2016

The theory and practice of Harmonic Serialism

John J. McCarthy, *University of Massachusetts - Amherst*

The theory and practice of Harmonic Serialism[1]

John J. McCarthy

1 Introduction

Harmonic Serialism (HS) is a variant of Optimality Theory. The difference between HS and standard OT is that HS is a *serial* model and standard OT is a *parallel* model. In this chapter, I will explain what that means, how they differ, and how HS works.

This chapter begins (section 2) with a simple example and an overview of HS's basic architecture, *modus operandi*, and provenance. Section 3 then provides a more extended illustration of HS, using the well-known syllabically conditioned alternations of Yawelmani Yokuts. Section 4 develops some results about the nature of HS's derivations, and section 5 briefly treats the principal consequences that HS has for the formulation of constraints. Section 6 explains a key element of the process of analysis in HS: the trade-off between GEN and CON as the locus of linguistic explanation. Comparison of HS and parallel OT P-OT on empirical grounds is the focus of section 7, and finally section 8 addresses the question of whether HS's derivations make it a more successful theory of opacity than P-OT.

2 An overview of Harmonic Serialism

In P-OT, GEN can make many changes at once when it produces a candidate. This means that competing candidates can differ in many ways from each other and from the underlying representation. For example, Yawelmani Yokuts has an underlying→surface mapping in which /taxa:-kˀa/ 'bring!' becomes *ta.xakˀ* by apocope and closed syllable shortening (Newman 1944). (The period/full stop marks a syllable boundary.) In a P-OT analysis of Yawelmani, the candidates from underlying /taxa:-kˀa/ include some that differ from it by a single change (such as *ta.xa:kˀ*, which differs only by virtue of a single deletion) and some that differ by two or more changes (such as *ta.xakˀ*, which differs by both deletion and shortening, or *taxa*, which differs by two deletions and a shortening process). In P-OT, GEN applies one or more operations together, in parallel, when it generates candidates. In consequence, P-OT's candidate sets are quite diverse and involve competition between candidates that may not resemble each other very much.

In HS, GEN is limited to making just one change at a time. The candidate set from underlying /taxa:-kˀa/ consists of all of the ways of making no more than a single change in this form: *ta.xa:kˀ*, *ta.xa.kˀa*, and so on, but not *ta.xakˀ* or *taxa*, which require two or more changes. In consequence, HS's candidate sets are less diverse than P-OT's, and they involve competition between candidates that are not very different from one another. This property of HS's GEN is called *gradualness*.

HS would be unworkable as a theory of language if all it did was to impose this limitation on GEN. That is because underlying and surface representations can in fact differ from one another in several ways. For instance, the surface form from underlying /taxa:-kˀa/ is *ta.xakˀ*, which shows the effects of both vowel deletion and vowel shortening. As we just saw, *ta.xakˀ* is not one of the candidates that HS's GEN produces from /taxa:-kˀa/. So how does the grammar get to it?

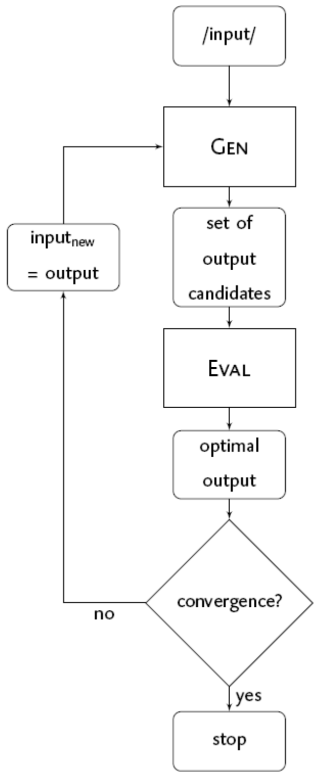
The answer is that HS has a loop. In HS, the optimal candidate chosen by EVAL becomes a new input to GEN, which forms a candidate set that goes to EVAL, and so on. The loop continues until EVAL picks an optimum that is identical with the most recent input to GEN. So the HS derivation for /taxaː-k^{ʔ}a/ goes something like this:

(1) Derivation for /taxaː-k^{ʔ}a/ in HS

/taxaː-k^{ʔ}a/
↓
GEN
↓
ta.xaː.k^{ʔ}a, ta.xaːk^{ʔ}, ta.xa.k^{ʔ}a, …
↓
EVAL
↓
ta.xaːk^{ʔ}
↓
GEN
↓
ta.xaːk^{ʔ}, ta.xak^{ʔ}, ta.xaː, …
↓
EVAL
↓
ta.xak^{ʔ}
↓
GEN
↓
ta.xak^{ʔ}, ta.xa, ta.xa.k^{ʔ}i, …
↓
EVAL
↓
ta.xak^{ʔ}
Convergence

Each pass through GEN and EVAL is called a *step*. The candidate set at each step includes the unchanged input to that step and all of the ways of making a single change in that input. (Syllabification, including resyllabification, gets special treatment. More about this in section 6.) The derivation in (1) is over when a step begins and ends with the same form, *converging* on *ta.xak^{ʔ}*. This is the final output of the grammar — the surface form for underlying /taxaː-k^{ʔ}a/. The whole model is summarized in the flowchart in (2).

(2) HS flowchart[2]



I will first illustrate HS with an example that is simpler than Yawelmani, but we will return in section 3 to look in detail at how the derivation in (1) comes about.

In Classical Arabic, word-initial consonant clusters are prohibited. When they occur in underlying representations, glottal stop and a high vowel are preposed: /fʕal/ → *ʔif.ʕal* ‘do!’. Two of the markedness constraints involved are *COMPLEX-ONSET, which is violated by faithful *fʕal*, and ONSET, which is violated by *if.ʕal*, with epenthetic *i* but not epenthetic *ʔ*.[3] In a P-OT analysis, both of these constraints are ranked at the top of the hierarchy, dominating DEP:[4]

(3) P-OT analysis of Arabic /fʕal/ → *ʔif.ʕal*

/fʕal/	*COMP-ONSET	ONS	MAX	CONTIG	DEP
a. → ʔif.ʕal					2
b. fʕal	1 W				L
c. if.ʕal		1 W			1 L
d. ʕal			1 W		L
e. fi.ʕal				1 W	1 L

Tableau (3) includes two other constraints that are also active. Because MAX dominates DEP, epenthesis is favored over deletion. And because CONTIGUITY (Kenstowicz 1994, McCarthy and Prince 1995, 1999) dominates DEP, initial epenthesis of *i* is preferred to medial epenthesis, even though initial epenthesis of *i* ends up requiring epenthesis of *ʔ* as well, because of ONSET.

Now we will look at the HS analysis, followed by comparison of the two. HS's GEN cannot epenthesize two segments at once, so at the beginning of the derivation the ultimate winner *ʔif.ʕal* is not in the candidate set. Instead, the candidate set is limited to forms that differ from /fʕal/ by at most one change: faithful *fʕal* and unfaithful *if.ʕal*, *ʕal*, *fi.ʕal*, etc. We want *if.ʕal* to win at this step of the derivation, because it is the only candidate that will get us eventually to *ʔif.ʕal*. And for *if.ʕal* to win, *COMPLEX-ONSET, MAX, and CONTIGUITY have to dominate ONSET:

(4) HS analysis of /fʕal/ → *ʔif.ʕal* — Step 1

/fʕal/	*COMP-ONSET	MAX	CONTIG	ONS	DEP
a. → if.ʕal				1	1
b. fʕal	1 W			L	L
c. fal		1 W		L	L
d. fi.ʕal			1 W	L	1

At the next step of the derivation, the input to GEN is *if.ʕal*, and the candidate set includes *if.ʕal* and all of the ways of effecting a single change in it: *fʕal*, *i.fi.ʕal*, *ʔif.ʕal*, etc. The same grammar is applied to this new candidate set and chooses *ʔif.ʕal* as the optimum because ONSET dominates DEP:

(5) HS analysis of /fʕal/ → *ʔif.ʕal* — Step 2

if.ʕal	*COMP-ONSET	MAX	CONTIG	ONS	DEP
a. → ʔif.ʕal					1
b. if.ʕal				1 W	L
c. fʕal	1 W	1 W			L

Observe that the faithfulness violations are determined relative to the input to the current step of the derivation, not to the underlying representation. Hence, *ʔif.ʕal* receives only one violation mark from DEP in (5), and *if.ʕal* gets none at all. This is consistent with HS's basic assumptions and turns out to solve a problem with positional faithfulness, as shown by Jesney (2011). (See section 7.1 for a brief summary of Jesney's argument.)

Although *we* know that we have reached the desired surface form, *the grammar* does not yet know that. So it submits the most recent optimum, *ʔif.ʕal*, as input for another pass through GEN and EVAL. Once again, the candidate set consists of *ʔif.ʕal* and all of the other forms that are one change away from it. None of the changed forms wins, and *ʔif.ʕal* once again emerges as victorious:

(6) HS analysis of /fʕal/ → *ʔif.ʕal* — Step 3

ʔif.ʕal	*COMP-ONSET	MAX	CONTIG	ONS	DEP
a. → ʔif.ʕal					
b. ʔi.fi.ʕal					1 W
c. ʔif.ʕa		1 W			

At this point, the derivation has converged on the final output *ʔif.ʕal*. The GEN/EVAL loop terminates.

It is instructive to compare the P-OT and HS analyses of these facts. The P-OT analysis in (3) does not assert any ranking between *COMPLEX-ONSET, MAX, and CONTIGUITY on the one hand and ONSET on the other. All four of these constraints are unviolated in all surface forms, and surface-unviolated constraints do not conflict in P-OT. Thus, they are unrankable.

In the HS analysis, however, *COMPLEX-ONSET, MAX, and CONTIGUITY must dominate ONSET. This ranking is necessary because, although ONSET is not violated by the faithful candidate at the beginning of the derivation or the optimal candidate at the end of the derivation, it is violated by the intermediate winner *if.ʕal*. This is by no means an unusual situation in HS. In fact, it is exactly what we should expect to see in a derivational theory with ranked, violable constraints. Constraint conflicts can emerge in the course of the derivation, and those conflicts are resolved in the usual OT fashion: the conflicting constraints are ranked, and the lower-ranking constraint is violated to spare a violation of the higher-ranking one. See McCarthy, Pater, and Pruitt (this volume) for further discussion.

I will conclude this brief introduction to HS with some remarks on its provenance. From the outset, it is important to understand that HS is really just a version of OT, not a full-blown alternative to it. Indeed, HS was first proposed in OT's *locus classicus*, Prince and Smolensky (1993/2004), but it was not pursued there and was in fact rejected in favor of P-OT. The case for HS was reopened in McCarthy (2000, 2002: 159-163, 2007b), where some general consequences of this theory are identified and discussed.

