

## The Power of Weighted Constraints

Joe Pater, University of Massachusetts, Amherst

Prince and Smolensky (1993/2004) adopt constraint ranking because of concerns about the power of weighted constraints (cf. Harmonic Grammar; Legendre *et al.* 1990). However, there has been little explicit comparison of the sets of linguistic patterns the two theories generate (cf. Prince 2002, Legendre *et al.* 2006a). The conclusion here is that HG is a viable theory of phonological typology if it incorporates proposed amendments to OT that make it properly local. I also discuss some of the complexities of doing linguistic analysis in HG, showing how it is aided by an implementation with Linear Programming, Potts *et al.*'s (2006) *HaLP*.

### 1. Pairwise Ranking vs. Weighting

#### 1.1 Optimization and trade-offs

I adopt a version of weighted constraint evaluation suggested as a strawman by Prince and Smolensky (1993/2004: 236); see also Flemming (2001), Prince (2002), Keller (2006) and Pater *et al.* (2006):

- (1) *Candidate evaluation by weighted constraints*
  - a. Multiply violation scores by weights
  - b. Sum weighted violation scores
  - c. Candidate is optimal *iff* score is lower than every other candidate

#### (2) A weighted constraint tableau

Weight	2	1	Σ
Input	Constraint 1	Constraint 2	
☞ Output 1		*	1
Output 2	*		2

One might think that a weighted constraint system could generate the following pattern by weighting NoCODA appropriately w.r.t. MAX (cf. Prince and Smolensky 1997, Prince 2002):

- (3) A word can have at most one coda; a second coda is deleted

That this is impossible can be shown by considering the weighting conditions under which each outcome is optimal.

- (4)

/batan/	MAX	NoCODA
☞ ba.tan		1
ba.ta	1	

Weighting condition:  $w(\text{MAX}) > w(\text{NoCODA})$

- (5)

/bantān/	MAX	NoCODA
ban.tān		2
☞ bā.tān	1	1

Weighting condition:  $2w(\text{NoCODA}) > 1w(\text{NoCODA}) + 1w(\text{MAX})$   
 $=w(\text{NoCODA}) > w(\text{MAX})$

Logical entailment with strict inequality:

- (6) If  $A > B$  then not  $B > A$

Morals:

- (7) Optimization  $\neq$  Numerical cut-off  
 As in OT, optimality is relative to the input; there is no numerical upper bound on well-formedness
- (8) Nature of trade-offs  
 Here violations of NoCODA and MAX trade off one-to-one; to get a difference between OT and HG, we need *asymmetric trade-offs* in constraint violations between candidates (Prince 2002)

An example of an asymmetric trade-off:

#### (9) Weighting $\neq$ Ranking

Weight	1.5	1	Σ
Input	Constraint 1	Constraint 2	
Output 1		2	2
☞ Output 2	1		1.5

A non-entailment with strict inequalities:

- (10) From  $A > B$  it does not follow that  $A > n*B$

To find differences between OT and HG, we can look for examples of the violation pattern in (9) - cases where one violation of one constraint trades-off against multiple violations of another

One can also get asymmetric trade-offs involving more than two constraints; we'll turn to those in section 2.

## 1.2 Gradient Alignment

Legendre *et al.* (2006a) show how the interaction of the following constraints yields distinct results in HG and OT

- (11) ALIGN-HEAD-R  
Assess one violation mark for every syllable intervening between the main stress and the right edge of the word
- WEIGHT-TO-STRESS  
Assess one violation mark for every unstressed heavy syllable

The following tableaux show how a weighting creates a system that places stress on a heavy syllable if it is at most 3 syllables away from the right edge, and on the final syllable if the heavy syllable is further away

### (12) A problem for HG

Weight:	3.5	1	Σ
/bantanama/	W-S	ALIGN-HD-R	
ban.ta.na.má	1		3.5
☞ bán.ta.na.ma		3	3

Weight:	3.5	1	Σ
/bantanavama/	W-S	ALIGN-HD-R	
☞ ban.ta.na.va.má	1		3.5
bán.ta.na.va.ma		4	4

Further weightings will create an infinite number of further systems, with the upper limit bounded only by word length. Stress systems are well-studied typologically, and while many examples of three-syllable windows have been found, not a single example of a stress window of any larger size exists.

Instead of abandoning weighting, another way of avoiding this result is to abandon gradient alignment (Kager 2001, McCarthy 2003a). McCarthy (2003a) derives the absence of this constraint type from two principles:

- (13) a. *Categorality*: Constraints assign a single mark for each locus of violation
- b. *Locality*: Constraints are limited to mentioning two constituents: the locus of violation, and its context (McCarthy 2003a: 80, based on Smolensky p.c.)

