Harmonic Grammar with Harmonic Serialism*

Joe Pater

UMass Colloq, Nov. 9, 2007

1 Introduction

Prince and Smolensky (1993/2004: 236) consider, but reject, a version of OT with weighted constraints, as in its predecessor Harmonic Grammar (HG: Legendre et al. 1990; Smolensky and Legendre 2006). In this talk, I start by discussing three attractive properties of weighted constraints. They offer:

1. *A retrictive theory of cumulative constraint interaction* (cf. Local Constraint Conjunction)
2. *Gradual, convergent learning algorithms*
3. *Straightforward extensions of the grammar and learning models to variation*

I will illustrate these properties with an analysis of a pattern of cumulative constraint interaction in the phonology of Japanese loanwords (Nishimura 2003, Kawahara 2006).

We will then turn to Prince and Smolensky's concern about weighted constraints. They address this "fear of optimization" (p. 232):

> *Loss of restrictiveness:* "In order to handle optimality, you must use numbers and use counting. . . The result will be a system of complicated trade-offs. . . giving tremendous descriptive flexibility and no hope of principled explanation. Therefore, the main goal of generative grammatical investigation is irredemeemably undermined."

Their "reassurance" (p. 233):

> *Loss of restrictiveness through arithmetic:* Concern is well-founded here. As we have shown, however, recourse to the full-blown power of numerical optimization is not required. . . In Optimality Theory, constraints are ranked, not weighted: harmonic evaluation involves the abstract algebra of order relations rather than numerical adjudication between quantities.

*This research would not have been possible, or as enjoyable, without the contributions of my collaborators: Michael Becker, Rajesh Bhatt, Paul Boersma, Karen Jesney, Chris Potts and Patrick Pratt. It also would not have been possible without discussion with many other members of this department.

Prince and Smolensky offer no examples of unwanted results of numerical optimization: the only one in the literature may be an example that uses gradient alignment (Legendre et al. 2006). I will demonstrate new software that allows us to compare the predictions of OT and HG (OT-Help; Becker et al. 2007), and admit that Prince and Smolensky were right. But, I will start to make a case that they are only right if OT computes well-formedness globally and in parallel, which may well be wrong (Wilson 2003; McCarthy 2006a; 2006b; 2007).

2 The theory

2.1 Harmonic Grammar: A linear model of grammar

The model of grammar I will assume is OT translated into a linear system (Prince and Smolensky 1993/2004: 236; see Smolensky and Legendre 2006 and Pater et al. 2007a for further discussion and references). It is a linear system because the Harmony function uses a linear equation. Because this Harmony function comes from Harmonic Grammar, I refer to this theory as HG, but it is different in many ways from the original 1990 proposal.

(1) *Optimality*

The optimal candidate has greater Harmony than its competitors

Harmony

The harmony of a candidate is the sum of its weighted constraint scores

OT constraints that assign violations assign negative penalties. For example:

(2) *CODA-VOICE: Assign a penalty of -1 to each voiced coda

IDENT-VOICE: Assign a penalty of -1 to each input-output pair of corresponding segments that differ in [voice] specification

The following tableau shows the outcome when *CODA-VOICE has a greater weight than IDENT-VOICE:

(3) *Coda devoicing*

Weight	2	1	$\mathcal{H}$
/bad/	*CODA-VOICE	IDENT-VOICE	
☞ bat		-1	-1
bad	-1		-2

Differences from an OT tableau:

(4) i. The constraint weights are given in the top row, above the constraints to which they apply

ii. The final column shows each candidates' Harmony

The interaction between *Coda-Voice and Ident-Voice is an example of what Prince (2002) calls an *Anything Goes* scenario:

(5) To translate Con-1 $>>$ Con-2 into an equivalent weighting, assign as weights any positive values that satisfy the inequality w(Con-1) $>$ w(Con-2)

The outcome in (3) will obtain with any set of positive weighting values in which w(*Coda-Voice) $>$ w(Ident-Voice). With w(Ident-Voice) $>$ w(*Coda-Voice), voiced codas are allowed:

(6) *Voiced codas allowed*

Weight	2	1	$\mathcal{H}$
/bad/	Ident-Voice	*Coda-Voice	
bat	-1		-2
☞ bad		-1	-1

Prince (2002) does not comment on the fact that when constraints interact in this way, the typologies produced by HG and OT are indistinguishable (see Pater et al. 2007a for discussion). An important point made by Prince (2002), as well as Prince and Smolensky (1993/2004: 236):

(7) Any linguistic pattern that can be analyzed by a constraint ranking can be analyzed by a weighting of the same set of constraints

As in Keller (2000; 2006) and Prince (2002), I will ban negative weights, so that constraints that assign penalties cannot reward candidates that contain the dispreferred structure.

