Serial Harmonic Grammar

Joe Pater
University of Massachusetts, Amherst

Colloquium, University of Southern California,
October 6th 2008

In Optimality Theory (Prince and Smolensky 1993/2004) ***ranked*** constraints evaluate a candidate set that includes the results of all the possible sets of representational changes to the input – evaluation is in ***parallel***.

In Serial Harmonic Grammar, ***weighted*** constraints evaluate a ***series*** of minimal representational changes. The aim of this talk is to introduce S-HG, and to illustrate some of its advantages over the original version of OT. I will do this using the same case that PS used to introduce OT: syllabification in Imdlawn Tashlhiyt Berber

Glossary: "Harmonic Serialism" = serial OT or HG
"Harmonic Grammar" = serial or parallel HG

1. Serialism illustrated: Berber syllabification

Prince and Smolensky (1993/2004: ch. 2):

(1) i. Analysis of Berber syllabification (following Dell and Elmedlaoui 1985 *et seq*.; henceforth DE85 *et seq*.)

ii. Introduction of Optimality Theory

iii. Introduction of Harmonic Serialism

Though PS reject Harmonic Serialism in favor of parallelism, it has recently been revived in McCarthy's (2000, 2006 *et seq*.) recent work.

Here I provide a slightly revised version of PS's serial analysis of Berber, to illustrate how the theory works, and provide a further argument for it.

The first amazing fact about syllabification in Imdlawn Tashlhiyt Berber (ITB):

(2) *Any* consonant can be the nucleus of a syllable

Some alternations from DE85:

(3)	/i + *root*/	/t + *root*/	
	.il.di.	.tL.di.	‘pull’
	.ir.ba.	.tR.ba.	‘carry on one’s back’
	.in.da.	.tN.da.	‘shake (milk)
	.if.si.	.tF.si.	‘untie’
	.ix.si.	.tX.si.	‘go out (fire)’

Amazing fact number 2 about ITB syllabification:

(4) Despite the fact that any consonant can be a nucleus, syllabification shows a preference for high sonority nuclei.

A famous example (DE85, PS):

(5) .rat.lult. *vs.* *.ra.tL.wLt. ‘you will be born’

Other languages display parallel “hard” or absolute restrictions on the sonority of nuclei:

(6) *English:* Only sonorants can occupy nucleus position

French: Only vowels can occupy nucleus position

The main analytic goal here is to derive the hard restrictions in English and French and the soft restriction in Berber from the same constraint(s).

This goal seems not yet to have been met, even in OT, which has in general been successful in reducing parallel hard and soft restrictions to a single motivating constraint. For example, a preference for intervocalic onsets can be derived from a seemingly “hard” Onset constraint:

(7)

/an/	DEP	ONSET
☞ an		*
?an	* !	

/aba/	DEP	ONSET
☞ a.ba		
ab.a		*

1.1 DE 1985

DE’s sonority scale:

(8) low vowel > high vowel > liquid > nasal
> voiced fricative > voiceless fricative
> voiced stop > voiceless stop

DE’s syllabification algorithm (slightly reworded):

(9) i. Locate highest sonority unsyllabified segment with no syllabified segment preceding it - N

ii. Form core syllable (Onset-Nuc) with N as nucleus

iii. Iterate until no more segments available

iv. Coda adjunction

v. Other rules

(10)	(ra)tlult	Build O-N, N = low vowel
	(ra)t(lu)lt	Build O-N, N = high vowel
	(ra)t(lu)(lT)	Build O-N, N = stop
	(rat)(lu)(lT)	Coda adjunction
	(rat)(lult)	“Prepausal annexation”

The DE (1985) analysis is an example of “markedness-based serialism” (MBS):

(11) The unmarked structure is built earlier in the derivation than the marked one, thus accounting for the preference for the unmarked structure

(note that this is how the onset preference is sometimes handled too)

Problem:

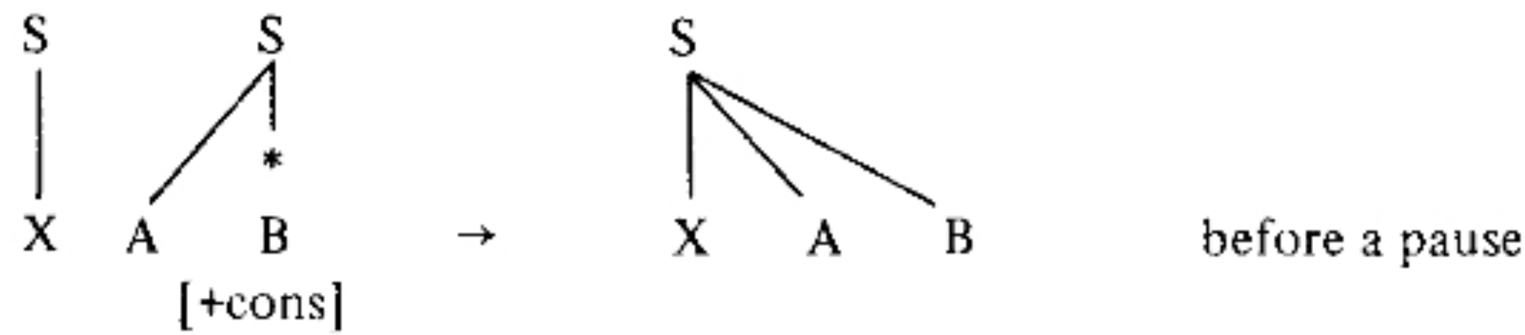
(12) What is the formal connection between the MBS account of the preference for high sonority nuclei, and related absolute restrictions?

ITB in fact has an absolute sonority-based restriction:

(13) In phrase-final position, obstruent nuclei are banned

DE's rule of pre-pausal annexation (DE p. 119) is simply stipulated to apply obligatorily only to obstruents:

(34)



e.g. .rat.lu.IT. → .rat.lult.

1.2 A Serial OT alternative

PS propose a serial OT analysis using the "Nuclear Harmony Constraint":

(14) H-NUC

A higher sonority nucleus is more harmonic than one of lower sonority.

Other constraints in PS, and almost all OT constraints since, assign violation scores. McCarthy (2003) suggests the following restatement of H-NUC, which I call *C-NUC (using an *n*-ary sonority feature; Hankamer and Aissen 1974, Selkirk 1982)

(15) *CONSONANTAL-NUCLEUS (*C-NUC)

Assign a violation score to a nucleus for each degree of sonority separating it from [a]

(16) *C-NUC(a) = 0 *C-NUC(i) = 1 *C-NUC(r) = 2,...

For our examples, we can simplify:

(17) **C-NUC violations*

Vowel = 0, Sonorant C = 1, Fricative = 2, Stop = 3

McCarthy (2003) rejects *C-NUC as a parallel OT constraint for reasons we'll soon see – we'll also see that it works fine in serial HG.

Along with *C-NUC, we need a constraint against hiatus to favor CV syllabification (= the 'ONSET' constraint in PS):

(18) *HIATUS

Assign a violation mark to a sequence of adjacent nuclei

In Harmonic Serialism, Gen produces a set of candidates minimally differing from the input (from PS):

(19) Gen ($input_i$):

The set of (partial) syllabifications of $input_i$ which differ from $input_i$ in no more than one syllabic adjunction.

I will assume the following “adjunctions”:

(20) *Operations*

HEAD PROJECTION (NUCLEUS)

e.g. a → (a) m → (M)

DEPENDENT ADJUNCTION (ONSET/CODA)

e.g. t(a) → (ta) (M)t → (Mt)

I assume an undominated Parse constraint, or feature of Gen, that forces syllabification at the expense of *C-NUC

(21) *Step 1*

/kšm/	*HIATUS	*C-NUC
(K)šm		3!
k(Š)m		2!
☞ kš(M)		1

I also assume that all previously assigned structure is retained. Only unshared violations in the tableaux (= new ones)

(22) *Step 2*

/kšM/	*HIATUS	*C-NUC
(K)š(M)		3!
k(Š)(M)	1!	2
☞ k(šM)		

In the final step, we're forced to put the stop in the nucleus.