I adopt a version of (13b), which is sufficient to rule out gradient alignment, while allowing other types of gradient constraints, which are compatible with HG

Constraint schema enforcing locality (something like this is implicit in most research in autosegmental and metrical phonology)

- (14) \*XY  
Assign a violation mark to every minimal string in containing XY, where X and Y are adjacent elements on some phonological tier

X and Y may contain information about projections onto higher tiers (perhaps only immediately adjacent tiers, as in 19; cf. 15 and text below). They can also be indexed for position in a domain (morphological and/or prosodic)

Note that in HG, a double-sided context (e.g. intervocalic spirantization) can be analyzed as the result of two single-sided constraints (see Flemming 2001), thus potentially eliminating one counterexample to (13b) and (14); ultimately it may be necessary to allow constraints to have a focus and right and left contexts.

I adopt a grid to represent prominence in the metrical representation (Lieberman 1975, see esp. Hayes 1995 on prominence). A ternary weight distinction, as in Kelkar's Hindi (Kelkar 1968; Prince and Smolensky 1993/2004):

### (15) Metrical representation of prominence distinctions

		x
	x	x
x	x	x
CV	CV:	CV:C

Since languages differ in terms of the prominence distinctions they make (syllable weight, vowel sonority), there are language-specific constraints relating these factors to grid projections. For present purposes, I will take the representations in (15) as given.

Stress is directly marked on the prominence-based grid with an 'o'; grid elements with different phonetic interpretations co-exist on a single plane (cf. Hayes 1995). The rightmost heavy pattern favors (16a) over (16b):

(16)	o	o	
	x	x	
	x	x	
	x	x	
a.	ban.tán.na.ma	b.	ban.tán.na.ma

'End-rule' constraint (adapting Prince 1983, see also Hayes and Wilson 2006; cf. Bakovic 1998, McCarthy 2003a for a prosodic alternative):

- (17) END-RIGHT  $x]_{FINAL}$

Assign a violation mark to a grid element that is final on its tier and that is not dominated by 'o'

To distinguish a language that stresses the rightmost heavy, and one that stresses the rightmost syllable, END-RIGHT is parameterized to grid level:

- (18) END-BASE-R  
 ...grid element of the lowest level that is final in its tier...  
 END-PEAK-R  
 ...grid element of the highest level that is final in its tier...

With the further parameterization to yield 'left' versions of the constraints in (18), we have a constraint set that would get both 'default-to-same' and 'default-to-opposite' stress systems (cf. Gordon 2000)

*Problem:* how do we get the apparently gradient effect displayed when END-PEAK-RIGHT is dominated by NONFINALITY, as in Prince and Smolensky's (1993/2004) analysis of Hindi (see also Walker 1997)? That is, if final stress is not an option, what prefers (19a) to (19b), since both violate END-PEAK-R?

- (19)
- |   |   |   |  |   |   |   |
|---|---|---|--|---|---|---|
|   | o |   |  | o |   |   |
| x | x | x |  | x | x | x |
| x | x | x |  | x | x | x |
- a. dan.bán.tan b. dán.ban.tan

One possibility: final syllables project differently from non-final syllables, perhaps due to some type of final consonant 'extrametricality' (note that final length is perceived as less prominent due to compensation for final lengthening; Nooteboom and Doodeman 1980; see Lunden 2006 for related discussion):

- (20)
- |   |   |   |  |   |   |   |
|---|---|---|--|---|---|---|
|   | o |   |  | o |   |   |
| x | x |   |  | x | x |   |
| x | x | x |  | x | x | x |
- a. dan.bán.tan b. dán.ban.tan

Another possibility is that the following constraint bans a final lapse at the highest grid level:

- (21) \*xx]<sub>FINAL</sub>  
 Assign a violation mark to an adjacent pair of grid marks at the highest grid level in which the second is the final element on the tier and neither is dominated by o

While further work is needed, it appears that a theory in which constraints obey the locality restriction in (14) is sufficiently expressive to provide a reanalysis of systems that have been previously analyzed using head alignment (and some that challenge it, if the 'default-to-opposite' analysis goes through)

For present purposes, the important point is both HG and OT generate just the two desired patterns with END-BASE-R and END-PEAK-R

- (22) *Unattested set of optima*

/bantinama/	END-BASE-R	END-PEAK-R
ban.ta.na.má		1
☞ bán.ta.na.ma	1	

/bantinama/	END-BASE-R	END-PEAK-R
☞ ban.ta.na.va.má		1
bán.ta.na.va.ma	1	

These obviously require contradictory ranking conditions:

- (23) ☞ bán.ta.na.ma      w(END-PEAK-R) > w(END-BASE-R)  
 ☞ ban.ta.na.va.má      w(END-BASE-R) > w(END-PEAK-R)

Legendre *et al.*'s (2006a) alignment example is to my knowledge the only published case of an unattested language generated by HG but not OT.