2.2 A gradual learning algorithm for HG

Similarities with OT learners in Tesar (1995 *et seq.*) and Boersma (1997 *et seq.*):

(8) *On-line error driven learner*

The learner is supplied with one correct (input, output) pair at a time.

The learner finds uses the current state of its grammar to find the optimal output for the same input (input, output').

If the two outputs do not match (output $\neq$ output'), the learners' own mapping is labeled an *error*, and the grammar is adjusted based on a comparison to the correct form.

The difference (this is the Perceptron update rule of Rosenblatt 1958; it is also used in other learning algorithms and statistical estimation procedures - see Pater 2007 for references, especially to previous linguistic applications):

(9) *HG update rule*

For each constraint, subtract the (unweighted) score of the error from that of the correct form.

Multiply each of these by a constant n ($> 0, \geq 1$), and add the results to the weights of the constraints.

For example, given the pair (/bad/, [bad]), a learner with the grammar in (3) would make an error. We first get the difference between the correct form and the error:

(10) *Creation of a learning vector*

		IDENT-VOICE	*CODA-VOICE
Correct	/bad/ [bad]		-1
Error	/bad/ [bat]	-1	
Correct - Error		**1**	**-1**

If we set the constant n at 1, the result of adding the values from (10) to (3) is as shown in the following tableau:

(11) *Updated grammar on next voiced coda*

Weight	2	1	$\mathcal{H}$
/bad/	IDENT-VOICE	*CODA-VOICE	
bat	-1		-2
☞ bad		-1	-1

This learner combines the advantages of the OT learners (see Boersma and Pater 2007 on (12a), and Jesney and Tessier 2007 on (12b)):

(12) a. Like Tesar's Constraint Demotion Algorithm (CDA), it is guaranteed to find a correct grammar for a set of data if one exists (and if full information about structure is provided); this is not true of Boersma's Gradual Learning Algorithm for stochastic OT (GLA; see Pater 2007)

b. Like the GLA, this learner can learn gradually, thus making it suitable to model the course of human language acquisition (see Jesney and Tessier on advantages over the stochastic OT/GLA combination); this is not true of the CDA (cf. Tessier 2006).

And with simple elaborations, it may have an advantage over both the CDA and the GLA (Jesney and Tessier 2007, Jesney tomorrow, Magri tomorrow):

(13) With a bias for low faithfulness, this learner appears to do better on subset problems than the CDA or GLA, even with equivalent, or more elaborate biases in those other learners

2.3 Noisy Harmonic Grammar

The linear model can be turned into a stochastic model of grammar by adding evaluation noise (an adaptation of stochastic OT; Boersma 1998; Boersma and Hayes 2001):

(14) i. Each time the grammar evaluates a candidate set, weighting values are perturbed by noise (sampled from a normal distribution with $SD = 2$)

ii. In repeated evaluations, this can produce variation in the choice of optima

If *Coda-Voice and Ident-Voice are given equal weight, the result will be approximately 50% devoicing:

(15) *Variable coda devoicing in Noisy HG*

Mean Weight		2	2	$\mathcal{H}$
	/bad/	*Coda-Voice	Ident-Voice	
50%	bat		-1	μ = -2
50%	bad	-1		μ = -2

The surprising finding of Boersma and Pater (2007) is that with evaluation noise in the grammar model, the learning algorithm in 2.2 continues to find correct weightings for categorical systems.

There are other approaches to stochastic HG, and other learning algorithms: Johnson (2002); Goldwater and Johnson (2003); Jäger (2006); Wilson (2006) propose log-linear models of grammar with associated learning algorithms (see Jesney 2007 for some comparisons with noisy HG), and Soderstrom et al. (2006) propose a connectionist learning implementation.

3 An application

To see how the HG model of grammar differs from that of OT, we can consider a slightly more complicated example in the phonology of voicing.[1] In Japanese loanwords (but not native words), pairs of voiced obstruents are permitted (Nishimura2003, Kawahara 2006; all data are from the latter source):

(16) Violations of Lyman's Law in loanwords

bagii 'buggy'	bagu 'bug'
bogii 'bogey'	dagu 'Doug'
bobu 'Bob'	giga 'giga'

Japanese also has a restriction against voicing in geminate (doubled) obstruents. But again, in loanwords, these occur:

(17) Voiced/voiceless obstruent geminate near-minimal pairs in Japanese loanwords

webbu 'web'	wippu 'whipped (cream)'
sunobbu 'snob'	sutoppu 'stop'
habburu 'Hubble'	kappuru 'couple'
kiddo 'kid'	kitto 'kit'
reddo 'red'	autoretto 'outlet'
heddo 'head'	metto 'helmet'

However, when a word contains both a voiced geminate and a voiced obstruent, the geminate is optionally, but categorically, devoiced:

[1]Thanks to Shigeto Kawahara for pointing out that this case could be dealt with in terms of HG.