HS is distinct from, though related to, OT with candidate chains (OT-CC), in which an HS-like system is used to construct derivations that then compete against one another (McCarthy 2007a, Wolf 2008, this volume). Other efforts to implement OT with derivations could also be mentioned (Black 1993, Chen 1999, Kiparsky 2000, Norton 2003, Rubach 2000), as well as non-OT theories of derivational constraint satisfaction, such as Harmonic Phonology (Goldsmith 1990, 1993) and the Theory of Constraints and Repair Strategies (Paradis 1988a, b).

It is often asked how HS differs from rule-based phonology (RBP) in the tradition of Chomsky and Halle (1968). It would be better to ask how they are alike, because there are many differences and few similarities. In fact, there are just two shared properties: both theories posit derivations with intermediate representations; and both place limits on how much can change from one step of a derivation to the next.[5] The differences consist of all of the other ways that OT differs from RBP: in RBP, a grammar is a list of language-particular rules; in OT, a grammar is a ranking of universal constraints; in RBP, rules change one representation into another; in OT, constraints compare candidates; and so on. All of the arguments that support OT over RBP (e.g., McCarthy 2002: 66-138), other than arguments that crucially depend on parallelism, apply with equal force to HS. For more about this comparison, see McCarthy, Pater, and Pruitt (this volume).

3 Doing Harmonic Serialism

Doing analysis in OT is hard because the required constraints are not always known in advance, so the analyst's task involves more than just finding a ranking. Indeed, one of the principal reasons for doing analysis in OT is to investigate the properties of the constraint component CON, because typologically informed analysis is the only way of studying CON.

Doing HS is harder than doing P-OT because the properties of GEN are so much more important in HS, and those properties are not always known in advance either. The goal of typologically informed analysis in HS is to investigate GEN, CON, and the relationship between the two. That sort of investigation is undertaken in section 6 and in many of the other contributions to this volume. But first, the basic analytic techniques must be mastered. Chapter 2 of McCarthy (2008a) uses syllabically conditioned alternations in Yawelmani Yokuts (Newman 1944) to explain how to do analysis in P-OT. This section uses the same example, to facilitate comparison.

The syllabic phonology of Yawelmani is mainly guided by two surface-unviolated constraints against syllables bigger than CVː or CVC. One of these constraints, *SUPERHEAVY, is violated by both CVːC or CVCC syllables. The other, *COMPLEX-CODA, is violated by CVCC syllables only. (On why both constraints are needed, see section 7 of McCarthy, Pater, and Pruitt (this volume).) One effect of *SUPERHEAVY is that long vowels are shortened in closed syllables:

(7) Closed syllable shortening (Kenstowicz and Kisseberth 1979: 83)

/laːn-hin/	lan.hin	'hear (nonfuture)'	cf. *laː.nal* 'id. (dubitative)'
/ṣaːp-hin/	ṣap.hin	'burn (nonfuture)'	cf. *ṣaː.nal* 'id. (dubitative)'

Another effect attributable to *SUPERHEAVY or *COMPLEX-CODA is that medial triconsonantal clusters are broken up by inserting *i* before the second consonant.

(8) Vowel epenthesis

/ʔilk-hin/	ʔi.lik.hin	'sing (nonfuture)'	cf. *ʔil.kal* 'id. (dubitative)'
/lihm-hin/	li.him.hin	'run (nonfuture)'	cf. *lih.mal* 'id. (dubitative)'

These examples are a good place to begin our HS analysis because the derivations are so shallow: the underlying and surface forms in each derivation differ by the effect of just one operation. The operations involved are mora deletion in (7) and vowel epenthesis in (8). Because the derivations are so shallow, the P-OT and HS tableaux are nearly identical. (On why it is possible to both syllabify and shorten/epenthesize/delete at Step 1, see section 6.)

(9) Step 1 of /laːn-hin/ → *lan.hin*[6]

/laːn-hin/	*COMP-CODA	*SUPER	PARSE-SEG	MAX	DEP	MAX-μ
a.→ lan.hin						1
b. laːn.hin		1 W				L
c. laː.ni.hin					1 W	L
d. laː.hin				1 W		L
e. laː.*n*.hin			1 W			L

(10) Step 1 of /ʔilk-hin/ → *ʔi.lik.hin*

/ʔilk-hin/	*COMP-CODA	*SUPER	PARSE-SEG	MAX	DEP	MAX-μ
a.→ ʔi.lik.hin					1	
b. ʔilk.hin	1 W	1 W			L	
c. ʔil.hin				1 W	L	
d. ʔil.*k*.hin			1 W		L	

In tableau (9), the faithful syllabification *laːn.hin* (9b) has a violation of *SUPERHEAVY. That violation could be avoided in various ways: by vowel shortening (9a), which violates MAX-μ, by vowel epenthesis (9c), which violates DEP, by consonant deletion (9d), which violates MAX, or by leaving *n* unsyllabified (9e), which violates PARSE-SEGMENT. Because the faithful candidate and the other alternatives violate constraints ranked higher than MAX-μ, (9a) is the winner. (Other candidates, not shown, might parse the *n* as a nucleus or as the onset of a degenerate syllable. They too violate constraints ranked higher than MAX-μ.)

In tableau (10), the situation is much the same, except that there is no long vowel to shorten. Because all of the other constraints dominate DEP, the winning candidate is (10a), with vowel epenthesis. (An alternative locus of vowel epenthesis, **ʔil.ki.hin*, is ruled out by ALIGN-RIGHT(stem, σ) (McCarthy and Prince 1993).)

At step 2, each of these derivations continues by submitting its output for another pass through GEN and EVAL. Readers can verify for themselves that the now-faithful candidates *lan.hin* and *ʔi.lik.hin* are more harmonic than any other candidates that can be reached in a single step from them, so the derivations converge on these final outputs.

Yawelmani also has an apocope process. It is blocked when it would create a final consonant cluster:

(11) Final vowel deletion blocked (Kenstowicz and Kisseberth 1979: 98)

/xat-kˀa/	xat.kˀa, *xatkˀ	'eat!'
/xat-mi/	xat.mi, *xatm	'having eaten'

The constraint responsible for apocope — call it *V# — must dominate MAX to be active. But it must itself be dominated by *SUPERHEAVY or *COMPLEX-CODA and PARSE-SEGMENT, to account for the absence of apocope in (11). This ranking is shown in (12):

(12) Step 1 of /xat-kˀa/ → *xat.kˀa*

/xat-kˀa/	*COMP-CODA	*SUPER	PARSE-SEG	*V#	MAX	DEP	MAX-μ
a.→ xat.kˀa				1			
b. xatkˀ	1 W	1 W		L	1 W		
c. xat.*kˀ*			1 W	L	1 W		

The candidate set in (12) was obtained under the same assumptions about GEN that produced the candidate sets in (9) and (10): it can (re)syllabify in parallel with unfaithful operations like deletion.

The candidate set in (12) is perhaps more noteworthy for what it does not contain as what it does. Missing are two candidates that might seriously challenge the winner: *xat*, with double deletion, and *xatikˀ*, with deletion and epenthesis. Neither of these forms is in the candidate set at step 1 because each is two steps away from the input. A P-OT analysis must take additional precautions to defeat these candidates, because unlike the intended winner they satisfy all four high-ranking markedness constraints, violating only low-ranking MAX and DEP.

Apocope may feed vowel shortening:

(13) Final vowel deletion with shortening (Kenstowicz and Kisseberth 1979: 98)

/taxaː-mi/	ta.xam	'having brought'
/taxaː-kˀa/	ta.xakˀ	'bring!'

The derivations of these forms involve two processes, segment deletion and mora deletion, and thus the derivations are longer than we have seen so far in Yawelmani.

For the HS analyst, the key questions in situations like this are:

i. Does it matter which process happens first?
ii. If it does matter, what is the order?
iii. What constraint ranking is necessary to force that order?

The answers for Yawelmani are:

i. Yes, it matters, because apocope creates the condition (a long vowel in a closed syllable) that necessitates shortening.
ii. For the reason just given, apocope must precede shortening.
iii. The ranking has to permit apocope even though it creates a configuration that requires subsequent shortening.

This ranking is a refinement of the ranking seen in the previous tableaux. *V# must dominate *SUPERHEAVY, the markedness constraint that CVːC syllables violate. Tableaux (14) and (15) show the first two steps of the apocope-shortening derivation under this ranking:

(14) Step 1 of /taxa:-mi/ → *ta.xam*

/taxa:-mi/	*COMP-CODA	PARSE-SEG	*V#	*SUPER	MAX	DEP	MAX-μ
a.→ ta.xa:m				1	1		
b. ta.xa:.mi			1 W	L	L		
c. ta.xa.mi			1 W	L	L		1 W

(15) Step 2 of /taxa:-mi/ → *ta.xam*

ta.xa:m	*COMP-CODA	PARSE-SEG	*V#	*SUPER	MAX	DEP	MAX-μ
a.→ ta.xam							1
b. ta.xa:m				1 W			L

The derivation then converges at step 3.

The situation in tableau (14) recalls a point made in the discussion of Arabic in section 2. It is not uncommon in HS to have derivational steps where a violation of a surface-true markedness constraint is introduced and later eliminated. For the analyst, it is important always to keep this possibility in mind, because without it many multi-process phonological derivations become unanalyzable. At step 1, Yawelmani swaps a violation of *V# for a violation of lower-ranking *SUPERHEAVY. That violation is eliminated at step 2, when the vowel shortens. The temporary violation of *SUPERHEAVY has to be tolerated because the ultimate winner, *ta.xam*, is not among the candidates that GEN produces at step 1.

4 Derivations in Harmonic Serialism

Derivations in HS are guaranteed to be finite and to show steady harmonic improvement until convergence. These two properties of HS derivations are related.

In HS, as in OT generally, harmony is the property that EVAL selects for. Given a constraint hierarchy H and two forms *A* and *B*, *A* is more harmonic than *B* if and only if the highest ranking constraint in H that distinguishes between *A* and *B* favors *A* over *B*. In other words, *A* is more harmonic than *B* relative to some constraint hierarchy if *A* satisfies that hierarchy better than *B* does.

Derivations in HS have to show steady harmonic improvement until convergence. The winner at step n has to be more harmonic than the winner at step $n-1$ (unless they are identical) because the step $n-1$ winner is a candidate at step n. That is because the step $n-1$ winner is the faithful candidate at step n, and by a standard assumption about GEN the faithful form is always a candidate.

Harmonic improvement tableaux like (16) are useful tools for studying derivations in HS. A harmonic improvement tableau shows the faithful candidate at the beginning and the winners at each successive step until convergence. By comparing each line in a harmonic improvement tableau with the one immediately above it, we can see why making the specified change is better than doing nothing at all. For example, at step 1 in (16) a violation of *COMPLEX-ONSET is eliminated at the expense of introducing violations of lower-ranking ONSET and DEP. At step 2, the ONSET violation is removed at the expense of introducing another DEP violation. At step 3, there is no further harmonic improvement, so the derivation has converged and terminates.