We now turn to two further cases of pairwise asymmetric trade-off. One favors the greater power of HG, the other is problematic for both HG and OT.

### 1.3 Scalar Constraints

Prince and Smolensky's (1993/2004) analysis of Berber syllabification invokes H-Nuc 'a higher sonority nucleus is more harmonic than one of lower sonority'. McCarthy (2003a: 82) translates it into now-standard OT formalism as follows:

- (24) H-NUC  
 Assign a nucleus one violation-mark for each degree of sonority lower than that of *a*

Unlike Alignment, McCarthy's argument against scalar gradience (building on Prince and Smolensky's 1993/2004 reanalysis of Berber in terms of a fixed ranking) is not that it overgenerates, but that it *undergenerates* in OT.

Consider a language like English, in which nuclei allow sonorant consonants, but no obstruents.

Abstracting from the English stress facts, we can use simple three-degree scale:

- (25) *Sonority scale*  
 Vowel = 2    Sonorant Consonant = 1    Obstruent Consonant = 0

The tableaux in (26) show the effects of the two rankings of H-NUC and the faithfulness constraint violated by epenthesis, DEP (McCarthy and Prince 1999).

(26) *The failure of scalar constraints in OT*

a. H-NUC >> DEP

/ta/	H-NUC	DEP	/t/	H-NUC	DEP	/ts/	H-NUC	DEP
☞ta			tl	1		ts	2	
ta.ə		1	☞təl		1	☞təs		1

b. DEP >> H-NUC

/ta/	DEP	H-NUC	/t/	DEP	H-NUC	/ts/	DEP	H-NUC
☞ta			☞tl		1	☞ts		2
ta.ə	1		təl	1		təs	1	

OT generates only a language that permits all degrees of violation of the constraint (e.g. Berber), and one that tolerates no violation (e.g. French), but not a language that permits an intermediate degree of violation (e.g. English)

In HG, however, an appropriate weighting of H-NUC and DEP will yield an inventory of intermediate size:

(27) *An intermediate inventory in HG (impossible in OT)*

	3	2	Σ		3	2	Σ
/t/	DEP	H-NUC		/ts/	DEP	H-NUC	
☞tl		1	2	ts		2	4
təl	1		3	☞təs	1		3

HG does not generate the 'gapped' inventory:

(28) *An impossible inventory in HG (and OT)*

/t/	DEP	H-NUC	/ts/	DEP	H-NUC
tl		1	☞ts		2
☞təl	1		təs	1	

- (29) ☞təl     w(H-NUC) > DEP  
 ☞ts        w(DEP) > 2w(H-NUC)

The differences between the predictions of OT fixed rankings derived from scales (or constraints in a stringency relation) and HG scalar constraints are more subtle and interesting, and need further examination. Scalar conflation (de Lacy 2004) seems problematic if HG scalar violations increase linearly, as assumed in the present example (cf. Flemming 2001). See Flemming (2001) for arguments for the use of scalar HG for phonetic trade-offs (cf. Boersma 1998:211).

1.4 *Multiple violations of a single categorical constraint*

As the NoCODA and MAX example in (4) and (5) showed, violations of pairs of categorical constraints normally trade-off one-to-one. However, McCarthy (2006a,b) draws attention to the fact that ONSET can demand an unlimited number of LINEARITY violations

(30) LINEARITY

Assign a violation mark for each precedence relation in the input that is reversed in the output

As shown in (30), with the ranking ONSET >> LINEARITY, an input segment at the right edge of the word will appear at the left edge in the output in order to satisfy ONSET, regardless of the number of LINEARITY violations this incurs. The optimal form (31b) incurs five violations of Linearity because [k] follows all of the other segments in the input, and precedes them in the output:

(31) *Non-local metathesis*

anatak	ONSET	LINEARITY
a. anatak	1	
b. kanata		5

As McCarthy notes, this pattern of long-distance metathesis is unattested.

At first glance it might seem that this is a problem with strict domination, but it is in fact aggravated in HG, since varying the weights of ONSET and LINEARITY can produce an infinite set of languages fitting the following description:

- (32) To satisfy ONSET, a segment is permitted to reverse precedence relations with *n* segments, where *n* is any non-negative integer.

