /V/, /CV/, /CVV/, and the like, which annotate as ▸V, CV, and CV▸V, respectively. Three constraints are involved, ONS, MAX, DEP. (Candidates which have NOCODA marks will be simply-bounded harmonically by [3] above, losers to another candidate under any weighting, so we may ignore them.) Any coda-free candidate will split its marks between the three constraints. The remarkable fact is that an *optimum* must have the violation profile (0,0,k): no matter how the constraints are ranked, the lowest-ranked of the three will incur the entire accumulation. From this it immediately follows that such candidate sets are AGF. The profile (0,0,k) has minimal first and second accumulations, and since all candidates accumulate to at least k over ONS, MAX, DEP, its third accumulation is also minimal. Because of symmetry, this argument applies to all rankings of the three constraints.

Now for the more complicated reckoning with C◂. Many candidate sets involving such coda-relevant marks will indeed show the (0,0,k) pattern: consider for example /CVC/ = CVC◂. (The three possible optima are shown as (a),(b),(c); some harmonically-bounded candidates are appended to give a sense of how they fail.)

(27) /CVC/ = CVC◂

| /CVC/ | Ons | NoCoda | Max | Dep |
|---|---|---|---|---|
| a. .CVC. | 0 | 1 | 0 | 0 |
| b. .CV.Cv. | 0 | 0 | 0 | 1 |
| c. .CV.■ | 0 | 0 | 1 | 0 |
| d. ■■■ | 0 | 0 | 3 | 0 |
| e vC.V■ | 1 | 1 | 1 | 1 |

But the behavior of C in the basic theory is intrinsically more complex than the behavior of V. Whereas all instances ▸V are disposed of uniformly in any optimum, according to the (0,0,k) pattern, the fate of the various C◂ can be split. The reason is that C has two structural hosts — onset and coda — whereas V has but one. Consider the following candidates from /CC/ = C◂C◂.

(28) /CC/ = C◂C◂

| C◂C◂ | Ons | NoCoda | Max | Dep |
|---|---|---|---|---|
| a. ■■ | 0 | 0 | 2 | 0 |
| b. .CvC. | 0 | 1 | 0 | 1 |
| c. .Cv.Cv. | 0 | 0 | 0 | 2 |
| d. .Cv.■ = ■.Cv. | 0 | 0 | 1 | 1 |

Candidates (d) are not distinguished in the BST: their shared violation profile is collectively bounded by (a) and (c).

The OT optima from this input are (a),(b),(c). Candidates (a) and (c) follow the (0,0,k) mode; they represent the uniform deletion solution and the uniform epenthesis solution, respectively . The most interesting candidate is (b), which successfully splits its necessary two marks between DEP and NOCODA, resolving each C◂ differently. This candidate would be collectively bounded, if there were a candidate with just 2 NoCoda violations and no others. But no such candidate exists.

The split pattern (b) leads directly to a ranking that fails to meet the requirements of the AG theorem (11).

(29) CC → Ø DEP>>MAX>>NOCODA

| C◂C◂ | DEP | MAX | NOCODA | ONS |
|---|---|---|---|---|
| a. ☞ ■■ | 0 | 2 | 0 | 0 |
| b. .CvC. | 1 | 0 | 1 | 0 |
| c. .Cv.Cv. | 2 | 0 | 0 | 0 |

Candidate (a) is clearly the OT winner by virtue of its performance on DEP. But it is overtaken by (b) in the second accumulation. In essence, we are able at this point in the hierarchy to trade two C-deletions for one V-insertion, dashing hopes of AG behavior.

## References


ROA = Rutgers Optimality Archive, http://roa.rutgers.edu

RuCCS-TR = Technical Reports of the Rutgers Center for Cognitive Science, http://ruccs.rutgers.edu/publicationsreports.html

Legendre, G., Y. Miyata, & P. Smolensky.1990a. Can connectionism contribute to syntax? Harmonic Grammar, with an application. *Proceedings of the 26th Meeting of the Chicago Linguistic Society*. Chicago, IL.

Legendre, G., Y. Miyata, & P. Smolensky.1990b. Harmonic Grammar – A formal multi-level connectionist theory of linguistic well-formedness: An application. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, Cambridge, MA. p. 884-891.

Legendre, G., Y. Miyata, & P. Smolensky. 1990c. Harmonic Grammar – A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, Cambridge, MA. p. 388-395.

McCarthy, John and Alan Prince. 1995. Faithfulness and Reduplicative Identity. In University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory, ed. Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk, 249-384. Amherst: GLSA. ROA-60.

Prince, Alan. 1997 *et seq*. Lectures on Paninian and AntiPaninian Relations. LSA Summer Institute (1997), U. Mass, Amherst (1997), Marburg (1998), PLC (2000), WCCFL (2001). http://ling.rutgers.edu/people/faculty/prince.html.

Prince, Alan. 2002. Entailed Ranking Arguments. ROA-500.

Prince, Alan & Vieri Samek-Lodovici. In Prep. Of Perfection.

Prince, Alan & Paul Smolensky 1993/2002. *Optimality Theory: Constraint Interaction in Generative Grammar*. RuCCS TR-2, CU-CS-696-93. ROA-537.

Samek-Lodovici, Vieri & Alan Prince. 1999. Optima. ROA-363.

Samek-Lodovici, Vieri & Alan Prince. 2002. Fundamental Properties of Harmonic Bounding. RuCCS TR-71.

Smolensky, Paul, and Géraldine Legendre. To Appear. *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*. Cambridge, MA: Blackwell.