(16) Harmonic improvement tableau for /fʕal/ → *ʔif.ʕal*

	*COMP-ONSET	MAX	CONTIG	ONS	DEP	
Faithful fʕal	1					
Step 1 if.ʕal				1	1	
Step 2 ʔif.ʕal					1	
Step 3 ʔif.ʕal						*convergence*

The harmonic improvement tableau is not a substitute for regular tableaux, because it does not show that the change made at each step is the best change. But the harmonic improvement tableau is useful for showing that harmony does indeed improve at each step of the HS derivation until convergence.

Because HS derivations must show steady harmonic improvement, they are guaranteed to converge in a finite number of steps. Under the standard OT assumption that all constraints either evaluate outputs (markedness) or require input-output identity (faithfulness), harmony cannot improve without limit (Moreton 2000, 2003). This result presupposes, of course, that the harmony criterion — that is, the ranking — does not change from one step to the next. If ranking can vary between steps, then eventual convergence is highly probable but not assured (Kimper 2011, Staubs and Pater this volume).

To see this more intuitively, it may be helpful to think about what it would take to produce a HS derivation that never converges. One imaginable way of getting a non-convergent derivation is to add structure without limit: /ta/ → *ta.ə* → *ta.ə.ə* → …. Another is to have a loop in the derivation, where a change is made and then immediately undone: /ta/ → *te* → *ta* → …. Neither is in fact possible in HS under standard assumptions about constraints and the constraint hierarchy.

For structure to be added without limit, there must be some markedness constraint that always favors a form with more structure over a form with less. Although there are certainly markedness constraints that favor having more structure under some circumstances — ONSET is an example — there is not and cannot be any markedness constraint that favors having more structure under *all* circumstances, regardless of the length of the input. Anti-faithfulness constraints have the capacity to force the addition of structure to any input (cf. Alderete 1998, 2001), but they lie well outside the standard assumptions about what constraints can do. More to the point, it is clear that anti-faithfulness constraints are incompatible with HS precisely because they undermine the convergence guarantee. Positive constraints also have this capacity, under certain assumptions — see Kimper (this volume) for discussion.

For a HS derivation to have a loop like /ta/ → *te* → *ta* → …, it must be the case that *ta* is more harmonic than *te*, and *te* is more harmonic than *ta*, according to the same constraint hierarchy. In a standard OT constraint hierarchy consisting only of markedness and faithfulness constraints, this is not possible. For the mapping /ta/ → *te* to occur, *te* must be less marked than the more faithful *ta*. And for the mapping *te* → *ta* to occur, *ta* must be less marked than the more faithful *te*. There is an obvious contradiction here, so the loop is an impossibility in HS. (Again, anti-faithfulness constraints would subvert this result.)

When a process is equally applicable in more than one location, HS derivations may exhibit a kind of arbitrariness that might at first seem disconcerting. Because HS's GEN is gradual, the process can only apply once in each step, though eventually it will apply everywhere. For example, Cairene Arabic has a process that shortens long vowels in unstressed syllables:

(17) Unstressed syllable shortening in Cairene

/itnaːʔiʃ-na/	ʔitnaʔíʃ-na	'we discussed'
/ʃaːf-it-ak/	ʃafítak	'she saw you (m. sg.)'
/ma-ʃaːf-uː-niː-ʃ/	maʃafuníːʃ	'they didn't see me'

As the last example shows, when there are several unstressed long vowels, all of them shorten.

I will ignore the part of the grammar that assigns stress and focus exclusively on the shortening process. Vowel shortening is a consequence of satisfying a constraint called WEIGHT-TO-STRESS (abbreviated WSP), which is violated by unstressed heavy syllables (Prince 1990). WSP dominates MAX-μ in Cairene:

(18) WSP dominates MAX-μ (P-OT analysis)

/ma-ʃaːf-uː-niː-ʃ/	WSP	MAX-μ
a. → maʃafuníːʃ		2
b. maʃaːfuːníːʃ	2 W	L
c. maʃaːfuníːʃ	1 W	1 L
d. maʃafuːníːʃ	1 W	1 L

Now suppose we are analyzing the same data in HS. We will enter the derivation at step 2, after stress has been assigned. There are three relevant candidates from the input *maʃaːfuːníːʃ*: the unchanged input itself, *maʃafuːníːʃ*, and *maʃaːfuníːʃ*. The actual surface form is not a candidate, because it is two changes away from the current input. The candidates that have shortened a vowel are tied as the intermediate winners:

(19) Step 2 of HS analysis

maʃaːfuːníːʃ	WSP	MAX-μ
a. → maʃafuːníːʃ	1	1
b. → maʃaːfuníːʃ	1	1
c. maʃaːfuːníːʃ	2 W	L

Choose either one of the intermediate winners and input it to GEN. The winning candidate is one where both vowels have shortened, as desired:

(20) Step 3 with *maʃafuːníːʃ* as intermediate winner

	maʃafuːníːʃ	WSP	MAX-μ
a. →	maʃafuníːʃ		1
b.	maʃafuːníːʃ	1 W	L

(21) Step 3 with *maʃaːfuníːʃ* as intermediate winner

	maʃaːfuníːʃ	WSP	MAX-μ
a. →	maʃafuníːʃ		1
b.	maʃaːfuníːʃ	1 W	L

Intermediate *maʃafuːníːʃ* and *maʃaːfuníːʃ* are said to be in a *convergent tie*: they tie at an intermediate step of the derivation, and the ultimate result is the same either way. For further information about ties in HS, see Mullin et al. (2010: 7-10).

This example can also be used to illustrate one final point about HS derivations: the intermediate winner may be a form that is harmonically bounded in the P-OT analysis. Observe from (18) that *maʃafuːníːʃ* and *maʃaːfuníːʃ* are both harmonically bounded in P-OT under this constraint set. Yet both are intermediate winners in HS. This situation is by no means unexpected. The intermediate steps of HS derivations are local optima: they are the most harmonic member of HS's limited candidate set. HS has a different sense of harmonic bounding: *maʃafuːníːʃ* and *maʃaːfuníːʃ* are harmonically bounded because the derivation will not converge on them under either ranking of the two constraints. For further discussion of local and global optima, see section 7.1.

5 Constraints in Harmonic Serialism

HS permits a rather different conception of faithfulness constraints than is standard in P-OT. Because the input and the output candidates can differ in unlimited ways in P-OT, Correspondence Theory was introduced to keep track of these differences and to assign faithfulness violations correctly (McCarthy and Prince 1995, 1999). In Correspondence Theory, the candidate set for input *in* consists, at a minimum, of all the strings over some alphabet and all of the ways of coindexing each of those strings with *in*. Thus, $d_1a_2g_3$ is a member of the candidate set for /$k_1æ_2t_3$/, as are $d_1a_3g_2$, d_1a_2g, and so on. Correspondence Theory's faithfulness constraints are defined in terms of these indices: the *d* of d_1a_2g is coindexed with /k/, so it violates IDENT(place) and IDENT(voice); the *g* of d_1a_2g has no index, so it violates DEP, and so on.

Although HS retains Correspondence Theory's *names* for the faithfulness constraints, the *definitions* of these constraints are quite different. For example, under a particular conception of HS's GEN, one of the candidates from input /kæt/ is *kæ*, which was produced from the input by applying a deletion operation. It is the application of this deletion operation that incurs the violation of MAX. This is how faithfulness is implemented in the OT-Help program (Staubs et al. 2010): each faithfulness constraint is associated with some operation(s) in GEN. Applying an operation incurs violations of its associated faithfulness constraint(s). Because faithfulness constraints look back only as far as the input to the most recent application of GEN, and not to the underlying representation,[7] the reasons why Correspondence Theory is required in P-OT do not carry over to HS.

All well-established faithfulness constraints are amenable to being redefined operationally. Violations of DEP and MAX are keyed to segment insertion and deletion operations. Applying insertion or deletion medially violates CONTIGUITY; applying these options peripherally violates ANCHOR. An operation that transposes adjacent segments is responsible for LINEARITY violations (McCarthy 2007b). The operations on distinctive features may involve changing feature values, thereby violating IDENT, or deleting and inserting featural elements, thereby violating DEP(feature) and MAX(feature) (see McCarthy 2008b for discussion).[8] This operational theory of faithfulness has consequences for the formulation of positional faithfulness constraints (see section 7). Together with gradualness, it also limits the usefulness of local conjunction of faithfulness constraints (Kirchner 1996, Moreton and Smolensky 2002). See section 8 for further discussion.

HS also has some consequences for markedness constraints; see Pruitt (2012) on *LAPSE and Pater (2012) on Prince and Smolensky's (1993/2004) scalar sonority constraint HNUC.

6 GEN in Harmonic Serialism[9]

Because of the gradualness requirement on GEN, HS's candidate sets are finite. Finiteness is assured as long as GEN contains no intrinsically iterative or recursive operations — that is, no operations that can add unbounded amounts of structure in a single step. This means that the effects of iteration or recursion have to be obtained over multiple steps of a derivation rather than all at once.

HS's GEN is gradual, but what exactly is meant by "gradualness"? This is a central research question in HS. The claim is that GEN can make only one change at a time, so the candidate set from input *in* consists of *in* itself and all the forms differing from *in* by a single change. There are, however, important issues in specifying exactly what constitutes a single change. Fortunately, it is clear how we should go about finding an answer to this question, although the answer itself is not yet complete. The technique can be explained with an example.

In Cairene Arabic, syncope affects short high vowels in unstressed non-final open syllables in internal and external sandhi (see (22)). The consonant that preceded the deleted vowel is resyllabified as a coda:

(22) Cairene syncope (data from Watson 2002: 70-72)

/wiħiʃ-a/	ˈwiħ.ʃa	'bad (f. sg.)'
/xulusˤ-it/	ˈxul.sˤit	'she finished'
/tˤardi kibiːr/	ˈtˤar.dik.ˈbiːr	'my parcel is big'

Do syncope and resyllabification occur in separate steps? Is GEN able to resyllabify an onset at the same time that it deletes its nucleus? Which of the derivations in (23) is the right one?[10]

(23) Two possible derivations for /wiħiʃ-a/ → *ˈwiħ.ʃa*

a. /wiħiʃ-a/ → wi.ħi.ʃa → ˈwi.ħi.ʃa → ˈwi.*ħ*.ʃa → ˈwiħ.ʃa

b. a. /wiħiʃ-a/ → wi.ħi.ʃa → ˈwi.ħi.ʃa → ˈwiħ.ʃa

Both derivations begin with initial syllabification followed by stress assignment. They then diverge, because each presupposes a different theory of GEN. The GEN that produces (23a) does syncope and resyllabification serially, in separate steps. The GEN that produces (23b) does syncope and resyllabification in parallel, in a single step.