McCarthy (2006a,b) shows that the adoption of harmonic serialism solves this problem; this holds for both OT and HG:

(33) *Gen/Eval interaction in Harmonic Serialism*

- a. Gen performs a single operation at a time
- b. Each operation must result in harmonic improvement according to Eval

To illustrate the workings of harmonic serialism, I'll use a tableau format that incorporates the derivational process.

(34) *Tableaux conventions for Harmonic Serialism*

a. Representations ordered according to the order of operations with the UR/Input at the top, followed by a fully faithful parse (McCarthy 2006b). The last representation in the SR/Output.

b. Each representation differs from the immediately preceding one in terms of one operation (one 'basic' faithfulness violation in McCarthy 2006b,c)

c. Each representation is more harmonic than the immediately preceding one

d. A representation that violates b. or c. is indicated with a bomb.

The attested pattern:

(35) *Local metathesis in standard OT and harmonic serialism*

atka	ONSET	LINEARITY
a. atka	1	
☛ b. taka		1

Ruling out the unattested pattern:

(36) *A violation of gradualness in non-local metathesis in harmonic serialism*

anatak	ONSET	LINEARITY
a. anatak	1	
☛ b. kanata		5

(37) *Violations of harmonic improvement*

a. anatak	ONSET	LINEARITY
anatak	1	
☛ anataka	1	1

b. anatak	ONSET	LINEARITY
anatak	1	
☛ naatka	1	1

For expository ease I have presented McCarthy's solution to the Onset/Linearity problem in terms of OT, but the violations of harmonic improvement in (37) would hold equally in Harmonic Grammar, since any pairwise Harmonic Bounding pattern in OT holds in HG (Prince 2002). As we will see, the gradualness requirement has important consequences for typology in HG.

## 2. Independence in Constraint Conflict

Another type of asymmetric trade-off:

(38) *Weighting ≠ Ranking*

Weight	1.5	1	1	Σ
Input	Con1	Con2	Con3	
Output 1		1	1	2
☛ Output 2	1			1.5

Another non-entailment:

(39) From  $A > B$  and  $A > C$ , it does not follow that  $A > B+C$

Shigeto Kawahara has brought the following Japanese loanword phonology example to my attention (Nishimura 2003, Kawahara to appear):

(40) *An attested cumulative constraint interaction*

- Voiced obstruents are parsed faithfully [bagii] 'buggy'
- Voiced geminates are parsed faithfully [reddo] 'red'
- Voiced geminates are devoiced in the context of another voiced obstruent [beddo] 'betto'

(41) *Cumulative effect of OCP-VOICE and \*VCE-GEM*

Weight	1.5	1	1	Σ
/beddo/ 'bed'	Id-VCE	OCP-VOICE	*VCE-GEM	
[beddo]		1	1	2
☛ [betto]	1			1.5

Weight	1.5	1	1	Σ
/bagii/ 'buggy'	Id-VCE	OCP-VOICE	*VCE-GEM	
☛ [bagii]		1		1
[bakii]	1			1.5

<i>Weight</i>	1.5	1	1	Σ
/reddo/ 'red'	Id-VCE	OCP- VOICE	*VCE- GEM	
☞ [reddo]			1	1
[retto]	1			1.5

This pattern can be captured in standard OT by increasing the size of the constraint set (see Nishimura 2003 for a Local Conjunction analysis of this case, and Kawahara to appear for one employing geminate-specific faithfulness). However, given the set of constraints in (41), no ranking will produce this pattern.

Because this asymmetric trade-off pattern involves violations of pair of constraints trading-off against a one of a single competing constraint, there are strict restrictions on the types of gang effect that can be captured.

(42) *A putative gang effect*

A vowel undergoes backness harmony *iff* it is nasal

Constraints at issue:

- (43) \*V-NAS Assign a violation mark to a nasal vowel  
 AGREE-BACK Adjacent vowels agree in [+/-back]  
 IDENT-BACK A vowels' input and output backness features are identical

The desired optima and constraint violations:

(44) *An unattested pattern of vowel harmony*

/idũ/	ID-BACK	*V-NAS	AGR- BACK
☞ [idỹ]	1	1	
[idũ]		1	1

/idu/	ID-BACK	*V-NAS	AGR- BACK
[idy]	1		
☞ [idu]			1

There is clearly no weighting of the constraints that will yield the pattern in (44)

(45) It follows from  $A + B > C + B$  that  $A > C$

The fact that this is impossible in HG highlights the fact that it has very little of the power of OT with Local Conjunction (Legendre *et al.* 2006, Pater 2006)

A problem: if IDENT-BACK were replaced by MAX, the pattern of violation marks instantiates the asymmetric trade-off in (36) and (41).