(23) *Step 3*

/k(šM)/	*HIATUS	*C-NUC
☞ (K)(šM)		3

We have just replicated the result of DE's algorithm.

Interestingly, we get a different result with parallel OT
(no matter how we count *C-NUC violations)

(24) *Parallel OT*

/kšm/	*HIATUS	*C-NUC
(K)(šM)		3! (+1 = 4!)
☞ (kŠm)		2

The serial OT result is correct (DE1988).

In one parallel analysis, PS (ch.2) use H-NUC, which positively rewards high sonority nuclei through comparison

(25)

/kšm/	*HIATUS	H-NUC
☞ (K)(šM)		
(kŠm)		M > Š !

In the other (ch. 8), they use Margin constraints

(26)

/kšm/	*MARGIN-NASAL	*MARGIN-FRICATIVE
☞ (K)(šM)		*
(kŠm)	* !	

Margin constraints have been abandoned in subsequent OT syllabification studies (e.g. Gnanadeskian 1995, Baertsch 2002), since there is no cross-linguistic preference for low sonority codas (a problem recognized in PS ch. 8)

The point that rising sonority sequences like /kšm/ require *MARGIN constraints rather than *PEAK constraints (≈ *C-NUC) is due Donca Steriade (p.c.).

PS (1993/2004) and Clements (1997) focus on falling sonority sequences, which pose the same problem:

(27) *Parallel OT*

/ršq/	*HIATUS	*C-NUC
(R)(šQ)		3! (+1 = 4!)
☞ (rŠq)		2

The correct form is (R)(šQ), which we again get from the serial OT derivation. Here we see the effect of *HIATUS and NOCODA in getting the correct result.

(28) *Serial OT Derivation*

/ršq/	*HIATUS	NOCODA	*C-NUC
rš(Q)			3!
r(Š)q			2!
☞ (R)šq			1

(R)šq	*HIATUS	NOCODA	*C-NUC
(Rš)q		1!	
(R)(Š)q	1!		2
☞ (R)š(Q)			3

(R)š(Q)	*HIATUS	NOCODA	*C-NUC
(Rš)(Q)		1!	
(R)(Š)(Q)	2!		2
☞ (R)(šQ)			3

2. Harmonic Grammar and ITB

We have just seen that a serial OT analysis of ITB allows us to use *C-Nuc, a constraint that unlike H-Nuc assigns violation scores, and that unlike the Margin constraints is typologically well supported. This provides an argument for serial OT, alongside those developed in McCarthy's recent work (see also forthcoming UMass Occasional Papers on Harmonic Serialism).

As well as having typological problems (see also Clements 1997), *MARGIN constraints cannot account for English or French restrictions on possible nuclei, since an obstruent nucleus does not violate the constraint. We will now see how *C-Nuc can account for these hard restrictions, along with the Berber one, if we adopt weighted constraints.

Though McCarthy (2003) proposes *C-NUC, he also quickly rejects it, because it doesn't work with ranked constraints

English allows sonorant consonants as nuclei (in unstressed syllables), but it doesn't allow lower sonority nuclei

If we consider the interaction of *C-NUC with a constraint against epenthesis (DEP), we only get two languages in OT:

(29) DEP >> *C-NUC
All segments can be nuclei (Berber)

(30) *C-NUC >> DEP
Only vowels (low vowels?) can be nuclei (French?)

English is impossible with this constraint set in OT.

Harmonic Grammar

(Legendre, Miyata and Smolensky 1990, Flemming 2001, Smolensky and Legendre 2006, Pater 2008, Potts et al. 2008)

(31) The well-formedness of a representation is the sum of weighted violation scores (Harmony)

With weighted constraints, *C-Nuc can account for English

(violations converted to negative integers, since they are penalties – the optimum has maximal harmony)

(32)

/tn/	DEP 1.5	*C-NUC 1	
☞ tN		–1	–1
tVn	–1		–1.5

/ts/	DEP 1.5	*C-NUC 1	
tS		–2	–2
☞ tVs	–1		–1.5

In Berber, DEP is weighted high enough to force any degree of *C-Nuc violation (*e.g.* 4 if T=3):