I will argue that syncope and resyllabification must be simultaneous, as in (23)b, by showing that the sequential derivation in (23a) is inconsistent with other facts of the language. The problem is that syncope is blocked after a consonant cluster: /ħagar kibiːr/ → *ˈħa.gar.ki.ˈbiːr* 'big stone'. The standard explanation for why syncope is blocked in this situation is that it would leave the *k* unsyllabified, because there is no way to adjoin it to the preceding or following syllable without creating a forbidden phrase-internal complex coda or complex onset: **ˈħa.gark.ˈbiːr*, **ˈħa.gar.ˈkbiːr*. The sequential theory is unable to express this insight, however. In the sequential theory, the derivation reaches a point where the choice is between *ħa.gar.ki.ˈbiːr* with no syncope and **ħa.gar.k.ˈbiːr* with syncope. To block syncope as required, the latter candidate needs to be ruled out, and a constraint against unsyllabified consonants could do that. But this same constraint would also wrongly block syncope in (23a), where the derivation has to pass through a stage with an unsyllabified consonant.

This particular serial analysis has failed because HS has no ability to look ahead. The decision about whether or not to syncopate has to be made based on the conditions obtaining at that point in the derivation. Thus, in the sequential theory, there is no way of knowing whether it will eventually be possible to attach the unsyllabified consonant to a nearby syllable. The simultaneous theory does not have this problem because it does resyllabification in parallel with syncope. It can therefore use the markedness constraint against unsyllabified consonants to rule out **ħa.gar.k.ˈbiːr* without compromising the derivation of *wiħ.ʃa*.

This example illustrates a general method for deciding what GEN can and cannot do in a single step. GEN determines how much and what kind of information is available to EVAL at each step of the derivation. Since there is no look-ahead, all of the information necessary to determine whether the right candidate wins has to be available at the point where it is crucial for that candidate to win. In the case of Cairene, if a particular theory of GEN segregates syncope and syncope-triggered resyllabification into different steps, then information about the ultimate consequences of syncope for syllabification is unavailable at the point in the derivation where the syncope decision must be made. As a result, under this version of GEN it is not possible to capture a very familiar phonological generalization: syncope occurs unless it would leave an unsyllabifiable consonant (cf. Kisseberth 1970). As we saw, that generalization is attainable under a different theory of GEN in which syncope and resyllabification can be simultaneous. It is attainable because the consequences of syncope for syllabification are known at the derivational step where the syncope choice has to be made.

It is important to realize that arguments of this type depend on the details of the theory of CON as well as GEN. For example, if there were a constraint that specifically prohibited an unsyllabified consonant after a closed syllable, then the sequential theory would be viable, because now there would be a constraint that correctly allows *ˈwi.ħ.ʃa* and correctly blocks **ħa.gar.k.ˈbiːr*. On the trade-offs between GEN and CON and how to resolve them, see section 7 of McCarthy, Pater, and Pruitt (this volume).

Although we have been focusing here on understanding a technique for studying GEN rather than drawing conclusions about GEN itself, we should not ignore the broader consequences of the argument presented above. Why is it possible to perform syncope and resyllabification in a single pass through GEN?

In McCarthy (2007a), I propose that GEN is limited to a single unfaithful operation at a time, but there is no limit on faithful operations. Syncope, epenthesis, feature change, and so on are unfaithful operations, so each of them requires a separate derivational step. But (re)syllabification is a faithful operation. It is therefore possible to combine syncope and (re)syllabification into a single derivational step.[11] This hypothesis about GEN can be tested by applying these techniques to other presumptively faithful operations, such as adjunction of an unstressed syllable to a foot, pruning of an empty node, or parsing a lexical word into a prosodic word.

7 Arguments for Harmonic Serialism[12]

Many of the arguments for HS over P-OT can be divided into two groups. Some arguments are based on the fact that HS, but not P-OT, has forms that are intermediate in a derivation. Others are based on differences between HS and P-OT in the typologies that they generate for a given constraint set. We will look at arguments of both types.

7.1. Evidence for HS from intermediate forms

The existence of intermediate derivational steps is one of the two characteristics of HS that distinguish it from P-OT. In this section, I briefly summarize two arguments for HS that are based on this difference. Several others are noted at the end of the section. These arguments share a common premise: certain generalizations cannot be expressed in underlying or surface representation, but those are the only two levels of representation that P-OT has. These generalizations are expressible in HS's intermediate representations, however.

The first argument comes from McCarthy (2008c). In many languages, some or all unstressed vowels delete. This simple generalization proves to be difficult to express in a P-OT analysis. The problem is that the generalization is inherently derivational: stress is assigned and then unstressed vowels are deleted. A P-OT grammar must optimize the effects of stress assignment and syncope simultaneously, and this turns out to be inadequate both descriptively and typologically.

In Macushi Carib (Hawkins 1950: 87), for example, words are parsed into iambic feet from left to right, and only then are unstressed vowels deleted:[13]

(24) Stress-syncope interaction in Macushi Carib

Underlying	Stress	Syncope	
piripi	(pirí)(pí)	(prí)(pí)	'spindle'
wanamari	(waná)(marí)	(wná)(mrí)	'mirror'
u-manari-rɨ	(umá)(narí)(rɨ́)	(má)(nrí)(rɨ́)	'my cassava grater'
u-wanamari-rɨ	(uwá)(namá)(rirɨ́)	(wá)(nmá)(rrɨ́)	'my mirror'

In HS, these are exactly the steps that the derivation follows. Because of gradualness, stress assignment and syncope cannot occur simultaneously. Stress assignment occurs first because syncope is *intrinsically ordered* after stress. Two processes are said to be intrinsically ordered if the applicability of one depends on the prior application of the other. In HS, this occurs when the markedness constraint implicated in the second process is not violated until the first process has applied. In the case of syncope and stress, the markedness constraint that is responsible for syncope, *V-PLACE$_{weak}$, is violated by a vowel in the weak syllable of a foot. (In other words, vowel place features are not licensed in this weak position.) Before foot structure has been assigned, all vowels vacuously satisfy this constraint. Therefore, *V-PLACE$_{weak}$ is not active until stress has been assigned, so stress is intrinsically ordered before syncope.

Let us assume that GEN includes operations that create a foot, remove a foot, or delete a vowel. This assumption about gradualness means that the candidate set at step 1, shown in tableau (25), includes candidates with foot parsing or syncope but not both. The imperative to parse into feet is provided by the constraint PARSE-SYLLABLE, which is violated by any unfooted syllable. (Prior or simultaneous syllabification may be assumed; see section 6.)

(25) Step 1 of /wanamari/ → *(wná)(mrí)*

/wanamari/	PARSE-SYLL	*V-PL$_{weak}$	MAX
a. → (waná)mari	2	1	
b. wanamari	4 W	L	
c. wanmari	3 W	L	1 W

Tableau (25) shows how stress assignment is intrinsically ordered before syncope. Syncope prior to foot parsing, as in (25c), improves performance on PARSE-SYLLABLE, but less so than foot parsing in (25a). Foot parsing introduces a violation of *V-PLACE$_{weak}$, as in (25a), because *V-PLACE$_{weak}$ is vacuously satisfied by unfooted vowels. For that reason, PARSE-SYLLABLE must be ranked higher than *V-PLACE$_{weak}$.[14]

At step 2, satisfaction of PARSE-SYLLABLE is still the prime directive, so syncope is once again postponed:

(26) Step 2 of /wanamari/ → *(wná)(mrí)*

/wanamari/	PARSE-SYLL	*V-PL$_{weak}$	MAX
a. → (waná)(marí)		2	
b. (waná)mari	2 W	1 L	
c. (wná)mari	2 W	L	1 W
d. (waná)mri	1 W	1 L	1 W

In a quadrisyllable like /wanamari/, full satisfaction of PARSE-SYLLABLE has been achieved by the end of step 2, so it is finally possible to attend to the requirements of *V-PLACE$_{weak}$, which is the next markedness constraint in the ranking. One of the unstressed, footed vowels deletes at step 3, with the other deleting at step 4 (not shown).

(27) Step 3 of /wanamari/ → *(wná)(mrí)*

(waná)(marí)	PARSE-SYLL	*V-PL$_{weak}$	MAX
a. → (wná)(marí)		1	1
b. (waná)(marí)		2 W	L

Which vowel deletes first is unimportant, since ultimately both delete. As it happens, the constraint responsible for left-to-right foot parsing, ALIGN-L(foot, word), also favors deleting from left to right — hence *(wná)(marí)* rather than *(waná)(mrí)* is shown as the winner in (27).

Finally, the derivation converges at step 5, with input and winner identical to one another. Alternatives to the intended winner, such as those in (28b) and (28c), reintroduce violations of the top-ranked markedness constraints or violate faithfulness constraints gratuitously:

(28) Step 5 of /wanamari/ → *(wná)(mrí)* — Convergence

	(wná)(mrí)	PARSE-SYLL	*V-PL$_{weak}$	MAX
a. →	(wná)(mrí)			
b.	wna(mrí)	1 W		
c.	(wnmrí)			1 W

This analysis shows that HS offers a viable approach to stress-syncope interactions. P-OT does not. The problem is that P-OT lacks the intermediate representation in which stress has been assigned prior to syncope. The P-OT analysis must therefore distinguish the intended winners from losing candidates that never even arise in the HS analysis, and this proves to be impossible. Here are some examples:

(29) Problematic losers in P-OT

Underlying	Intended winner	Problematic losers
piripi	(prí)(pí)	*(pí)(rpí)
wanamari	(wná)(mrí)	*(wá)(nmá)(rí)
u-manari-rɨ	(má)(nrí)(rɨ́)	*(ú)(mná)(rrɨ́)
u-wanamari-rɨ	(wá)(nmá)(rrɨ́)	*(ú)(wná)(mrí)(rɨ́)

The intended winners respect the generalization that syncope affects the odd-numbered non-final syllables — i.e., exactly the syllables that are left unstressed after the left-to-right iambic parse. The problematic losers follow the generalization that syncope affects the *even*-numbered non-final syllables — i.e., exactly the syllables that would be left unstressed by a left-to-right *trochaic* parse. The problem for P-OT is that no markedness constraint evaluating surface forms can systematically distinguish the two patterns of syncope. The reason for this failure is that P-OT has only two levels of representation to work with, underlying and surface, but capturing the generalization about which vowels are targeted for syncope requires an intermediate level, post-stress and pre-syncope.

Another argument for HS's intermediate levels of representation comes from Jesney (2011). Positional faithfulness constraints are like other faithfulness constraints except that their scope of action is limited to certain prominent positions, such as stressed syllables (Beckman 1998). For example, the positional faithfulness constraint IDENT$_{stress}$(nasal) is protective of nasalization contrasts in stressed syllables. When ranked above *V$_{nasal}$, which itself dominates the position-insensitive faithfulness constraint IDENT(nasal), the result is a language like Nancowry (Radhakrishnan 1981), where phonemic vowel nasalization is maintained in stressed syllables but neutralized in unstressed ones. In the following schematized example, stress is assumed to be trochaic, so TROCHEE is undominated:

(30) Attested positional faithfulness effect (P-OT)

	bãdõ	ID$_{str}$(nas)	PARSE-SYLL	TROCH	*V$_{nas}$	ID(nas)
a. →	(bã́do)				1	1
b.	(bádo)	1 W			L	2 W
c.	(bã́dõ)				2 W	L
d.	(badṍ)			1 W	1	1
e.	bado		2 W		L	2 W

Because of *V$_{nasal}$, nasalized vowels are neutralized to oral in unstressed syllables, as in (30a). But there is no neutralization in stressed syllables (cf. (30b)), because of IDENT$_{stress}$(nasal).