(18) Optional devoicing of a geminate in Lyman's Law environment

guddo ~ gutto 'good'	doggu ~ dokku 'dog'
beddo ~ betto 'bed'	baggu ~ bakku 'bag'
doreddo ~ doretto 'dredlocks'	budda ~ butta 'Buddha'
baddo ~ batto 'bad'	doraggu ~ dorakku 'drug'
deibiddo ~ deibitto 'David'	biggu ~ bikku 'big'

According to Nishimura (2003) and Kawahara (2006), such devoicing is judged unacceptable in both (16) and (17).

The voicing pattern in Japanese loanwords can be analyzed as the cumulative effect of a constraint against multiple voiced obstruents (*2-Voice), and one against voiced geminates (*Vce-Gem):

(19) *Japanese loanword devoicing as cumulative constraint interaction*

	Weight	1.5	1	$\mathcal{H}$
	/bobu/	Ident-Voice	*2-Voice	
☞	[bobu]		-1	-1
	[bopu]	-1		-1.5

	Weight	1.5	1	$\mathcal{H}$
	/webbu/	Ident-Voice	*Vce-Gem	
☞	[webbu]		-1	-1
	[weppu]	-1		-1.5

	Weight	1.5	1	1	$\mathcal{H}$
	/doggu/	Ident-Voice	*Vce-Gem	*2-Voice	
	[doggu]		-1	-1	-2
☞	[dokku]	-1			-1.5

Due to the strict domination property of ranked constraints, which rules out cumulativity, or gang effects, no OT ranking of these constraints can generate this pattern (see Nishimura 2003 and Kawahara 2006 for expansions of the OT constraint set that deal with these data, and section 4 for a comparison with Nishimura's Local Conjunction approach).

Kawahara (2006) empasizes that the devoicing pattern is emergent: it has entered Japanese in recent borrowings, and there would have been no evidence in the learners' input for a difference between the forms that show devoicing, and those that do not. This pattern is in fact a predicted stage of gradual learning in HG.

We start with constraint values that are appropriate for the pattern of voicing in the native vocabulary, in which there is no voiced geminates, and only a single voiced obstruent in each word:

(20) *Voicing in native Japanese phonology*

	Mean weight	100	50	$\mathcal{H}$
	/bobu/	*2-Voice	Ident-Voice	
	[bobu]	-1		-100 (μ)
☞	[bopu]		-1	-50 (μ)

	Mean weight	100	50	$\mathcal{H}$
	/webbu/	*Vce-Gem	Ident-Voice	
	[webbu]	-1		-100 (μ)
☞	[weppu]		-1	-50 (μ)

Praat (Boersma and Weenink 2007) implements the grammar and learning model discussed above. The details of this simulation:

(21) i. Learning data: mappings /bobu/ →[bobu], /webbu/ →[webbu] and /doggu/ →[doggu] with equal frequency

ii. The initial weightings: as in (20)

iii. Learning rate (Praat's plasticity, the constant n in 2.2): 0.1.

After 1100 pieces of learning data, the weightings were as shown in (22):

(22) *Emergent variable Japanese loanword devoicing*

	Mean weight	107.72	53.65	$\mathcal{H}$
	/bobu/	Ident-Voice	*2-Voice	
100%	[bobu]		-1	-53.65 (μ)
	[bopu]	-1		-107.72 (μ)

	Mean weight	107.72	52.50	$\mathcal{H}$
	/webbu/	Ident-Voice	*Vce-Gem	
100%	[webbu]		-1	-52.50 (μ)
	[weppu]	-1		-107.72 (μ)

	Mean weight	107.72	52.50	53.65	$\mathcal{H}$
	/doggu/	Ident-Voice	*Vce-Gem	*2-Voice	
67.5%	[doggu]		-1	-1	-106.15 (μ)
32.5%	[dokku]	-1			-107.72 (μ)

This simulation is a gross abstraction from the actual process of loanword adaptation, and does not explain why this grammar seems to be stable in current Japanese. However, the fact that this pattern emerges as a stage in learning should make the necessary further elaboration tractable.