(33)

/tn/	DEP 4	*C-NUC 1	
☞ tN		–1	–1
tVn	–4		–4

/ts/	DEP 4	*C-NUC 1	
☞ tS		–2	–2
ts	–1		–4

In French, *C-NUC is weighted higher than DEP
(assuming it doesn't penalize vowels)

(34)

/tn/	*C-NUC 1.5	DEP 1	
tN	–1		–1.5
☞ tVn		–1	–1

/ts/	*C-NUC 1.5	DEP 1	
tS	–2		–3
☞ tVs		–1	–1

Outcome:

(35) A single constraint gets absolute restrictions on the sonority of nuclei in English and French as well as preference for highly sonorous nuclei in Berber

As far as I know, this is the first analysis to accomplish this. Clements (1997) and Dell and Elmedlaoui (2002) use a constraint for Berber that can be stated as follows:
(for our purposes - as elsewhere in this talk, I abstract from sonority “plateaux”)

(36) Sonority Peak
If a segment is of greater sonority than the segments adjacent to it, it is syllabified as a nucleus

The Sonority Peak constraint does not penalize a syllable like (tS), which is ill-formed in English and French.

As McCarthy (2003) shows (following PS ch. 8), an OT analysis of English or French would require splitting *C-NUC into different constraints for each step of the sonority scale. While this would work (as would constraints in a "stringency" relation), we can get further mileage out of HG interaction.

The Berber ban on phrase-final obstruent nuclei can also be attributed to *C-NUC in HG if we add a constraint penalizing final nuclei (cf. McCarthy and Prince's 1994 word-final-C constraint)

(37) *FINAL-N

Assign a violation mark to a phrase-final nucleus

The *additive* interaction of *C-NUC and *FINAL-N produce the desired result:

(38) *Final annexation as a gang effect*

(rat)(lu)lt	*CODA 3	*C-NUC 1	*FINAL-N 1	
☞ (rat)(lult)	–1			–3
(rat)(lu)(lT)		–3	–1	–4

Why a gang effect? Because *FINAL-N is not strong enough to beat *CODA on its own. For example, sonorant consonants can (optionally) surface as final nuclei.

(39) *Sonorants as final nuclei*

.i.gi.dR. ‘eagle’ .R.gL. ‘lock’ .du.mN. ‘they (m.) last’

The violation score of –1 on *C-NUC contributed by the sonorant is not sufficient to help *FINAL-N overcome *CODA.

(40) *Escape from final annexation*

(i)(gi)dr	*CODA 3	*C-NUC 1	*FINAL-N 1	
☞ (i)(gi)(dR)		–1	–1	–2
(i)(gidr)	–1			–3

We're now in a position to do a serial HG derivation for /ratlult/. As in Legendre *et al*. (2006) point out in their HG translation of the PS (1993/2004) analyses, *HIATUS must have a value greater than the highest possible score on *C-NUC. Since we are assuming T=3, then *HIATUS is set at 4.

(41)	*HIATUS	*CODA	*C-NUC	*FINAL-NUC
	4	3	1	1

Optimum	*Failed candidates*				
(ra)tlult (0,0,0,0)=0	r(aT)lult (0,0,3,0)=3	ra(tL)ult (0,0,1,0)=1	rat(lu)lt (0,0,0,0)=0	ratl(wL)t (0,0,1,0)=1	ratlu(lT) ... (0,0,3,1)=4
(ra)t(lu)lt (0,0,0,0)=0	(rat)lult ... (0,1,0,0)=3				
(rat)(lu)lt (0,1,0,0)=3	(ra)t(lu)(lT) (0,0,3,1)=4	(ra)t(lu)(L)t (1,0,1,0)=5			
(rat)(lul)t (0,1,0,0)=3	(rat)(lu)(lT) (0,0,3,1)=4				
(rat)(lult) (0,1,0,0)=3	(rat)(lul)(T) (0,0,3,0)=4				

Notes:

1. Nucleus projection and onset adjunction collapsed into one step
2. Parentheses contain violation vectors (multiplied by weights to get scores)
3. rat(lu)lt ties with optimum in first step because high and low vowels are treated uniformly by *C-Nuc

Most words with final sonorant consonants as nuclei optionally undergo final annexation:

(42) *Optional final annexation*

.i.gi.dR. / .i.gidr. .R.gL. / .Rgl. .du.mN. / .dumn.