P-OT's problem, which was first recognized by Rolf Noyer (cited in Beckman (1998: fn. 37)), is that positional faithfulness constraints work as intended only when the position of greater faithfulness is held constant in those candidates where the positional faithfulness constraint is making a crucial comparison. That is certainly true in (30): the surface reflex of /ã/ is stressed in both (30a) and (30b). Candidates that are stressed differently or not at all, such as (30d) and (30e), are ruled out by other constraints, so they do not depend on IDENT$_{stress}$(nasal) to exclude them.

Now consider what happens when stress is allowed to differ among the viable candidates. In (31), TROCHEE is ranked below *V$_{nasal}$. The result is that stress is shifted from an underlying nasalized vowel onto an underlying oral one. This happens because the positional faithfulness constraint is crucially comparing two candidates, (31a) and (31b), that differ in stress:

(31) Unattested positional faithfulness effect (P-OT)

	/pãko/	ID$_{str}$(nas)	PARSE-SYLL	*V$_{nas}$	ID(nas)	TROCH
a. →	(pakó)				1	1
b.	(páko)	1 W			1	L
c.	(pã́ko)			1 W	L	L
d.	pako		2 W		1	L

When this same grammar is presented with any other combination of nasalized and oral vowels (i.e., /bãdõ/, /sato/, or /kafõ/), it defaults to trochaic stress. Thus, in this hypothetical language, stress is normally on the penult, but it is on the ultima when the penult vowel is underlying nasal and the final vowel is underlying oral, though both end up oral at the surface. No real language does anything remotely like this.

What is the source of this problem? Positional faithfulness constraints are sensitive to structure that is assigned by the grammar, such as stress. Since the surface form is the only grammar-derived level of representation in P-OT, P-OT's positional faithfulness constraints have to be defined like this: "If a segment in the surface representation is in a stressed syllable, it must be faithful to its underlying correspondent". When positional faithfulness constraints are defined in this way, the problem in (31) is unavoidable.

As Jesney shows, this otherwise intractable problem is solved if HS is adopted and if positional faithfulness constraints are defined to refer to the prosodic structure of the *input*: if a segment in the input to GEN is in a stressed syllable, it must be faithful in [nasal]. In HS, the input to GEN is not necessarily the underlying representation, so it can have structure that has been assigned by the grammar. Since the input is the same for all candidates being compared, problems like (31) cannot arise.

The HS derivation of /pãko/ proceeds as follows. At step 1, there is a choice between assigning stress or denasalizing *ã*. If $*V_{nasal}$ dominates PARSE-SYLLABLE, then denasalization takes precedence, and we have a language without nasalized vowels and without a positional faithfulness effect. If PARSE-SYLLABLE is ranked higher, as in tableau (32), then stress is assigned first. Stress (re)assignment and denasalization cannot cooccur, of course, because of gradualness.

(32) Step 1 from /pãko/

/pãko/	ID_{str}(nas)	PARSE-SYLL	$*V_{nas}$	ID(nas)	TROCH
a. → (pã́ko)			1		
b. pako		2 W	L	1 W	
c. (pãkó)			1		1 W

The derivation then converges at step 2, shown in (33). Input *(pã́ko)* has a stressed nasalized vowel. Since this vowel is stressed in the input to this pass through the GEN-EVAL loop, redefined $IDENT_{stress}$(nasal) protects it from denasalization:

(33) Step 2 from /pãko/

(pã́ko)	ID_{str}(nas)	PARSE-SYLL	$*V_{nas}$	ID(nas)	TROCH
a. → (pã́ko)			1		
b. (páko)	1 W		L	1 W	
c. (pãkó)			1		1 W

The failure of the final-stressed candidate *(pãkó)* in (32c) and (33c) is crucial to this argument for HS. If this candidate were to survive, it would change into *(pakó)* at the next step of the derivation, and HS would be making the same bad prediction as P-OT. In fact, it does not survive. The reasoning goes like this:

> The P-OT tableau (31) shows that *(pakó)* is the *global optimum* for underlying /pãko/ under this grammar. Another way of saying the same thing is that *(pakó)* is the *global minimum of potential for harmonic improvement*. The derivation (32)–(33) shows that this global optimum is inaccessible in HS because there is no harmonically improving path to it. The HS derivation gets stuck at a *local optimum*, *(pã́ko)*. Equivalently, *(pã́ko)* is a *local minimum of*

potential for harmonic improvement. This happens because HS's GEN cannot simultaneously assign final stress and denasalize the penult, and final stress is not harmonically improving unless the penult is simultaneously denasalized.

The overall lesson here is that HS has no look-ahead capability; a candidate that fails to improve harmony at step *n* cannot win simply because it would lead to greater harmonic improvement at step *n + 1*. "Getting stuck" at a local optimum sounds like a bad outcome, but it can actually be a good one. Positional faithfulness and many other typological results of HS depend on this property of the theory.

The examples of stress-syncope interaction and positional faithfulness show that HS's intermediate levels of representation are necessary to capture some basic generalizations about phonology. Staubs (this volume) shows how HS solves a problem with positional markedness that is similar to the positional faithfulness problem just discussed. Other arguments for HS's intermediate levels are based on discussions of opacity (Elfner this volume), the interaction of stress and syllabification (Elsman this volume), compensatory lengthening (Torres-Tamarit this volume), phonetically grounded constraints (McCarthy 2011), scalar constraints (Pater 2012), and local variation (Kimper 2011).

7.2. Evidence for HS from language typology

Language typology is a central concern of research in OT. Because the same constraints can be ranked differently in different languages, any proposed constraint system constitutes an implicit claim about the range of permissible variation among languages. The logic of language typology in HS is explained in McCarthy (2007b, 2009) and summarized here.

For identical constraint systems, P-OT and HS may predict different typologies. The source of the difference is HS's core properties, gradualness and harmonic improvement. Does a given constraint system CON yield a language in which underlying /A/ maps to surface *B*? In P-OT, the answer is yes if and only if there is some ranking of CON where *B* is more harmonic than *A* and every other candidate derived from /A/. In HS, this answer is sufficient only if *B* and /A/ differ by a single change. If it requires more than one change to get from /A/ to *B*, then there must also be a harmonically improving path of winning intermediate steps from /A/ to *B*. Sometimes, there is no such path. That is when P-OT and HS make different typological predictions.

This reasoning is important in HS's solution to some *too many repairs* (TMR) problems. A TMR problem is the observation that the actually attested ways of satisfying a markedness constraint are often more limited than we would expect from ranking permutation (Blumenfeld 2006, Lombardi 1995/2001, Pater 1999, Steriade 2001/2008, Wilson 2001, and others). For example, the markedness constraint CODA-COND says that coda consonants do not license place of articulation (Goldsmith 1990: 123-128, Ito 1989). One way of satisfying this constraint is for a coda to share place with a following onset, since onset position does license place. This is the reason why place often assimilates in consonant clusters: in *pamta*, labial place is unlicensed in coda *m*, but in *pan͡ta* the *n*'s coronal place is licensed because it is shared with the onset (as indicated by the ligature). Unexplained is why place always assimilates from the onset to the coda and never the other way around: /pamta/ → *pan͡ta*, never *pam͡pa*. Tableau (34) illustrates the problem:

(34) /pamta/ → *pan͡ta*/*pam͡pa* in P-OT

/pamta/	Coda-Cond	Id(place)
a. → pan͡ta		1
b. → pam͡pa		1
c. pamta	1 W	L

P-OT predicts intra- or interlinguistic variation in direction of assimilation when violations of Coda-Cond are repaired, but the predicted variation is not usually observed.

HS offers an explanation for this asymmetry, once the process of place assimilation is properly understood in operational terms (McCarthy 2008b). Long before HS or even OT, it was claimed that place assimilation is a two-step process, with deletion of the unlicensed place feature prior to spreading of the licensed one (Cho 1990, Kiparsky 1993, Mascaró 1987, Poser 1982): /pamta/ → *paNta* → *pan͡ta*. (*N* denotes a placeless nasal.) If HS's Gen is restricted in this fashion, then the directional asymmetry in place assimilation follows automatically. At step 1, deletion of place from the coda consonant satisfies Coda-Cond, but deletion of place from the onset (yielding a placeless *ʔ*) does not:

(35) Step 1 of /pamta/ → *pan͡ta*

/pamta/	Coda-Cond	Have-Pl	Id(place)
a. → paNta		1	1
b. pamta	1 W	L	L
c. pamʔa	1 W	1	1

At step 2, placeless *N* becomes *n* by spreading place from the following *t*. This occurs to satisfy Have-Place, which *N* violates:

(36) Step 2 of /pamta/ → *pan͡ta*

paNta	Coda-Cond	Have-Pl	Id(place)
a. → pan͡ta			1
b. paNta		1 W	L

The derivation then converges at step 3 (not shown).

This example illustrates the point about harmonic improvement that was made earlier in this section. In P-OT, *pan͡ta* and *pam͡pa* are both possible surface results from underlying /pamta/, since both satisfy Coda-Cond and violate Ident(place) equally. In HS, though, it is not enough for a surface form to be a P-OT winner; it must also be linked with the underlying form by a chain of harmonically improving intermediate forms. That is not the case with *pam͡pa*; under the stated assumption about Gen, it requires an intermediate form, *pamʔa*, that does not improve harmony relative to Coda-Cond, as (35) shows. Because of gradualness and harmonic improvement, HS yields a more restrictive typology of place assimilation than P-OT does, all else being equal. This more restrictive typology better fits what we actually find in languages.

Another area where HS appears to have a typological advantage over P-OT is in accounting for locality effects. For example, Pruitt (2010) shows that metrical foot parsing exhibits locality effects that are hard to account for in P-OT but follow readily in HS from the assumption that GEN builds feet one at a time.

One such locality effect involves the interaction between foot parsing and vowel shortening. In quantity-sensitive languages, trochaic feet are usually limited to a pair of light syllables (ĹL) or a single heavy syllable (H́) (Hayes 1985, 1995, McCarthy and Prince 1986/1996, Prince 1990). (H́L) trochees are disfavored by a constraint called dubbed RHHRM (for rhythmic harmony) by Prince & Smolensky (1993/2004: 70-71). In the following discussion, I will assume a language with left-to-right quantity-sensitive trochees.