4 Harmonic Grammar versus Local Conjunction

Smolensky (2006) proposes a version of OT that includes a mechanism of local constraint conjunction (henceforth OT-LC), which allows OT to capture cumulative interactions between constraints. Conjunction takes two independent constraints and creates a new one that is violated *iff* both conjuncts are violated.

Japanese loanword devoicing as local conjunction (Nishimura 2003):

(23)

	/bobu/	IDENT-VOICE	*2-VOICE
☞	[bobu]		*
	[bopu]	*	

	/webbu/	IDENT-VOICE	*VCE-GEM
☞	[webbu]		*
	[weppu]	*	

	/doggu/	*VCE-GEM&*2-VOICE	IDENT-VOICE	*VCE-GEM	*2-VOICE
	[doggu]	*		*	*
☞	[dokku]		*		

The main issue with OT-LC is that it predicts a number of cumulative interactions that seem highly unlikely as sound patterns of natural languages (see esp. McCarthy 1999, 2003a). For example, as well as coda devoicing, many languages have processes in which consonants assimilate in place of articulation. No known language has coda devoicing only when the adjacent segment disagrees in place of articulation. OT-LC predicts that this pattern should occur.

(24)

	/bad/	IDENT-VOICE	*CODA-VOICE
	bat	*	
☞	bad		*

	/sat+ma/	IDENT-PLACE	AGREE-PLACE
☞	satma		*
	sapma	*	

	/bad+ma/	*CODA-VOICE&AGREE-PLACE	IDENT-VOICE	*CODA-VOICE
☞	batma		*	
	badma	*		*

This is a problem of *co-relevance* (McCarthy 2003a): two constraints that regulate independent dimensions of well-formedness interact in an implausible way.

In HG, however, this cumulative interaction is correctly predicted to be impossible. The following shows the violation marks of the unconjoined constraints, converted into negative integers, along with the optima for the pattern generated by OT-LC:

(25)

/bad/	Ident-Voice	*Coda-Voice
bat	-1	
☞ bad		-1

/sat+ma/	Ident-Place	Agree-Place
☞ satma		-1
sapma	-1	

/bad+ma/	Ident-Place	Ident-Voice	*Coda-Voice	Agree-Place
☞ batma		-1		-1
badma			-1	-1

For the optima to beat their competitors, their scores must be higher than that of their competitors. From that we can infer the following required inequalities in the weight values of the constraints:

(26) a. /bad/ → [bad]: w(Ident-Voice) > w(*Coda-Voice)

b. /bad+ma/→ [batma]: w(*Coda-Voice) + w(*Agree-Place) > w(Ident-Voice) + w(*Agree-Place)

After removing the term that appears on both sides of the inequality in (26b.), we can see that the two inequalities are inconsistent:

(27) a. /bad/ → [bad]: w(Ident-Voice) > w(*Coda-Voice)

b. /bad+ma/→ [batma]: w(*Coda-Voice) > w(Ident-Voice)

The shared Agree-Place violation in /bad+ma/→ [batma] and /bad+ma/→ [badma] is irrelevant in determining whether devoicing occurs:

(28) Co-relevance is entailed by the nature of cumulative interaction in HG

In the Japanese example above, a single Ident-Voice violation results in the satisfaction of both *2-Vce and *Vce-Gem. Because of this, the weighting conditions are not inconsistent:

(29) a. /bobu/ → [bobu]: w(Ident-Voice) > w(*2-Voice)

b. /webbu/→ [weppu]: w(Ident-Voice) > w(*Gem-Voice)

c. /doggu/ → [dokku]: w(*Gem-Voice) + w(*2-Voice) > w(Ident-Voice)

This comparison between HG and OT-LC shows that it is possible to reason about the predictions of weighted constraints (see also Legendre et al. 2006; Pater 2006; Pater et al. 2007a).

5 Computationally testing the typological predictions of HG

For the simple cases we have been discussing, it is relatively easy to infer the weighting conditions from the violation profiles of the candidates, and to find a set of weights that fits the inequalities, or to show that the inequalities are inconsistent.

For larger sets of inputs, output candidates, and constraints, hand-solving of HG problems quickly becomes very difficult. Furthermore, the task of finding the set of languages generated by a set of constraints might seem to be impossible, since even a single constraint has infinitely many weighting values.

While the number of rankings of a set of OT constraints is finite, it is still extremely large. With large constraint sets, calculating the predicted typology requires software help:

(30) OT-Soft (Hayes et al., 2003)

The HG solution:

(31) HaLP (Potts et al., 2007)

OT-Help (Becker et al., 2007)

These programs use the Linear Programming Simplex Algorithm to find a set of weights for a given set of optima, or to find out if no such weighting exists. For details on the translation from HG problems to LP systems see Pater et al. (2007b). To calculate the HG and OT typologies, OT-Help uses the same strategy as OT-Soft: tableaux are added one at a time and the sets of jointly feasible optima are gradually built up.

