2010

Harmonic Serialism Supplement to Doing Optimality Theory

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

Doing Optimality Theory
Supplement on Harmonic Serialism
John J. McCarthy
August 25, 2010

Preparation of this work was supported by grant BCS-0813829 from the National Science Foundation to the University of Massachusetts Amherst.

Section and page numbers refer to my *Doing Optimality Theory* (Blackwell, 2008).

Contents

1 New section after 2.5 (page 80) *Harmonic Serialism: Theory and Analysis*

Harmonic Serialism (HS) is a variant of OT. The difference between HS and standard OT is that HS is a **serial** model and standard OT is a **parallel** model. In this section, I'll explain what that means and show how to do analyses in HS. Later in this document, you will find additions to other parts of *Doing OT* that discuss further implications of HS and give some reasons for thinking that it might be superior to standard OT. At the end of this section, I have included some information on the origins of HS and suggestions for further reading.

In standard 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, in tableau (33) on page 64 of *Doing OT*, the candidates from underlying /taxaː-k^{ʔ}a/ include some that differ from it by a single change (such as *ta.xaːk^{ʔ}*, which differs only by virtue of a single deletion) and some that differ by two or more changes (such as *ta.xak^{ʔ}*, which differs by both deletion and shortening, or *tax*, which differs by two deletions). In standard OT, GEN applies one or more operations together, in parallel, when it generates candidates. In consequence, standard OT's candidate sets are quite diverse and involve competition between candidates that may not resemble each other very much.

(33) Summary tableau: /taxaː-k^{ʔ}a/ → [ta.xak^{ʔ}]

/taxaː-k^{ʔ}a/	*COMP-SYLL	*C^{unsyll}	*V#	MAX	IDENT(long)
a. → ta.xak^{ʔ}				*	*
b. ta.xaː.k^{ʔ}a			*W	L	L
c. ta.xaːk^{ʔ}	*W			*	L
d. ta.xaː.k^{ʔ}		*W		*	L
e. ta.xaː			*W	**W	L
f. tax				***W	L

In HS, GEN is limited to making just one change at a time. The candidate set from underlying /taxaː-k^{ʔ}a/ consists of all of the ways of making no more than a single change in this form: *ta.xaːk^{ʔ}*, *ta.xa.k^{ʔ}a*, and so on, but not *ta.xak^{ʔ}* or *tax*, which require two changes. In consequence, HS's candidate sets are less diverse than standard OT's, and they involve competition between candidates that are not very different. This property of HS's GEN is sometimes referred to as **gradualness**.

HS would be unworkable as a theory of language if all it did was to impose this limitation on GEN. That's because underlying and surface representations can differ from one another in several ways. For instance, the surface form from underlying /taxaː-k^{ʔ}a/ is *ta.xak^{ʔ}*, which shows the effects of both vowel deletion and vowel shortening. As we just saw, *ta.xak^{ʔ}* is not one of the candidates that GEN produces from /taxaː-k^{ʔ}a/. 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

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 later.) 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/.

I'll first illustrate the process of constructing an analysis in HS with an example that's simpler than Yawelmani, but then we'll return 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!'. The markedness constraints involved are *COMPLEX-ONSET, which is violated by faithful *fʕal*, and ONSET, which is violated by *if.ʕal*, which has an epenthetic *i* but no epenthetic *ʔ*. In a standard OT analysis, both of these constraints are ranked at the top of the hierarchy, dominating DEP:

(2) Standard OT analysis of Arabic /fʕal/ → *ʔif.ʕal*

	/fʕal/	*COMPLEX-ONSET	ONSET	MAX	CONTIGUITY	DEP
a. →	ʔif.ʕal					**
b.	fʕal	*W				L
c.	if.ʕal		*W			*L
d.	ʕal			*W		L
e.	fi.ʕal				*W	L

Tableau (2) includes a couple of other constraints that are also necessary for the analysis. Because MAX dominates DEP, epenthesis is favored over deletion. And because CONTIGUITY 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'll 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 has to dominate ONSET:

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

	/fʕal/	*COMPLEX-ONSET	MAX	CONTIGUITY	ONSET	DEP
a. →	if.ʕal				*	*
b.	fʕal	*W			L	L
c.	fal		*W		L	L
d.	fi.ʕal			*W	L	*

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.