If we increase the weight of *FINAL-NUC by 1, we get a tie in these cases:

(43) *Optional final annexation as a tie*

(i)(gi)dr	*CODA 3	*FINAL-N 2	*C-NUC 1	
☞ (i)(gidr)	–1			–3
☞ (i)(gi)(dR)		–1	–1	–3

This weighting yields a single choice with a final obstruent or final vocoid (i, u, a)

(44) *Obligatory final obstruent coda*

(rat)(lu)lt	*CODA 3	*FINAL-N 2	*C-NUC 1	
☞ (rat)(lult)	–1			–3
(rat)(lu)(lT)		–1	–3	–5

(45) *Obligatory final vocalic nucleus*

(tL)di	*CODA 3	*FINAL-N 2	*C-NUC 1	
(tLdj)	–1			–3
☞ (tL)(di)		–1		–2

Actual theories of variation in HG produce probability distributions over candidates:

(46) *a. Maximum Entropy Grammar*
Johnson (2002), Goldwater and Johnson (2003)

b. Noisy Harmonic Grammar
Boersma and Pater (2008)

A particularly attractive aspect of these theories is that they are accompanied by learning algorithms that can accurately learn categorical patterns (cf. Pater 2008 on OT-GLA) and also learn patterns of variation.

It is not yet known if these are compatible with HS, but early results using HG-GLA with categorical HS are promising (Pater 2008 ms., Pratt in prep.); see also Kimper (2008: ROA) for arguments for a serial approach to variation.

Though incomplete, this analysis of Berber illustrates three arguments for weighted constraints:

(47) i. Constraints like *C-NUC can state scalar generalizations directly, rather than in the indirect fashion required for OT's ranked constraints, in which fixed rankings or stringency relations amongst a set of constraints are derived from a scale

ii. Account of gang effects (AKA cumulative constraint interaction), such as between *FINAL-NUC and *C-NUC (For comparison with Smolensky's OT with Local Constraint conjunction, see Pater 2008 Harmonic Mind Review)

iii. Straightforward extensions to variation, and its learning (see Coetzee and Pater 2008 Handbook of Phonology for a comparison of OT and HG theories of variation)

3. HS and HG typological plausibility

Prince and Smolensky (1993/2004) address the following "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 irredeemably 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.

Prince and Smolensky (1993/2004) do not document the “complicated trade-offs” produced by weighted constraints.

An example of a system produced by parallel HG but not parallel OT (see Legendre et al. 2006, Pater 2008 ms. for others)

(48) ONSET is satisfied if it requires 2 instances of metathesis (each one violating LINEARITY) but not 3

(49)

adat	ONS 2.5	LIN 1	
☞ data		–2	–2
adat	-1		–2.5
adta	–1	–1	–3.5

/adatana/	ONS 2.5	LIN 1	
☞ adatana	–1		–2.5
datana		–3	–3
daatana	–1	–1	–3.5

The parallel OT ranking ONS >> LIN produces metathesis across the board, but even /adat/ → [data] is unattested (McCarthy 2006).

McCarthy (2006) points out that if each metathesis is a single operation, HS does not produce the unattested result, since the derivation will terminate on the first step:

(50)

adat	ONS	LIN
☞ adat	–1	
daat	–1	–1
adta	–1	–1

This result holds in both serial OT and serial HG, since pairwise harmonic bounding in OT in maintained in HG (Prince 2002).

Since HS is limited to a single application of each operation when it forms a candidate set, only a single violation of a given faithfulness constraint can be traded off against other violations. This inherently limits trade-offs (Pater 2008).