(46) *An unattested pattern of vowel deletion*

/idũ/	MAX	*V-NAS	AGR- BACK
☞ [id]	1		
[idũ]		1	1

/idu/	MAX	*V-NAS	AGR- BACK
[idy]	1		
☞ [idu]			1

The following builds on McCarthy's (2006c) Harmonic Serialist analysis of Wilson's (2003) C1-deletion generalization. IDENT-NASAL is replaced:

(47) MAX-NASAL An input [nasal] feature is preserved in the output

Violation of principles of harmonic serialism in unattested mapping:

(48) *Violation of gradualness*

/idũ/	MAX	MAX-NASAL	*V-NAS	AGR- BACK
[idũ]			1	1
☞ [id]	1	1		

(49) *Violation of harmonic improvement*

/idũ/	MAX	MAX-NASAL	*V-NAS	AGR- BACK
[idũ]			1	1
☞ [idu]		1		1

The derivational step in (49) must violate harmonic improvement because of the weighting condition imposed by the presence of nasal vowels in the inventory:

(50) /dũ/ → [dũ] \*[du] MAX-NASAL > \*V-NAS

Consequence of independence in constraint conflict (especially true in McCarthy's 2006b,c version of Harmonic Serialism):

(51) Where markedness constraints conflict with different (basic) faithfulness constraints, a gang effect cannot be produced in HG

### 3. Linear Programming and HG Analysis

We've thus far been looking at extremely simple patterns of constraint violation, involving at most three tableaux with just a few constraints and a couple of candidates at a time, each usually incurring a single violation. As we increase the number of each of these, it quickly becomes difficult to answer the question we've been posing:

(52) What weighting, if any, makes a particular set of candidates optimal?

Pater, Potts and Bhatt (2006) present an application of Linear Programming that can answer this question for larger problems; a web-based implementation is provided by Potts, Pater, Bhatt and Becker (2006).

The mathematician George Dantzig discovered the simplex algorithm in 1947, while working for the Air Force. He called it Linear Programming, where programming has an old military sense: troop movements, supply distribution, etc.

(53) The tremendous power of the simplex method is a constant surprise to me.  
— George Dantzig

Linear Programming optimizes a value, subject to a set of inequalities (which LP calls constraints - I'll avoid that term for obvious reasons). The value being optimized is expressed in an *objective function*.

In our application of the simplex algorithm, the objective function consists of the minimization of the summed weights of the constraints, and the inequalities are transformed weighting conditions

(54) *Our objective function*

Minimize  $w_1 + \dots + w_n$  (where  $n$  = number of constraints)

We minimize the values because otherwise our systems would all be unbounded, and there would be no optimal solution

(55) *Unboundedness of maximization*

Given a set of weights  $\{w_1 \dots w_n\}$  that satisfies the weighting conditions, these values can all be multiplied by a constant resulting in a sum  $w_1 + \dots w_n$  that is greater than the original one, and still satisfies the implications

Minimization is bounded because linear programming uses inequalities (greater or equal to), rather than strict inequalities, and we impose a minimum value of 1 on all weights ( $w_x \geq 1$ ).

The weighting conditions are transformed as follows:

(56) *Transformation of weighting condition to inequality*

$$w_1 > w_2 \rightarrow w_1 - w_2 > 0 \rightarrow w_1 - w_2 \geq 1$$

In practice, it is possible to construct an inequality directly from the violation pattern of an optimal form and its paired sub-optimal candidate:

(57) *Transformation of violation pattern to an inequality*

Input	C1	C2
☞ Output 1		1
Output 2	1	

$$\rightarrow w(C1) - w(C2) \geq 1$$

To answer the question in (52), we do the following:

- (58) i. Transform the violation patterns as described above for every optimum-sub-optimum (winner-loser) pair in each tableau to create a set of inequalities
- ii. For every constraint, impose the further inequality that  $w_x \geq 1$
- iii. Run the simplex to find the optimal value for the objective function (54)

Since we avoid unboundedness by minimizing, the simplex is guaranteed to provide one of two outputs:

- (59) i. A set of values for the objective function, which when used as weights for our constraints, gives the optimum the lowest score in its tableau
- ii. A verdict of *infeasible*

The infeasibility verdict indicates that there is a contradiction in the weighting conditions, that is, that *no weighting will pick that set of optima*.

### 4. Case Study: A Failed Analysis

Here I document a failed attempt to analyze what McCarthy (2003b) calls a Grandfather Effect by using HG to mimic the Local Conjunction analysis of Bakovic (2000) and Smolensky (2006). This highlights the usefulness of Linear Programming to HG analysis, but also some of its limits, as well as some of the complexities and limits of HG analysis itself.