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

	if.ʕal	*COMPLEX-ONSET	MAX	CONTIGUITY	ONSET	DEP
a. →	ʔif.ʕal					*
b.	if.ʕal				*W	L
c.	fʕal	*W	*W			L

Observe that the faithfulness violations are determined relative to the input to the current step of the derivation; faithfulness is not relative to the underlying representation. This is consistent with HS's basic assumptions and turns out to solve a problem with positional faithfulness, as we will see later (pages 19ff.).

Although **we** know that we have reached the desired surface form, **the grammar** doesn't 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:

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

	ʔif.ʕal	*COMPLEX-ONSET	MAX	CONTIGUITY	ONSET	DEP
a. →	ʔif.ʕal					
b.	ʔi.fi.ʕal					*W
c.	ʔif.ʕa		*W			

At this point, the derivation has converged on the final output *ʔif.ʕal*. That is the surface form, and the GEN/EVAL loop terminates.

Now we're ready to compare the standard OT and HS analyses of Arabic. The standard OT analysis in (2) does not assert any ranking between *COMPLEX-ONSET and ONSET. In fact, these constraints are **unrankable** in a standard OT analysis of Classical Arabic. *COMPLEX-ONSET and ONSET are unrankable because they are unviolated. Constraint conflict is the basis for constraint ranking, and surface-unviolated constraints do not conflict in standard OT.

But *COMPLEX-ONSET and ONSET **are** ranked in the HS analysis, with *COMPLEX-ONSET on top. 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's 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 time-honored OT fashion: the conflicting constraints are ranked, and the lower-ranking constraint is violated to spare a violation of the higher-ranking one.

There is a special kind of tableau that can be useful for studying derivations in HS. It shows the faithful candidate at the beginning and the winners at each successive step until convergence. It is called a harmonic improvement tableau for reasons that will be explained shortly:

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

		*COMPLEX-ONSET	MAX	CONTIGUITY	ONSET	DEP
Faithful	fʕal	*				
Step 1	if.ʕal				*	*
Step 2	ʔif.ʕal					*

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 (6) 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.

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, well, harmonic improvement. Review the explanation of harmony on page 21 of *Doing OT*: harmony is the property that EVAL selects for. Derivations in HS have to show steady harmonic improvement until convergence. The winner at each step of the derivation must be more harmonic than its predecessor. A harmonic improvement tableau shows that visually: high-ranking violation marks disappear as the derivation progresses, sometimes to be replaced by lower-ranking ones.

The harmonic improvement tableau and the idea of harmonic improvement in HS derivations can be put to immediate use in addressing an issue that was raised earlier but postponed: how does syllabification fit into the scheme of HS's gradual GEN? The answer is that (re)syllabification must not count as a change on its own, so it is always free to co-occur with some other change. In the first line of (6), the *f* of *fʕal* is parsed as part of a complex onset. In the second line, it is parsed as a coda. For epenthesis to improve harmony, it has to be possible to change the syllabification of *f* at the same time as *i* is epenthesized. If we tried to do these in separate steps, harmony would not improve at every step:

(7) No harmonic improvement if resyllabification is a separate step

		*COMPLEX-ONSET	MAX	CONTIGUITY	ONSET	DEP
Faithful	fʕal	*				
Step 1	i.fʕal	*			*	*
Step 2	if.ʕal				*	
Step 3	ʔif.ʕal					*

Step 1 does not improve harmony relative to the faithful candidate. This is impossible in HS, so this is an ill-formed derivation.

Why is (re)syllabification "for free", but other changes are mutually exclusive in HS's GEN? One possible answer is that other changes are unfaithful, but (re)syllabification never is. Syllabification does not seem to be contrastive in any language, which means there are no constraints demanding faithfulness to syllabification. So a reasonable hypothesis is that HS's GEN is able to make an unlimited number of faithful changes but only one unfaithful change when it generates a candidate.