The standard ranking for left-to-right foot parsing uses the constraint ALIGN-L(foot, word) (McCarthy and Prince 1993)to favor having all feet as far to the left as possible:

(37) Left-to-right parsing (P-OT)

salamataka	FOOT-BIN	PARSE-SYLL	AL-L (ft, wd)	AL-R (ft, wd)
a. → (sála)(máta)ka		1	2	4
b. salamataka		5 W	L	L
c. sa(láma)(táka)		1	4 W	2 L
d. (sála)(máta)(ká)	1 W	L	6 W	4

If we include RHHRM at the top of the hierarchy and allow shortening of long vowels by ranking MAX-μ low, we get a language in which a long vowel in the first syllable is shortened only if it is followed by an odd number of light syllables. Compare (38) with (39):

(38) Shortening before odd L sequence (P-OT)

pa:takasa	FOOT-BIN	RHHRM	PARSE-SYLL	AL-L (ft, wd)	MAX-μ
a. → (páta)(kása)				2	1
b. (pá:)(táka)sa			1 W	1 L	L
c. (pá:ta)(kása)		1 W		2	L

(39) No shortening before even L sequence (P-OT)

pa:takasafa	FOOT-BIN	RHHRM	PARSE-SYLL	AL-L (ft, wd)	MAX-μ
a. → (pá:)(táka)(sáfa)				4	
b. (páta)(kása)fa			1 W	2 L	1 W
c. (pá:ta)(kása)fa		1 W	1 W	2 L	

No known language behaves in this highly non-local fashion, where avoiding an unfooted syllable at one end of the word — i.e., final *sa* in *(pá:)(táka)sa* (38b) — triggers vowel shortening at the other end of the word. Optimization of metrical structure can cause vowels to shorten, as it does in Latin (Allen 1973, Mester 1994, Prince and Smolensky 1993/2004, McCarthy, Pater, and Pruitt this volume) or Fijian (Dixon 1988, Hayes 1995, Schütz 1985), but these effects are always strictly local, involving a pair of adjacent syllables.

This example reflects a more general problem with P-OT, highlighted in McCarthy (2007b, 2008c, 2010a) and Pruitt (2010) and already encountered in section 7.1: it has excessive power to do global optimization. The reason why (37)–(39) is predicted to be a possible language in P-OT is that P-OT's GEN builds complete and final surface candidates in which the effects of vowel shortening and full metrical parsing are evaluated together. Thus, EVAL gets to choose the best combination of shortening and parsing, no matter how distant the long vowel might be from the parsing problem.

In HS, if GEN is limited to building one foot at a time, then the language in (37)–(39) cannot be obtained with these constraints, as Pruitt (2010) demonstrates. We will first consider how iterative parsing works in HS with a word that contains no heavy syllables. At step 1 (tableau (40)), the best option is to build a disyllabic foot at the left edge of the word. Building no foot or a monosyllabic foot is disfavored by PARSE-SYLLABLE; the latter also violates FOOT-BINARITY. Building a foot non-initially violates ALIGN-L(foot, word):

(40) Step 1 of iterative parse

/salamataka/	FOOT-BIN	PARSE-SYLL	AL-L (ft, wd)	AL-R (ft, wd)
a. → (sála)mataka		3		3
b. salamataka		5 W		L
c. sa(láma)taka		3	1 W	2 L
d. salama(táka)		3	3 W	L

At step 2 (tableau (41)), the best option is to build a disyllabic foot as far to the left as possible. PARSE-SYLLABLE requires construction of an additional foot, and ALIGN-L determines where it is built. After this, the derivation converges, as tableau (42) shows. The only remaining unfooted syllable is the last one, and FOOT-BINARITY ensures that nothing can be done about it.

(41) Step 2 of iterative parse

(sála)mataka	FOOT-BIN	PARSE-SYLL	AL-L (ft, wd)	AL-R (ft, wd)
a. → (sála)(máta)ka		1	2	4
b. (sála)mataka		3 W	L	3 L
c. (sála)ma(táka)		1	3 W	3 L

(42) Step 3 of iterative parse (convergence)

	(sála)(máta)ka	FOOT-BIN	PARSE-SYLL	AL-L (ft, wd)	AL-R (ft, wd)
a. →	(sála)(máta)ka		1	2	4
b.	(sála)(máta)(ká)	1 W	L	6 W	4

What happens when the underlying representation contains an initial long vowel? If we apply the same ranking as (38) and (39), but within the HS architecture, the result does not depend on whether the long vowel is followed by an odd number (43) or even number (44) of light syllables: the leftmost pair of light syllables is parsed into a foot, since this option best satisfies RHHRM and PARSE-SYLLABLE:

(43) Step 1 from /pa:takasa/ (odd sequence of Ls)

	/pa:takasa/	FOOT-BIN	RHHRM	PARSE-SYLL	AL-L (ft, wd)	MAX-μ
a. →	pa:(táka)sa			2	1	
b.	pa:takasa			4 W	L	
c.	(pá:)takasa			3 W	L	
d.	(pá:ta)kasa		1 W	2	L	
e.	patakasa			4 W	L	1 W

(44) Step 1 from /pa:takasafa/ (even sequence of Ls)

	/pa:takasafa/	FOOT-BIN	RHHRM	PARSE-SYLL	AL-L (ft, wd)	MAX-μ
a. →	pa:(táka)safa			3	1	
b.	pa:takasafa			5 W	L	
c.	(pá:)takasafa			4 W	L	
d.	(pá:ta)kasafa		1 W	3	L	
e.	patakasafa			5 W	L	1 W

These derivations continue, parsing pairs of light syllables from left to right, and then returning to parse the initial heavy syllable into a foot of its own. The derivations converge on *(pá:)(táka)sa* and *(pá:)(táka)(sáfa)*. With this ranking, there is no shortening, and there is certainly no dependency of shortening on whether an odd or even number of light syllables follow.

There will be shortening if the ranking of RHHRM and ALIGN-L(foot, word) is reversed, but still there is no dependency of shortening on the following syllables. At step 1, the first two syllables are parsed into a RHHRM-violating foot: *(pá:ta)kasa*, *(pá:ta)kasafa*. Foot parsing continues at step 2, yielding *(pá:ta)(kása)* and *(pá:ta)(kása)fa*. At this point, PARSE-SYLLABLE is as well satisfied as it can get, and RHHRM gets its chance to compel shortening of the long vowel (see McCarthy, Pater, & Pruitt this volume), but both the odd and even length words are affected.

This is another clear point of difference between P-OT and HS. Because P-OT optimizes globally, it allows long-distance dependencies like the one in (37)–(39). HS does not permit this dependency, at least with the standard constraints and the unremarkable version of GEN that we have been assuming. HS is more limited in this respect because decisions about foot parsing and shortening are made one at a time.

This example hints at an important connection between locality and serialism. Long-distance effects are often produced by iterative application of a process; the construction of metrical feet is one example, and successive cyclic *wh*-movement (Chomsky 1977) is another. In P-OT, process iteration is invisible to EVAL because it takes place entirely in GEN. EVAL sees only the final result. In HS, however, process iteration is visible to EVAL because the results of each iteration are presented in the candidate set. This enforces a kind of locality because each iteration must improve harmony or the process terminates (McCarthy 2007b, 2008c, 2010a, Pruitt 2010). The claim implicit in HS (and in successive cyclic *wh*-movement) is that visible iteration and its concomitant locality effects are a better theory of language typology than the global alternative.

Although the logic of typological research in P-OT and HS is clear, the practice can be difficult. When CON, GEN, or the input set are even moderately complex, accurate determination of typologies by hand becomes impossible. For this reason, computational tools for conducting typological research have been developed. It is instructive to compare the procedure for computing typologies in the two main systems in current use, OTSoft for P-OT (Hayes, Tesar and Zuraw 2003) and OT-Help for HS (Staubs et al. 2010).

In OTSoft, a language is a set of underlying forms and the surface forms that they map to. A typology of a constraint system is a set of languages and the rankings that produce each of them. For each underlying form, the user supplies a tableau with various surface candidates and the violation marks they receive from each constraint. OTSoft then uses Recursive Constraint Demotion (RCD) (Tesar and Smolensky 1998) to determine the possible optima for each underlying form. This is done by iterating through the surface candidates from each underlying form, setting each candidate as optimum and then using RCD to find a ranking. If RCD fails to find a ranking, then that candidate is not a possible optimum; if it succeeds, then that candidate is a possible optimum. The typology is computed by trying all combinations of possible optima, one for each underlying form. If RCD succeeds in finding a ranking, then this combination — this language — is in the typology. Otherwise it is not.

In HS, and consequently in OT-Help, a language is not just a combination of underlying forms and the surface forms that they map to; it also includes the derivations that link the underlying and surface forms. Because different rankings produce different derivations, and because two different derivations can sometimes produce the same underlying→surface mapping, derivations as well as mappings must be included in the typology. OT-Help must therefore iterate the procedure used in OTSoft, finding the possible optima at step 1, then finding the possible optima at step 2 by setting each of the step-1 optima as an input, and so on. A further point of difference is that OT-Help provides users with a language for defining operations in GEN and constraints in CON. This allows OT-Help to compute its own candidates (relying on the finiteness of the candidate set in HS) and to determine their constraint violations. Users can quickly and easily test hypotheses about GEN and CON, to address questions like those discussed above.

The study of language typology in HS is still at a relatively early stage. Besides the papers already mentioned, there has been work on typology in relation to autosegmental spreading (McCarthy 2007b, 2010a), apocope and metathesis (McCarthy 2007b), lexical structure (McCarthy and Pruitt 2012), and reduplication (McCarthy, Kimper and Mullin 2012). In this volume, the chapters by Moore-Cantwell, Staubs, and Kimper also address typological issues from a HS perspective.

8 Opacity in Harmonic Serialism

Certain types of phonological opacity are amenable to analysis in HS — see Elfner (this volume) and Torres-Tamarit (this volume), for example. HS is not a general theory of phonological opacity, however. I will illustrate this point with a pair of examples.

The derivation in (45) exemplifies the type of opacity known as *counterbleeding*.

(45) Counterbleeding order in Bedouin Arabic

Underlying	/ħaːkim-iːn/	
Palatalization ($k \rightarrow k^j$/_i)	ħaːk^jimiːn	
Syncope (i→ Ø/VC_CV)	ħaːk^jmiːn	
Surface	ħaːk^jmiːn	'ruling (masculine plural)'

This example is problematic in P-OT because deleting /i/ will satisfy two markedness constraints at once: the one that favors palatalizing *k* before *i*, and the one that favors deleting short high vowels in open syllables. Thus, the winning candidate should be *ħaːkmiːn*, with unpalatalized *k*. The actual winner, *ħaːk^jmiːn*, seems needlessly unfaithful because it has palatalized *k* even though the *i* that triggers palatalization is absent from the surface form.

HS fares no better on this example. If *ħaːkmiːn* is a candidate at step 1, then it will win for the same reason that it wins in P-OT. The derivation will then converge on it at step 2. At step 1, *ħaːk^jimiːn* loses to *ħaːkmiːn* because it violates the constraint against short high vowels in open syllables and both candidates do equally well on the constraint against plain *k* before *i*. And at step 2 *ħaːk^jmiːn* loses to *ħaːkmiːn* because it is unfaithful.