#### 4.1 The attempted analysis

The data pattern is from Assamese ATR harmony (Mahanta 2007):

- (60) a. [+ATR] spreads regressively from the high vowels [i] and [u], yielding an allophonic alternation in the mid vowels, which appear as [+ATR] [e] and [o] in when they precede [+ATR] vowels, and as [-ATR] [ɛ] and [ɔ] elsewhere

e.g. [bɔsɔr] 'year'      [bosori] 'yearly'

- b. Underlying [-ATR] [ʊ] also neutralizes with [+ATR] [u] when it precedes a [+ATR] vowel.

e.g. [bul] 'daze'      [buluwa] 'mislead'

- c. Harmony is blocked by an intervening consonant cluster (C1 is a coda)

e.g. [gʊs.tʰi] 'clan'      [kɔl.ki] 'last incarnation of Vishnu'

- d. Underlying [+ATR] vowels surface faithfully in closed syllables

e.g. [gʊ.rut.tɔ] 'importance'

Constraints for harmony and mid vowel allophony (based on Mahanta 2007):

- (61) a. \*[-ATR][+ATR]  
Assign a violation mark to the minimal string containing a [-ATR] vowel followed by a [+ATR] vowel
- b. IDENT-ATR  
Assign a violation mark if a segments' input and output correspondents differ in [ATR]
- c. \*[+ATR, -high]  
Assign a violation mark to a segment specified as [+ATR, -high]

My aim was to analyze blocking with an appropriately weighted constraint against [+ATR] vowels in closed syllables (cf. Javanese, Québécois French).

- (62) \*+ATR-C  
Assign a violation mark to a [+ATR] vowel followed by a coda consonant

Grandfather effect weighting conditions (see Pater 2006, Pater *et al.* 2006):

(63) *A consistent set of weighting conditions*

$$a. w(\text{IDENT-ATR}) > w(*\text{+ATR-C}) \quad /guruttɔ/ \rightarrow [gʊ.rut.tɔ], *[gʊ.rut.tɔ] \quad (60d)$$

$$b. w(*[-\text{ATR}][+\text{ATR}]) > w(\text{IDENT-ATR}) \quad /bul+uwa/ \rightarrow [buluwa], *[bul+uwa] \quad (60b)$$

$$c. w(\text{IDENT-ATR}) + w(*\text{+ATR-C}) > w(*[-\text{ATR}][+\text{ATR}]) \quad /gʊs.tʰi/ \rightarrow [gʊs.tʰi], *[gʊs.tʰi] \quad (60c)$$

Weighting found by HaLP (Potts *et al.* 2006):

$$(64) \begin{aligned} w(\text{IDENT-ATR}) &= 3 \\ w(*[-\text{ATR}][+\text{ATR}]) &= 4 \\ w(*\text{+ATR-C}) &= 2 \end{aligned}$$

Allophony weighting conditions:

(65) *A consistent set of weighting conditions*

$$a. w(*[-\text{ATR}][+\text{ATR}]) > w(\text{IDENT-ATR}) + w(*[\text{+ATR}, -\text{high}]) \quad /bɔsɔr+i/ \rightarrow [bosori], *[bɔsɔri] \quad (60a)$$

$$b. w(*[\text{+ATR}, -\text{high}]) > w(\text{IDENT-ATR}) \quad /bo/ \rightarrow [bɔ], [bo] \quad (\text{Richness of Base})$$

Weighting found by HaLP:

$$(66) \begin{aligned} w(*[-\text{ATR}][+\text{ATR}]) &= 4 \\ w(\text{IDENT-ATR}) &= 1 \\ w(*[\text{+ATR}, -\text{high}]) &= 2 \end{aligned}$$

One might expect the patterns in (63) and (65) to be independent.

The answer delivered by HaLP when they are combined:

(67) *This system is infeasible*

Now comes the hard part:

(68) Why is it infeasible? What can I do to fix it?

## 4.2 Analyzing Inconsistency

The problem I started with was even more complicated, including the [a] blocking pattern, and a constraint against [+high, -ATR] vowels. My first step was to turn the weighting conditions into a set of 'A > B' statements, but that didn't help (me) to diagnose the source of inconsistency.

I eventually stumbled onto a strategy that turns out to be used in infeasibility analysis in Linear Programming - I tried to find the smallest infeasible system, that is, the smallest set of inconsistent weighting conditions. The system of weighting conditions in the previous section is close to being that set.