Compared with standard OT, doing analysis in HS is easier in some ways and harder in others. It is easier because the difficult problem addressed in section 2.5 of *Doing OT* — figuring out which candidates to worry about — is a lot easier in HS. HS's candidate sets are very narrowly circumscribed by gradualness. Standard OT's "specter of an unexpected competitor suddenly emerging to eliminate the desired winner" (in Karttunen's evocative phrase) is a mere will-o'-the-wisp as far as HS is concerned.

On the other hand, HS is harder because the derivation is another unknown, in addition to the ranking, that the analyst has to figure out. The task of figuring out the derivation is easier than it seems, however. The first step is to inventory all of the unfaithful changes that occur on the path from underlying to surface representation. In /fʕal/ → *ʔif.ʕal*, those changes are obviously the two epenthesis operations. Then, for each pair of operations, ask whether one of them has to occur first. The criterion for whether an operation has to occur first is harmonic improvement. There are two logically possible ways of getting from /fʕal/ to *ʔif.ʕal*. If the derivation is /fʕal/ → *if.ʕal* → *ʔif.ʕal*, with vowel epenthesis first, then we know even before doing the ranking that harmonic improvement is at least a possibility: epenthesizing *i* eliminates an initial cluster, and epenthesizing *ʔ* eliminates the onsetless syllable caused by epenthesis of *i*. But if the derivation is /fʕal/ → *ʔfʕal* → *ʔif.ʕal*, with consonant epenthesis first, then harmonic improvement is out of the question: epenthesizing *ʔ* makes a bad situation even worse.

It sometimes happens that the order in which two changes occur doesn't matter. This is often the case when the two changes are the result of the same operation applied in two different places. For example, exercise 9 on page 41 of *Doing OT* illustrates a process of vowel reduction in Palauan. In /keri-mam/ → *kərəmám*, there are three changes: reduction of *e*, reduction of *i*, and assignment of stress to *a*. Because only unstressed vowels reduce, stress assignment clearly has to precede reduction. But it doesn't matter whether *e* or *i* reduces first. In fact, it might be literally impossible to determine the order in which these vowels reduce, if *kərimám*

Underlying root	*No suffix*	*Root+/k/ 'my'*	*Root+/mam/ 'our'*	
/ʔabu/	[ʔáb]	[ʔəbúk]	[ʔəbəmám]	'ashes'
/mada/	[mád]	[mədák]	[mədəmám]	'eyes'
/keri/	[kér]	[kərík]	[kərəmám]	'question'
/ʔuri/	[ʔúr]	[ʔərík]	[ʔərəmám]	'laughter'
/ʔara/	[ʔár]	[ʔərák]	[ʔərəmám]	'price'
/buʔi/	[búʔ]	[bəʔík]	[bəʔəmám]	'spouse'
/duʔa/	[dúʔ]	[dəʔák]	[dəʔəmám]	'skill'
/badu/	[bád]	[bədúk]	[bədəmám]	'rock'

and *kerəmám* are tied at step 2 of the derivation. It is safe to arbitrarily pick one of them as the winner, because the other unstressed vowel will reduce at step 3 and this temporary difference will disappear.

Another way that HS can be harder than standard OT is that can require a more specific constraint system. Yawelmani /taxa:-k^{ʔ}a/ → *ta.xak^{ʔ}* illustrates this.

In the standard OT analysis of Yawelmani in section 2.3 of *Doing OT*, there is a single markedness constraint, *COMP-SYLL, that is violated by both CVCC and CV:C syllables.[1] As *Doing OT* tableau (33) shows, *COMP-SYLL combines with *V# to force vowel shortening in /taxa:-k^{ʔ}a/ → *ta.xak^{ʔ}*. But, as tableau (34) shows, it also blocks apocope in /xat-k^{ʔ}a/ → *xat.k^{ʔ}a*. This difference in *COMP-SYLL's role is determined primarily by the ranking of two faithfulness constraints. *COMP-SYLL causes shortening in /taxa:-k^{ʔ}a/ → *ta.xak^{ʔ}* because it and *V# dominate IDENT(long). But it blocks apocope in /xat-k^{ʔ}a/ → *xat.k^{ʔ}a* because it and DEP dominate *V#. (On the ranking of DEP, see exercise 17 on p. 65 of *Doing OT*.)