Counterfeeding opacity is also problematic in HS. Recall from (17) that Cairene Arabic shortens long vowels in unstressed syllables. Cairene also has a syncope process that deletes short high vowels in a VC__CV context, but short vowels derived from long ones do not delete:

(46) Cairene chain shift

a. Short high vowels delete in VC__CV

/fihim-uː/	fíhmu	'they understood' (cf. *fíhim* 'he understood')

b. But not if they are derived from long vowels

/ji-ʃiːl-uː-na/	jiʃilúːna	'they ask us' (cf. *jiʃíːl* 'he asks')
	*jiʃlúːna	

This combination of processes is called a *chain shift* because the output of one process is identical with the input to another: A → B and B → C in identical or overlapping contexts. This traditional name is somewhat misleading because the point is that the processes do not in fact chain together — underlying A does not map to surface C.

In P-OT, deletion of long vowels can be prevented by the constraint MAX(Vː) (Gouskova 2003). This constraint is defined so that it checks a vowel's length in the input, not the output. Therefore, it protects underlying long vowels from deletion even when they have been shortened.

This move does not carry over to HS, however. The problem is that an underlying long vowel, once it has been shortened in the course of the derivation, is indistinguishable from an underlying short vowel. There is no obvious way of ruling out a derivation where a vowel first shortens (*jiʃiːlúːna* → *jiʃilúːna*) and then deletes (*jiʃilúːna* → **jiʃlúːna*). As far as MAX(Vː) is concerned, the short *i* from /iː/ is indistinguishable from the short *i* in the middle syllable of /fihim-u/. Since /fihim-u/ → *fíhmu*, we expect /jiʃiːluːna/ → **jiʃlúːna*.

One solution would be to take a different view of faithfulness constraints in HS. If MAX(Vː) always looked back at the underlying representation instead of the input to the current derivational step, then it would see /ji-ʃiːl-uː-na/'s long /iː/ in the second syllable. The problem with this move is that is inconsistent with the improved theory of positional faithfulness in HS (see section 7.1).

Another approach, equally applicable in P-OT, is to reexamine purported chain shifts with an eye toward determining whether the B that is the output of the A → B mapping is truly identical with the B that is the input to the B → C mapping. If they are merely similar and not identical, a HS analysis may be possible. In Cairene, for example, shortened *i* is tense, like long *iː*, but *i* not derived from a long vowel is lax (Mitchell 1956: 10, 112). Also see Gouskova and Hall (2009) on possible phonetic differences between underlying and epenthetic vowels.

The most radical solution to the opacity problem is to adopt something like OT-CC (McCarthy 2007a, Wolf 2008, 2011). OT-CC is specifically a theory of opacity, based on evaluating derivations. It uses something like HS as its GEN, and it compares derivations using constraints on the order of operations. For further information, see Wolf (this volume).

9 Conclusion

As I noted earlier, OT is hard, and HS makes it even harder. In parallel OT, typology follows from hypotheses about the constraint set. In HS, typology follows from a combination of hypotheses about GEN and the constraint set. The results so far, in this volume and elsewhere, suggest that HS is worth this extra effort.

Notes

[1] This research was supported by grant BCS-0813829 from the National Science Foundation to the University of Massachusetts Amherst. I am indebted to all of the participants in our weekly grant group for their advice about this chapter, with particular thanks going to Kathryn Pruitt and Joe Pater for their comments on the manuscript.

[2] Thanks to Kevin Mullin for creating the diagram in (2).

[3] Other markedness constraints rule out other parses of faithful *fʕal*, such as those where *f* is a syllable nucleus or an appendix to the word.

[4] Tableaux are in comparative format (Prince 2002). The winning candidate appears to the right of the arrow, and losers are in the rows below it. Integers stand for the number of violation marks incurred by a candidate, replacing the familiar strings of asterisks. In loser

rows, the effects of the constraints are indicated by W and L, W if the constraint favors the winner and L if it favors the loser.

[5] Limits on how much a single rule can do were a later development in RBP. Examples include Prince's (1983) *Move x* and Archangeli and Pulleyblank's (1994) *Insert path*.

[6] Unsyllabified consonants are italicized in the tableaux.

[7] In McCarthy (2008b, c), I assume that faithfulness constraints in HS refer to the underlying representation, but I also observe that the same results could be obtained with faithfulness constraints that refer to the input to the current step. The latter view has become standard in subsequent work in HS.

[8] Coalescence and breaking, which violate the correspondence constraints UNIFORMITY and INTEGRITY, have not yet been examined from an HS perspective.

[9] Portions of this section come from McCarthy (2009).

[10] The derivations in (23) do not exhaust the possibilities. For instance, (23a) could be elaborated further, with syllabification proceeding segment-by-segment or syllable-by-syllable, rather than all at once. Or (23b) could be further compressed, by allowing initial syllabification and assignment of stress to occur in a single step. Other contributions to this volume, particularly Moore-Cantwell's, explore other ideas about how syllabification is integrated into derivations.

[11] The argument that resyllabification is a faithful operation is based on an observation about syllabification. Although languages differ in how they syllabify (e.g., [qa**b.l**a] 'before' in Arabic vs. [ə.**bl**ɑjʤ] *oblige* in English), no known language has a contrast between monomorphemic [qab.la] and [qa.bla] (Blevins 1995: 221, Clements 1986: 318, Hayes 1989: 260, McCarthy 2003: 60-62). It is a basic tenet of OT that lack of contrast means lack of faithfulness.

[12] This section originally appeared in McCarthy (2010b).

[13] Not shown in (24) are lengthening of stressed vowels and main stress on the final foot.

[14] Other candidates, such as trochaic *(wána)mari* or right-to-left *wana(marí)*, are ruled out by constraints that are standard in the OT literature on stress systems. For textbook treatments, see Kager (1999: 142ff.) or McCarthy (2008a: 183ff.).

References

Alderete, John (1998) *Morphologically-Governed Accent in Optimality Theory*. Doctoral dissertation, University of Massachusetts Amherst, Amherst, MA. Available on Rutgers Optimality Archive, ROA-309. Published (2001) by Routledge.

Alderete, John (2001) Dominance effects as transderivational anti-faithfulness. *Phonology* 18: 201-253.

Allen, W. S. (1973) *Accent and Rhythm*. Cambridge: Cambridge University Press.

Archangeli, Diana and Douglas Pulleyblank (1994) *Grounded Phonology*. Cambridge, MA: MIT Press.

Beckman, Jill (1998) *Positional Faithfulness*. Doctoral dissertation, University of Massachusetts Amherst, Amherst, MA. Available on Rutgers Optimality Archive, ROA-234.

Black, H. Andrew (1993) *Constraint-Ranked Derivation: A Serial Approach to Optimization*. Doctoral dissertation, University of California, Santa Cruz. Available at http://www.sil.org/silepubs/abstract.asp?id=47751.

Blevins, Juliette (1995) The syllable in phonological theory. In John A. Goldsmith (ed.), *The Handbook of Phonological Theory*, 206-244. Cambridge, MA, and Oxford, UK: Blackwell.

Blumenfeld, Lev (2006) *Constraints on Phonological Interactions*. Doctoral dissertation, Stanford University, Stanford, CA. Available on Rutgers Optimality Archive, ROA-877.

Chen, Matthew (1999) Directionality constraints on derivation? In Ben Hermans and Marc van Oostendorp (eds.), *The Derivational Residue in Phonological Optimality Theory*, 105-127. Amsterdam: John Benjamins. Available on Rutgers Optimality Archive, ROA-453.

Cho, Young-Mee Yu (1990) *Parameters of Consonantal Assimilation*. Doctoral dissertation, Stanford University.

Chomsky, Noam (1977) On *Wh*-movement. In Peter Culicover, Thomas Wasow and Adrian Akmajian (eds.), *Formal Syntax*, 77-132. New York: Academic Press.

Chomsky, Noam and Morris Halle (1968) *The Sound Pattern of English*. New York: Harper & Row.

Clements, G. N. (1986) Syllabification and epenthesis in the Barra dialect of Gaelic. In Koen Bogers, Harry van der Hulst and Maarten Mous (eds.), *The Phonological Representation of Suprasegmentals*, 317-336. Dordrecht: Foris.

Dixon, Robert M. W. (1988) *A Grammar of Boumaa Fijian*. Chicago: University of Chicago Press.

Goldsmith, John (1990) *Autosegmental and Metrical Phonology*. Oxford and Cambridge, MA: Blackwell.

Goldsmith, John (1993) Harmonic phonology. In John Goldsmith (ed.), *The Last Phonological Rule: Reflections on Constraints and Derivations*, 21-60. Chicago: University of Chicago Press.

Gouskova, Maria (2003) *Deriving Economy: Syncope in Optimality Theory*. Doctoral dissertation, University of Massachusetts Amherst, Amherst, MA. Available on Rutgers Optimality Archive, ROA-610.

Gouskova, Maria and Nancy Hall (2009) Acoustics of unstressable vowels in Lebanese Arabic. In Steve Parker (ed.), *Phonological Argumentation: Essays on Evidence and Motivation*. London: Equinox Publishing.

Hawkins, W. N. (1950) Patterns of vowel loss in Macushi (Carib). *International Journal of American Linguistics* 16: 87-90.

Hayes, Bruce (1985) Iambic and trochaic rhythm in stress rules. In M. Niepokuj, M. VanClay, V. Nikiforidou and D. Jeder (eds.), *Proceedings of BLS 11: Parasession on Poetics, Metrics, and Prosody*, 429-46. Berkeley: Berkeley Linguistic Society.

Hayes, Bruce (1989) Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20: 253-306.

Hayes, Bruce (1995) *Metrical Stress Theory: Principles and Case Studies*. Chicago: The University of Chicago Press.

Hayes, Bruce, Bruce Tesar and Kie Zuraw. (2003) OTSoft 2.1. Unpublished paper, UCLA and Rutgers University, Los Angeles and New Brunswick,NJ. Software package available at http://www.linguistics.ucla.edu/people/hayes/otsoft/.

Ito, Junko (1989) A prosodic theory of epenthesis. *Natural Language & Linguistic Theory* 7: 217-259.

Jesney, Karen (2011) Positional faithfulness, non-locality, and the Harmonic Serialism solution. In Suzi Lima, Kevin Mullin and Brian Smith (eds.), *Proceedings of the 39th Annual Meeting of the North East Linguistic Society*, 239-440. Amherst, MA: GLSA Publications. Available on Rutgers Optimality Archive, ROA-1018.

Kager, René (1999) *Optimality Theory*. Cambridge: Cambridge University Press.

Kenstowicz, Michael (1994) Syllabification in Chukchee: A constraints-based analysis. In Alice Davison, Nicole Maier, Glaucia Silva and Wan Su Yan (eds.), *Proceedings of the Formal Linguistics Society of Mid-America 4*, 160-181. Iowa City: Department of Linguistics, University of Iowa. Available on Rutgers Optimality Archive, ROA-30.