Because OT-Help also uses the same input file format as OT-Soft, we can easily compare the results for OT and HG, using files that have been constructed for OT typological investigations.

The HG and OT figures in (32) were obtained by submitting OT-Soft files prepared by René Kager to OT-Help. Positive HG refers to a version of HG in which the constraint weights are limited to positive reals, as we have been assuming.

(32) *Number of predicted languages with Kager's (2006) gradient Alignment constraint set*

All logically possible combinations of optima (words 2-9 syllables in length, candidates are all footings with QI binary trochees): 685, 292, 000

Optimality Theory: 35

Positive Harmonic Grammar: 911

This typological explosion is expected - gradient Alignment constraints lead to especially unwelcome results in HG (Legendre et al. 2006; Pater et al. 2007b). A somewhat less dramatic increase also occurs with Kager's Lapse constraints:

(33) *Number of predicted languages with Kager's (2006) non-gradient Alignment + Lapse constraints*

All logically possible combinations of optima: 685, 292, 000

Optimality Theory: 25

Positive Harmonic Grammar: 85

But this is still a three-fold increase in the number of predicted languages.

A further problem:

(34) Pater et al. (2007a) identify several cases in which the interaction of categorical, local constraints on segmental and syllable structure produce an infinite number of languages in HG.

Infinity may or may not be a worry, but the languages in the space we can examine are bizarre:

(35) A string of *n* segments assimilate to a "trigger", but a string of *n+1* does not

A segment will metathesize with *n* segments to provide an onset, but will stay in place if the onset position is *n+1* segments away

Pater et al's (2007a) diagnosis of the source of these problems:

(36) Due to standard OT's global evaluation, not to numerically weighted constraints

OT is unique amongst generative constraint-based theories of phonology in that the entire representation is evaluated once and only once by the entire constraint set. As Wilson (2003) and McCarthy (2006b *et seq.*) show, this globality can lead to unwelcome results, even with ranked constraints.

6 Some consequences of Harmonic Serialism for HG

Here I will briefly show that a theory in which representations are evaluated incrementally and locally has desirable consequences not only for OT, but also for HG (see Pater et al. 2007a for other examples). The local, incremental theory:

(37) Harmonic Serialism (HS; Prince and Smolensky 1993/2004: ch. 2, McCarthy 2000; 2006b *et seq.*)

Fundamental principles of HS:

(38) *Gradualness* Gen is limited to a single operation at a time (AKA serialism or incrementalism)

Harmonic improvement Each operation must increase well-formedness as defined by the constraint set[2]

There are various ways that a theory of HS could be elaborated: McCarthy's (2007) OT with Candidate Chains (OT-CC) is the best-developed version (though see also Wilson 2003, which moves further from standard OT assumptions, as well as the early precedent in Goldsmith 1993).

[2]More precisely, a step in a derivation *S* involves the choice of the most harmonic member of a candidate set that includes the representation at step *S-1*, and all candidates that result from applying a single operation. "Single operation" could be defined in several ways: one instance of any operation (deletion, insertion, etc.), one instance of some defined operation, (e.g. one deletion) or any number of instances of a single operation (e.g. any number of deletions). We will only be considering a single instance of a single operation.

6.1 Serial HG in segmental phonology

To see how HS works, and its consequences for HG, we can start with a problem you may have noticed. If we replace IDENT-VOICE and IDENT-PLACE with a single MAX constraint, we get an unattested cumulative interaction between *CODA-VOICE and AGREE-PLACE, as shown in (39).

(39) *An unattested cumulative interaction produced by HG*

Weight	1.5	1	$\mathcal{H}$
/bad/	MAX	*CODA-VOICE	
ba	-1		-1.5
☞ bad		-1	-1

Weight	1.5	1	$\mathcal{H}$
/sat+ma/	MAX	AGREE-PLACE	
sama	-1		-1.5
☞ satma		-1	-1

Weight	1.5	1	1	$\mathcal{H}$
/bad+ma/	MAX	*CODA-VOICE	AGREE-PLACE	
☞ bama	-1			-1.5
badma		-1	-1	-2

This is in fact a special case of a broader problem:

(40) Why are AGREE-PLACE and *CODA-VOICE always repaired by featural change, rather than by segmental deletion?