### (69) Weighting conditions repeated and recoded

- a.  $w(\text{IDENT-ATR}) > w(*\text{+ATR-C})$   $A > B$
- b.  $w(*\text{-ATR}[+\text{ATR}]) > w(\text{IDENT-ATR})$   $C > A$
- c.  $w(\text{IDENT-ATR}) + w(*\text{+ATR-C}) > w(*\text{-ATR}[+\text{ATR}])$   $A + B > C$
- d.  $w(*\text{-ATR}[+\text{ATR}]) > w(\text{IDENT-ATR}) + w(*\text{+ATR, -high})$   
 $C > A + D$
- e.  $w(*\text{+ATR, -high}) > w(\text{IDENT-ATR})$   $D > A$

### (70) Transitivity applied (comma used in OT sense)

- a.  $D, C > A > B$
- b.  $A + B > C > A + D$

Contradictions revealed at last!

### (71) Contradictory entailments

- a.  $D > B$
- b.  $B > D$

Conclusion for HG analysis of Grandfather Effect:

(72) Weighting required for grandfather effect can lead to entailments that are inconsistent with the rest of the system

See Legendre *et al.* 2006a for a similar conclusion about a simpler failure of HG to replicate Local Conjunction

Strategies for dealing with complexities of HG analysis (besides implementing LP and OT filtering routines):

- (73) i. Build up problem gradually
- ii. Pull out constraints and winner-loser pairs and iteratively test
- iii. Use OT techniques for inconsistency detection and analysis (see Pater to appear for an extension of Tesar and Smolensky's RCD)

## 5. Conclusions

Restrictions on typology that might seem to stem from strict domination are in fact better understood as the result of the following:

### (74) Sources of 'strict domination'

- i. The structure of an optimization system (no numerical cut-off)
- ii. Constraint locality
- iii. Harmonic Serialism's gradualness
- iv. Independence of constraint conflicts

A framework for further exploration/development:

(75) Local Harmonic Grammar with Harmonic Serialism

Why use HG?

(76) a. Better suited for scalar constraints - no inventory problem

b. Better suited for gradual learning (see Boersma's 1998 hybrid model, see Jäger to appear and Pater *et al.* 2007 on HG models)

c. Better suited for variation (see again Boersma's 1998 hybrid, see Goldwater and Johnson 2003 for a HG model)

d. Better suited for gradient well-formedness (Legendre *et al.* 1990, 2006a; Keller 2006; Hayes and Wilson 2006, Coetzee and Pater in prep.; cf. Boersma and Hayes 2001, Boersma 2004)

e. Better suited for a unified model of phonetics and phonology (Flemming 2001; cf. Boersma 1998)

Why use OT?

(77) As a means of analyzing complex constraint interactions, in cases where the extra power of weighting is not required

Because any set of optima chosen by an OT ranking can be chosen by an HG weighting, an OT analysis is completely compatible with HG.