Kenstowicz, Michael and Charles Kisseberth (1979) *Generative Phonology: Description and Theory*. New York: Academic Press.

Kimper, Wendell (2011) Locality and globality in phonological variation. *Natural Language & Linguistic Theory* 29: 423-465.

Kiparsky, Paul (1993) Blocking in non-derived environments. In Sharon Hargus and Ellen Kaisse (eds.), *Studies in Lexical Phonology*, 277-313. San Diego: Academic Press.

Kiparsky, Paul (2000) Opacity and cyclicity. *The Linguistic Review* 17: 351-367.

Kirchner, Robert (1996) Synchronic chain shifts in Optimality Theory. *Linguistic Inquiry* 27: 341-350. Available on Rutgers Optimality Archive, ROA-66.

Kisseberth, Charles (1970) On the functional unity of phonological rules. *Linguistic Inquiry* 1: 291-306.

Lombardi, Linda (1995/2001) Why Place and Voice are different: Constraint-specific alternations in Optimality Theory. In Linda Lombardi (ed.), *Segmental Phonology in Optimality Theory: Constraints and Representations*, 13-45. Cambridge: Cambridge University Press. Originally circulated in 1995. Available on Rutgers Optimality Archive, ROA-105.

Mascaró, Joan. (1987) A reduction and spreading theory of voicing and other sound effects. Unpublished paper, Universitat Autònoma de Barcelona, Barcelona.

McCarthy, John J. (2000) Harmonic serialism and parallelism. In Masako Hirotani (ed.), *Proceedings of the North East Linguistics Society 30*, 501-524. Amherst, MA: GLSA Publications. Available on Rutgers Optimality Archive, ROA-357.

McCarthy, John J. (2002) *A Thematic Guide to Optimality Theory*. Cambridge: Cambridge University Press.

McCarthy, John J. (2003) Sympathy, cumulativity, and the Duke-of-York gambit. In Caroline Féry and Ruben van de Vijver (eds.), *The Syllable in Optimality Theory*, 23-76. Cambridge: Cambridge University Press.

McCarthy, John J. (2007a) *Hidden Generalizations: Phonological Opacity in Optimality Theory*. London: Equinox Publishing.

McCarthy, John J. (2007b) Restraint of analysis. In Sylvia Blaho, Patrik Bye and Martin Krämer (eds.), *Freedom of Analysis*, 203-231. Berlin and New York: Mouton de Gruyter. Also appears in Eric Bakovic, Junko Ito, and John J. McCarthy, eds. (2006) Wondering at the Natural Fecundity of Things: Essays in Honor of Alan Prince. Linguistics Research Center, UC Santa Cruz. Available at http://repositories.cdlib.org/lrc/prince/10.

McCarthy, John J. (2008a) *Doing Optimality Theory*. Malden, MA, and Oxford, UK: Blackwell.

McCarthy, John J. (2008b) The gradual path to cluster simplification. *Phonology* 25: 271-319. doi:10.1017/S0952675708001486.

McCarthy, John J. (2008c) The serial interaction of stress and syncope. *Natural Language & Linguistic Theory* 26: 499-546. DOI: 10.1007/s11049-008-9051-3.

McCarthy, John J. (2009) Studying Gen. *Journal of the Phonetic Society of Japan* 13: 3-12.

McCarthy, John J. (2010a) Autosegmental spreading in Optimality Theory. In John Goldsmith, Elizabeth Hume and Leo Wetzels (eds.), *Tones and Features*, 195-222. Berlin and New York: Mouton de Gruyter. Available at: http://works.bepress.com/john_j_mccarthy/100.

McCarthy, John J. (2010b) An introduction to Harmonic Serialism. *Language and Linguistics Compass* 4: 1001-1018.

McCarthy, John J. (2011) Perceptually grounded faithfulness in Harmonic Serialism. *Linguistic Inquiry* 42. Available at http://works.bepress.com/john_j_mccarthy/104/.

McCarthy, John J., Wendell Kimper and Kevin Mullin (2012) Reduplication in Harmonic Serialism. *Morphology* 22: 173-232.

McCarthy, John J. and Alan Prince (1986/1996) Prosodic Morphology 1986. New Brunswick, NJ: Rutgers University Center for Cognitive Science. Available at http://ruccs.rutgers.edu/pub/papers/pm86all.pdf.

McCarthy, John J. and Alan Prince (1993) Generalized Alignment. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology*, 79-153. Dordrecht: Kluwer. Available on Rutgers Optimality Archive, ROA-7.

McCarthy, John J. and Alan Prince (1995) Faithfulness and reduplicative identity. In Jill Beckman, Laura Walsh Dickey and Suzanne Urbanczyk (eds.), *University of Massachusetts Occasional Papers in Linguistics 18*, 249-384. Amherst, MA: GLSA Publications. Available on Rutgers Optimality Archive, ROA-103.

McCarthy, John J. and Alan Prince (1999) Faithfulness and identity in Prosodic Morphology. In René Kager, Harry van der Hulst and Wim Zonneveld (eds.), *The Prosody-Morphology Interface*, 218-309. Cambridge: Cambridge University Press.

McCarthy, John J. and Kathryn Pruitt (2012) Sources of phonological structure. In Ralf Vogel and Hans Broekhuis (eds.), *Linguistic Derivations and Filtering: Minimalism and Optimality Theory*. London: Equinox Publishing.

Mester, Armin (1994) The quantitative trochee in Latin. *Natural Language & Linguistic Theory* 12: 1-61.

Mitchell, T. F. (1956) *An Introduction to Egyptian Colloquial Arabic*. Oxford: Oxford University Press.

Moreton, Elliott. (2000) Faithfulness and potential. Unpublished paper, University of Massachusetts Amherst, Amherst, MA.

Moreton, Elliott (2003) Non-computable functions in Optimality Theory. In John J. McCarthy (ed.), *Optimality Theory in Phonology: A Reader*, 141-163. Malden, MA, and Oxford, UK: Blackwell. Available on Rutgers Optimality Archive, ROA-364.

Moreton, Elliott and Paul Smolensky (2002) Typological consequences of local constraint conjunction. In Line Mikkelsen and Christopher Potts (eds.), *Proceedings of the 21st West Coast Conference on Formal Linguistics*, 306-319. Cambridge, MA: Cascadilla Press. Available on Rutgers Optimality Archive, ROA-525.

Mullin, Kevin, Brian W. Smith, Joe Pater and John J. McCarthy. (2010) OT-Help 2.0 User Guide. Unpublished paper, University of Massachusetts Amherst, Amherst, MA. http://people.umass.edu/kmullin/OTHelp2man.pdf.

Newman, Stanley (1944) *Yokuts Language of California*. New York: Viking Fund.

Norton, Russell J. (2003) *Derivational Phonology and Optimality Phonology: Formal Comparison and Synthesis*. Doctoral dissertation, University of Essex, Colchester, Essex. Available on Rutgers Optimality Archive, ROA-613.

Paradis, Carole (1988a) On constraints and repair strategies. *The Linguistic Review* 6: 71-97.

Paradis, Carole (1988b) Towards a theory of constraint violations. *McGill Working Papers in Linguistics* 5: 1-43.

Pater, Joe (1999) Austronesian nasal substitution and other NC effects. In René Kager, Harry van der Hulst and Wim Zonneveld (eds.), *The Prosody-Morphology Interface*, 310-343. Cambridge: Cambridge University Press. Available on Rutgers Optimality Archive, ROA-160. Reprinted in *Optimality Theory in Phonology: A Reader*, ed. by John J. McCarthy, Malden, MA and Oxford, Blackwell (2004).

Pater, Joe (2012) Serial Harmonic Grammar and Berber syllabification. In Toni Borowsky, Shigeto Kawahara, Takahito Shinya and Mariko Sugahara (eds.), *Prosody Matters: Essays in Honor of Lisa Selkirk*. London: Equinox Publishing.

Poser, William J. (1982) Phonological representations and action-at-a-distance. In Harry van der Hulst and Norval Smith (eds.), *The Structure of Phonological Representations*, 121-158. Dordrecht: Foris.

Prince, Alan (1983) Relating to the grid. *Linguistic Inquiry* 14: 19-100.

Prince, Alan (1990) Quantitative consequences of rhythmic organization. In M. Ziolkowski, M. Noske and K. Deaton (eds.), *Parasession on the Syllable in Phonetics and Phonology*, 355-398. Chicago: Chicago Linguistic Society.

Prince, Alan (2002) Arguing optimality. In Angela Carpenter, Andries Coetzee and Paul de Lacy (eds.), *University of Massachusetts Occasional Papers in Linguistics 26: Papers in Optimality Theory II*, 269-304. Amherst, MA: GLSA. Available on Rutgers Optimality Archive, ROA-562.

Prince, Alan and Paul Smolensky (1993/2004) *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden, MA, and Oxford, UK: Blackwell.

Pruitt, Kathryn (2010) Serialism and locality in constraint-based metrical parsing. *Phonology* 27: 481-526.

Pruitt, Kathryn (2012) *Stress in Harmonic Serialism*. Doctoral dissertation, University of Massachusetts, Amherst, Amherst, MA.

Radhakrishnan, R. (1981) *The Nancowry Word: Phonology, Affixal Morphology and Roots of a Nicobarese Language*. Edmonton: Linguistic Research.

Rubach, Jerzy (2000) Glide and glottal stop insertion in Slavic languages: A DOT analysis. *Linguistic Inquiry* 31: 271-317.

Schütz, Albert J. (1985) *The Fijian Language*. Honolulu: University of Hawaii Press.

Staubs, Robert, Michael Becker, Christopher Potts, Patrick Pratt, John J. McCarthy and Joe Pater (2010) OT-Help 2.0 [computer program]. Computer program. Available at http://web.linguist.umass.edu/~OTHelp/.

Steriade, Donca (2001/2008) The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In Kristin Hanson and Sharon Inkelas (eds.), *The Nature of the Word: Studies in Honor of Paul Kiparsky*, 151-180. Cambridge, MA: MIT Press. Originally circulated in 2001.

Tesar, Bruce and Paul Smolensky (1998) Learnability in Optimality Theory. *Linguistic Inquiry* 29: 229-268. Available on Rutgers Optimality Archive, ROA-155. Reprinted in *Optimality Theory in Phonology: A Reader*, ed. by John J. McCarthy, Malden, MA and Oxford, Blackwell (2004).

Watson, Janet C. E. (2002) *The Phonology and Morphology of Arabic*. Oxford: Oxford University Press.

Wilson, Colin (2001) Consonant cluster neutralization and targeted constraints. *Phonology* 18: 147-197.

Wolf, Matthew (2008) *Optimal Interleaving: Serial Phonology-Morphology Interaction in a Constraint-Based Model*. Doctoral dissertation, University of Massachusetts Amherst, Amherst, MA. Available on Rutgers Optimality Archive, ROA-996.

Wolf, Matthew (2011) Limits on global rules in Optimality Theory with Candidate Chains. *Phonology* 28: 87-128.