McCarthy's (2006a) Harmonic Serialist account of this sort of gap:

(41) Segmental deletion is not a single operation, but is the result of a set of featural deletions

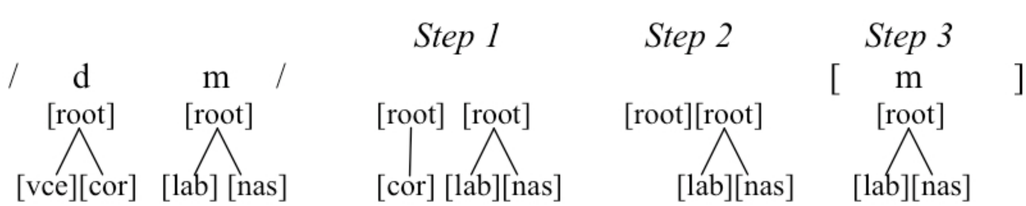
The following are serial HG reformulations of MAX-FEATURE constraints:

(42) *DELETE-VOICE Assign a penalty of -1 to an candidate produced by deleting a voice feature

*DELETE-PLACE Assign a penalty of -1 to a candidate produced by deleting a place feature

A derivation for /bad+ma/ → [bama] could proceed as follows:

(43) Gradual deletion in Harmonic Serialism



For Step 1 to be harmonically improving we need a weighting with w(*Coda-Voice) > w(*Delete-Voice), as in the following tableau:

(44)

Weight	1.5	1	$\mathcal{H}$
/bad+ma/	*Coda-Voice	*Delete-Voice	
☞ batma		-1	-1
badma	-1		-1.5

Which will equally lead to deletion of the [voice] feature in every coda:

(45)

Weight	1.5	1	$\mathcal{H}$
/bad/	*Coda-Voice	*Delete-Voice	
bad	-1		-1.5
☞ bat		-1	-1

*Coda-Voice has nothing to say about whether these representations continue along the chain to full deletion; this will depend on the weightings of constraints that penalize other aspects of the coda.

The result of adopting HS for HG:

(46) Because of gradualness, there is no way for Agree-Place and *Coda-Voice to gang up to force deletion of a segment

More broadly:

(47) Gradualness provides additonal limits on gang effects along with those that are inherent to HG

6.2 Serial HG in metrical phonology

A proposal from Pruitt (in prep.):

(48) Metrical structure is built incrementally in accordance with harmonic improvement

To see the consequences of this for HG, we can first consider why the HG + gradient Alignment combination is such a disaster

In the following tableaux, foot edges are indicated by parentheses, and prosodic word edges by square brackets. Align(Ft, Wd, L) demands that the left edge of every foot be aligned with the left edge of the word, and is violated by each syllable intervening between these two edges. Parse-Syl is violated by every syllable that fails to be parsed into a foot.

(49) *Violation profiles*

	Align(Ft, Wd, L)	Parse-Syl
a. [(ta.ta)(ta.ta)(ta.ta)]	2 + 4 = 6	0
b. [(ta.ta)(ta.ta)ta.ta]	2	2
c. [(ta.ta)ta.ta.ta.ta]	0	4
d. [ta.ta.ta.ta.ta.ta]	0	6

Align(Ft, Wd, L) and Parse-Syl conflict in that every foot added after the leftmost one satisfies Parse-Syl at the cost of violating Alignment. This cost increases as feet are added: the second foot from the left adds two violations, the third one adds four, and so on. This type of cost increase can interact with weighting to produce results impossible in OT (see Prince 2002a on collective harmonic bounding and HG/OT differences):

(50) *An optimum that can be picked by HG but not OT*

Weight	1	1.5	$\mathcal{H}$
	Align(Ft, Wd, L)	Parse-Syl	
a. [(ta.ta)(ta.ta)(ta.ta)]	-6	0	-6
☞ b. [(ta.ta)(ta.ta)ta.ta]	-2	-2	-5
c. [(ta.ta)ta.ta.ta.ta]	0	-4	-6
d. [ta.ta.ta.ta.ta.ta]	0	-6	-9

HG in fact predicts an infinite number of possible languages: any maximum of n feet can be imposed (see also Legendre et al. 2006; Pater et al. 2007a).