## References

- Bakovic, Eric. 2000. Harmony, Dominance and Control. Doctoral dissertation, Rutgers University. ROA-360
- Boersma, Paul. 1998. *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. Doctoral dissertation, University of Amsterdam.
- Boersma, Paul. 2004. A stochastic OT account of paralinguistic tasks such as grammaticality and prototypicality judgments. Ms., University of Amsterdam. [ROA-648].
- Boersma, Paul, and Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32: 45-86. [ROA-348].
- Coetzee, Andries, and Joe Pater. In prep. Weighted OCP-Place Constraints in Arabic and Muna.
- De Lacy, Paul. 2004. Markedness conflation in Optimality Theory. *Phonology* 21.2: 145-199.
- Flemming, Edward. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18. 7-44.
- Goldwater, Sharon, and Mark Johnson. 2003. Learning OT Constraint Rankings Using a Maximum Entropy Model. In *Proceedings of the Workshop on Variation within Optimality Theory*, ed. Jennifer Spenader, Anders Eriksson, and Östen Dahl, 111-120. Stockholm University.
- Gordon, Matthew. 2000. Re-examining default-to-opposite stress. *Berkeley Linguistics Society* 26, 101-112.
- Jäger, Gerhard. To appear. In *Maximum entropy models and Stochastic Optimality Theory*, ed. Jane Grimshaw, Joan Maling, Chris Manning, Jane Simpson, and Annie Zaenen. Stanford, CA: CSLI. [ROA-625].
- Kawahara, Shigeto. To appear. A faithfulness ranking projected from a perceptibility scale: The case of [+voice] in Japanese. In *Language*.
- Legendre, Géraldine, Yoshiro Miyata, and Paul Smolensky. 1990. Harmonic Grammar – a formal multi-level connectionist theory of linguistic wellformedness: An application. In *Proceedings of the twelfth annual conference of the Cognitive Science Society*, 884–891. Cambridge, MA: Lawrence Erlbaum.
- Legendre, Géraldine, Antonella Sorace, and Paul Smolensky. 2006a. The Optimality Theory–Harmonic Grammar connection. In Smolensky and Legendre (2006), 903–966.
- Legendre, Géraldine, Yoshiro Miyata, and Paul Smolensky. 2006b. The interaction of syntax and semantics: A Harmonic Grammar account of split intransitivity. In Smolensky and Legendre (2006), 417–452.
- Lunden, Anya. 2006. Weight, final lengthening and stress: A phonetic and phonological case study of Norwegian. Ph. D. Dissertation, University of California, Santa Cruz.
- Mahanta, Shakuntala. 2007. On Consonantal Intervention in Vowel Harmony. Handout from OCP-4, University of the Aegean.
- McCarthy, J. 2003a. OT constraints are categorical. *Phonology* 20, 75-138.
- McCarthy, J. 2003b. Comparative markedness. *Theoretical Linguistics* 29. 1-51.
- McCarthy, John J. 2006a. Restraint of analysis. In *Wondering at the natural fecundity of things: Essays in honor of Alan Prince*. <http://repositories.cdlib.org/lrc/prince/10>.
- McCarthy, John J. 2006b. Hidden Generalizations: Phonological Opacity in Optimality Theory. London: Equinox Publishing.
- McCarthy, John. 2006c. Slouching Toward Optimality: Coda-Reduction in OT-CC. ROA-878.
- Nooteboom, S. G and G. J. N. Doodeman. 1980. Production and perception of vowel length in spoken sentences. *Journal of Acoustical Society of America* 67: 276-286.
- Pater, Joe. 2006. Additive optimization and phonological typology. URL <http://people.umass.edu/pater/AO+typology.pdf>, Handout from a talk presented at the MIT/UMass Phonology Workshop,
- Pater, Joe. 2007. Morpheme-Specific Phonology: Constraint Indexation and Inconsistency Resolution. To appear in S. Parker (ed.) *Phonological Argumentation*. Equinox Press.
- Pater, Joe, Christopher Potts, and Rajesh Bhatt. 2006. Harmonic Grammar with Linear Programming. Ms, University of Massachusetts, Amherst. ROA-872.
- Pater, Joe, Karen Jesney and Anne-Michelle Tessier. 2007. Phonological Acquisition as Weighted Constraint Interaction. In *Proceedings of the Conference on Generative Approaches to Language Acquisition – North America (GALANA 2)*, ed. Alyona Beliokova, Luisa Meroni, and Mari Umeda. Somerville, MA: Cascadilla Proceedings Project.
- Potts, Christopher, Joe Pater, Rajesh Bhatt, and Michael Becker. 2006. HaLP: Harmonic grammar with linear programming, version 1. Software available online at <http://web.linguist.umass.edu/~halp/>.
- Prince, Alan. 2002a. Anything Goes. In *New century of phonology and phonological theory*, ed. Takeru Honma, Masao Okazaki, Toshiyuki Tabata, and Shin ichi Tanaka, 66–90. Tokyo: Kaitakusha. ROA-536.
- Prince, Alan. 2002b. Arguing Optimality. In *University of Massachusetts Occasional Papers in Linguistics: Papers in Optimality Theory II*, ed. Angela Carpenter, Andries Coetzee, Paul de Lacy, 269-304. Amherst, MA: GLSA. [ROA-562].
- Prince, Alan, and Paul Smolensky. 1993/2004. *Optimality Theory: Constraint interaction in generative grammar*. Technical Report, Rutgers University and University of Colorado at Boulder, 1993. Revised version published by Blackwell, 2004. [ROA-537].
- Smolensky, Paul. 2006. Optimality in phonology II: Harmonic completeness, local constraint conjunction, and feature domain markedness. In Paul Smolensky and Geraldine Legendre (eds.) *The harmonic mind: From neural computation to Optimality-Theoretic grammar*. Cambridge, MA: MIT Press. 27–160.
- Smolensky, Paul, and Geraldine Legendre. 2006. *The harmonic mind: From neural computation to Optimality-Theoretic grammar*. Cambridge, MA: MIT Press.
- Tesar, Bruce and Paul Smolensky. 1998. Learnability in Optimality Theory. *Linguistic Inquiry* 29: 229-268.
- Walker, Rachel. 1997. Mongolian stress, licensing, and factorial typology. Ms. USC [ROA].