When violations trade off one-to-one, ranking and weighting are indistinguishable (Prince 2002, Pater 2008), and they are also indistinguishable in many cases of one-to-many trade-off (Tesar 2007, Pater 2008).

(51) *Lardil tableau (=P&S 1993/2004: (183 A))*

/yiliyili/	FREE-V	ALIGN	PARSE	NOCODA
i.☞ .yi.li.yil.<i>		1	1	1
ii. .yi.li.yi.li.	1			

Both HG and OT produce two languages: one like Lardil in which NOCODA is violated at the cost of satisfying FREE-V, and one in which FREE-V satisfaction is blocked by NOCODA violation. Any "gang effect" between NOCODA and the other constraints would be *vacuous*, since FREE-V satisfaction entails violation of ALIGN and PARSE.

4. The Big Picture

(At least) four versions of OT currently on the table:

(52) Parallel OT, Serial OT, Parallel HG, Serial HG

Deciding between them is complicated by this fact:

(53) Each theory can/must employ different constraint sets to deal with attested phenomena

As an example, note that *C-NUC gets Berber, English and French only in Serial HG. The other theories need different constraint sets for these attested languages.

The upshot: (see Potts et al. 2008 for a parallel OT/HG example)

(54) None of the theories is likely in a subset relationship with any of the others in terms of what they generate; no one theory is inherently more restrictive

To determine which of these theories can best fit the typological data, software help is essential. Luckily, some of that help is already available, and more is on the way.

	OT-Soft	OT-Help 1	OT-Help 2
Parallel OT	✓	✓	✓
Parallel HG		✓	✓
Serial OT/HG			✓

OT-Soft: Hayes, Tesar and Zuraw (2003)

OT-Help 1: Becker, Pater, and Potts (2007)

OT-Help 2: Becker, Pater, Potts, and Pratt (in prep.)

Selected References

Boersma, Paul and Joe Pater. 2008. Convergence Properties of a Gradual Learning Algorithm for Harmonic Grammar. Ms., University of Amsterdam and University of Massachusetts, Amherst (roa.rutgers.edu).

Coetzee, Andries and Joe Pater. 2008. The place of variation in phonological theory. To appear in John Goldsmith, Jason Riggle, and Alan Yu (eds.), The Handbook of Phonological Theory (2nd ed.). Blackwell. (roa.rutgers.edu).

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.

Legendre, Géraldine, Antonella Sorace, and Paul Smolensky. 2006. The Optimality Theory–Harmonic Grammar connection. In Smolensky and Legendre (2006), 903–966.

McCarthy, John J. 2006. 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. To appear. The serial interaction of stress and syncope. http://people.umass.edu/jjmccart/metrically-conditioned-syncope.pdf

Pater, Joe. 2008. Review of Smolensky and Legendre (2006). The Harmonic Mind. To appear in Phonology. (roa.rutgers.edu).

Pater, Joe. 2008. Optimization and Linguistic Typology. Ms., University of Massachusetts, Amherst. (roa.rutgers.edu).
Pater, Joe. 2008. Gradual learning and convergence. Linguistic Inquiry 39/2. 334-345.
Potts, Christopher, Joe Pater, Rajesh Bhatt and Michael Becker. 2008. Harmonic Grammar with Linear Programming: From linear systems to linguistic typology Ms, University of Massachusetts, Amherst. (roa.rutgers.edu)
Pratt, Patrick. In prep. Computational Harmonic Serialism. Honor’s Thesis, University of Massachusetts, Amherst.
Prince, Alan. 2002. Anything goes. In *New century of phonology and phonological theory*, ed. by Takeru Honma, Masao Okazaki, Toshiyuki Tabata, and Shin-ichi Tanaka, 66–90. Tokyo: Kaitakusha (roa.rutgers.edu).
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.rutgers.edu).
Smolensky, Paul, and Geraldine Legendre. 2006. *The harmonic mind: From neural computation to Optimality-Theoretic grammar*. Cambridge, MA: MIT Press.
Tesar, Bruce. 2007. A comparison of lexicographic and linear numeric optimization using violation difference ratios. Ms, Rutgers University. (roa.rutgers.edu).