If we add one foot at a time, it is not at all obvious that we need gradient alignment.[3] For example, one (probably too) simple approach to left-to-right footing would make use of this constraint:

(51) *x(Assign a penalty of -1 to a foot preceded by an unfooted syllable

As the following table shows, misalignment of the first foot assigned, or any subsequent foot, gets just one violation:

(52) *Violation profiles for incremental foot assignment*

		*x(
Assignment of first foot	[(ta.ta)ta.ta.ta.ta]	0
	[ta.(ta.ta)ta.ta.ta]	-1
Assignment of second foot	[(ta.ta)(ta.ta).ta.ta.ta]	0
	[(ta.ta)ta.(ta.ta).ta.ta]	-1
Assignment of third foot	[(ta.ta)(ta.ta)(ta.ta)ta]	0
	[(ta.ta)(ta.ta)ta(tata)]	-1

The consequence for HG/OT differences:

(53) When each foot can get at most one violation of the "Alignment" constraint, we cannot get the HG/OT difference illustrated in (50)

[3] McCarthy's 2003b proposes that OT constraints are non-gradient, arguing that stress typology is better handled by Kager's Lapse constraints. There remain some issues, and some reasons to think that something like gradient Alignment may be necessary in standard OT: see Alber 2005a; 2005b and Hyde 2007.

The broader point:

(54) When constraints evaluate locally, we may not get the "complicated trade-offs" Prince and Smolensky (1993/2004: 232) worry about

In other words, when evaluation is local, it becomes unclear whether the added restrictiveness of ranking is necessary.

7 Conclusions

Much remains to be done in terms of further developing and evaluating HG as a theory of phonological typology, learning and gradience. However, the results we have thus far indicate that it has significant advantages over OT in modeling learning and gradience, and that it is also viable as a framework for typological study. This viability appears to be dependent on adopting Harmonic Serialism's model of Gen/Eval interaction (or some other mode of local evaluation). Much also remains to be done in developing that model (including learning theories), and assessing HG/OT differences within it.

References

Alber, Birgit. 2005a. Clash, Lapse and Directionality. *Natural Language and Linguistic Theory* 23:485–542.

Alber, Birgit. 2005b. Introduction to metrical theory. Handout from ESSLLI 2005., July 2005.

Becker, Michael, Joe Pater, and Christopher Potts. 2007. OT-Help: Java tools for Optimality Theory Software available on-line at http://web.linguist.umass.edu/ OTHelp/.

Boersma, Paul. 1997. How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* 21:43–58. Available on Rutgers Optimality Archive, http://ruccs.rutgers.edu/roa.html.

Boersma, Paul. 1998. *Functional Phonology: Formalizing the Interaction Between Articulatory and Perceptual Drives*. The Hague: Holland Academic Graphics.

Boersma, Paul, and Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32:45–86. Available on Rutgers Optimality Archive, http://ruccs.rutgers.edu/roa.html.

Boersma, Paul, and Joe Pater. 2007. On the convergence properties of a gradual learning algorithm for Harmonic Grammar. Ms, University of Amsterdam and UMass Amherst.

Boersma, Paul, and David Weenink. 2007. Praat: doing phonetics by computer (Version 4.6) [Computer program]. Retrieved May 16, 2007 from http://www.praat.org/. Developed at the Institute of Phonetic Sciences, University of Amsterdam.

Goldsmith, John. 1993. Harmonic phonology. In *The last phonological rule*, ed. John Goldsmith, 21–60. Chicago: University of Chicago Press.

Goldwater, Sharon, and Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. In *Proceedings of the stockholm workshop on variation within optimality theory*, ed. Jennifer Spenader, Anders Eriksson, and Östen Dahl, 111–120. Stockholm: Stockholm University.

Hayes, Bruce, Bruce Tesar, and Kie Zuraw. 2003. OTSoft 2.1. URL `http://www.linguistics.ucla.edu/people/hayes/otsoft/`, Software package developed at UCLA.

Hyde, Brett. 2007. Nonfinality and weight sensitivity. *Phonology* To appear.

Jäger, Gerhard. 2006. Maximum entropy models and Stochastic Optimality Theory. In *Architectures, rules, and preferences: A festschrift for Joan Bresnan*, ed. Jane Grimshaw, Joan Maling, Chris Manning, Jane Simpson, and Annie Zaenen. Stanford, CA: CSLI. To appear.

Jesney, Karen. 2007. The locus of variation in weighted constraint grammars. In *Paper presented at the Stanford Workshop on Variation, Gradience and Frequency in Phonology*.

Jesney, Karen, and Anne-Michelle Tessier. 2007. Re-evaluating learning biases in Harmonic Grammar. In *University of massachusetts occasional papers in linguistics 36: Papers in theoretical and computational phonology.*, ed. Michael Becker. GLSA.

Johnson, Mark. 2002. Optimality-theoretic lexical functional grammar. In *The lexical basis of syntactic processing: Formal, computational and experimental issues*, ed. Suzanne Stevenson and Paola Merlo, 59–73. John Benjamins.

Kager, René. 2006. Rhythmic licensing: An extended typology. In *Proceedings of the 3rd International Conference on Phonology*. Seoul, Korea: The Phonology-Morphology Circle of Korea. Available from.

Kawahara, Shigeto. 2006. A faithfulness ranking projected from a perceptibility scale: The case of [+voice] in Japanese. *Language* 82:536–574.

Keller, Frank. 2000. Gradience in grammar: Experimental and computational aspects of degrees of grammaticality. Doctoral Dissertation, University of Edinburgh.

Keller, Frank. 2006. Linear optimality theory as a model of gradience in grammar. In *Gradience in grammar: Generative perspectives*, ed. Gisbert Fanselow, Caroline Féry, Ralph Vogel, and Matthias Schlesewsky. Oxford: Oxford University Press.

Legendre, Geraldine, Yoshiro Miyata, and Paul Smolensky. 1990. Harmonic Grammar – A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. In *Proceedings of the twelfth annual conference of the cognitive science society*, 388–395. Mahwah, NJ: Lawrence Erlbaum Associates.

Legendre, Géraldine, Antonella Sorace, and Paul Smolensky. 2006. The Optimality Theory–Harmonic Grammar connection. In Smolensky and Legendre (2006), 339–402.

McCarthy, John. 2006a. Slouching towards optimality: coda reduction in OT-CC. Available as ROA-878 on the Rutgers Optimality Archive, http://roa.rutgers.edu.

McCarthy, John. 2007. *Hidden generalizations: Phonological opacity in optimality theory*. Advances in Optimality Theory. London, UK: Equinox.

McCarthy, John J. 1999. Sympathy and phonological opacity. *Phonology* 16:331–399.

McCarthy, John J. 2000. Harmonic serialism and parallelism. In *Proceedings of the north east linguistics society 30*, ed. Masako Hirotani, 501–524. Amherst, Mass.: GLSA Publications. Can be downloaded from http://www-unix.oit.umass.edu/ jjmccart/.

McCarthy, John J. 2003a. Comparative Markedness. *Theoretical Linguistics* 29:1–51.

McCarthy, John J. 2003b. OT constraints are categorical. *Phonology* 20:75–138.

McCarthy, John J. 2006b. Restraint of analysis. In *Wondering at the natural fecundity of things: Essays in honor of Alan Prince*, chapter 10. Linguistics Research Center. URL `http://repositories.cdlib.org/lrc/prince/10`.

Nishimura, Kohei. 2003. Lyman's law in loanwords. Ms, University of Tokyo.

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, February 11.

Pater, Joe. 2007. Gradual learning and convergence. *Linguistic Inquiry* To appear.

Pater, Joe, Rajesh Bhatt, and Christopher Potts. 2007a. Linguistic optimization. Ms, University of Massachusetts, Amherst.

Pater, Joe, Christopher Potts, and Rajesh Bhatt. 2007b. Harmonic Grammar with Linear Programming. Ms., UMass Amherst.

Potts, Christopher, Michael Becker, Rajesh Bhatt, and Joe Pater. 2007. HaLP: Harmonic grammar with linear programming, version 2. Software available online at http://web.linguist.umass.edu/˜halp/.

Prince, Alan. 2002. 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, and Paul Smolensky. 1993/2004. Optimality Theory: Constraint interaction in generative grammar. RuCCS Technical Report 2, Rutgers University, Piscateway, NJ: Rutgers University Center for Cognitive Science. Revised version published 2004 by Blackwell. Page references to the 2004 version.

Rosenblatt, Frank. 1958. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review* 65:386–408.

Smolensky, Paul. 2006. Optimality in phonology II: Harmonic completeness, local constraint conjunction, and feature domain markedness. In Smolensky and Legendre (2006), 27–160.

Smolensky, Paul, and Géraldine Legendre. 2006. *The harmonic mind: From neural computation to Optimality-Theoretic grammar*. Cambridge, MA: MIT Press.

Soderstrom, Melanie, Donald W. Mathis, and Paul Smolensky. 2006. Abstract genomic encoding of universal grammar in optimality theory. In *The harmonic mind: From neural computation to optimality-theoretic grammar. volume 2: Linguistic and philosophical implications.*, ed. Paul Smolensky and GÈraldine Legendre, 403–471. MIT Press.

Tesar, Bruce. 1995. Computational Optimality Theory. Doctoral Dissertation, University of Colorado, Boulder, CO.

Tessier, Anne-Michelle. 2006. Biases and stages in phonological acquisition. Doctoral Dissertation, UMass Amherst. ROA-833.

Wilson, Colin. 2003. Analyzing unbounded spreading with constraints: marks, targets and derivations. *Unpublished ms, UCLA* .

Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30:945–982.