(34) Summary tableau: /xat-k^{ʔ}a/ → [xat.k^{ʔ}a]

/xat-k^{ʔ}a/	*COMP-SYLL	*C^{unsyll}	*V#	MAX	IDENT(long)
a. → xat.k^{ʔ}a			*		
b. xatk^{ʔ}	*W		L	*W	
c. xat.k^{ʔ}		*W	L	*W	

17 Using the loser *[xa.tik^{ʔ}], with epenthetic [i], give a ranking argument for DEP and some other constraint in (30). Could you have made this argument using the loser *[xat.k^{ʔ}i] instead?

*COMP-SYLL, functioning in this dual role, is insufficiently specific for the HS analysis. Step 1 of the HS analysis of /xat-k^{ʔ}a/ looks exactly like tableau (34) — even with respect to the omission of the candidate *xa.tik^{ʔ}* mentioned in exercise 17. But if we apply this same ranking to step 1 of /taxa:-k^{ʔ}a/, we get blocking of apocope where blocking is not wanted:

(8) Unwanted blocking of apocope in HS

/taxa:-k^{ʔ}a/	*COMP-SYLL	*V#	MAX	IDENT(long)
a. → ta.xa:k^{ʔ}	*		*	
b. ta.xa:.k^{ʔ}a	L	*W	L	

Tableau (8) is ill-formed because it has an undominated loser-favoring constraint. The difference between the standard OT tableau (33) and the HS tableau (8) is that the form with both apocope and shortening, *ta.xak^{ʔ}*, is not a candidate at step 1 because it is two changes away from the underlying representation. The only way to satisfy *COMP-SYLL and also have apocope in /taxa:-k^{ʔ}a/ is to shorten the penultimate vowel at the same time as when the final vowel is deleted. But HS's one-change-at-a-time GEN cannot do that.

The solution to this conundrum is to introduce an additional constraint that blocks apocope with /xat-k^{ʔ}a/ but not /taxa:-k^{ʔ}a/. This constraint, *COMPLEX-CODA, rules out the CVCC syllable of *xatk^{ʔ}*, but it says nothing about the CV:C syllable of *ta.xa:k^{ʔ}*. It replaces *COMP-SYLL at the top of the hierarchy, and *COMP-SYLL is instead ranked between *V# and IDENT(long).

With this revised grammar, apocope is still correctly blocked at step 1 of /xat-k^{ʔ}a/ → *xat.k^{ʔ}a*:

[1] The standard OT analysis also has a constraint *C^{unsyll} that rules out parsing CVCC and CV:C as a heavy syllable followed by an extrasyllabic consonant. I will ignore *C^{unsyll} and these other parses because they don't really add anything to the comparison of standard OT and HS.

(9) Step 1 of /xat-kˀa/ → *xat.kˀa*

	/xat-kˀa/	*COMPLEX-CODA	*V#	*COMP-SYLL	MAX	IDENT(long)
a.→	xat.kˀa		*			
b.	xatkˀ	*W	L	*W	*W	

But the more specific constraint *COMPLEX-CODA does not block apocope at the first step of /taxaː-kˀa/ → *ta.xaːkˀ* → *ta.xakˀ*:

(10) Step 1 of /taxaː-kˀa/ → *ta.xaːkˀ* → *ta.xakˀ*

	/taxaː-kˀa/	*COMPLEX-CODA	*V#	*COMP-SYLL	MAX	IDENT(long)
a.→	ta.xaːkˀ			*	*	
b.	ta.xaː.kˀa		*W	L	L	

The derivation continues with shortening at step 2:

(11) Step 2 of /taxaː-kˀa/ → *ta.xaːkˀ* → *ta.xakˀ*

	ta.xaːkˀ	*COMPLEX-CODA	*V#	*COMP-SYLL	MAX	IDENT(long)
a.→	ta.xakˀ					*
b.	ta.xaːkˀ			*W		L

And it converges at step 3 (not shown).

Adding a constraint to CON is not something that we want to do just to salvage an analysis in HS. We're obliged to come up with independent motivation for the new constraint. Motivating new constraints is the topic of chapters 4 and 5 of *Doing OT*, where the topic will be treated in detail, but it's not too soon to take a stab at it now. The independent motivation for *COMPLEX-CODA comes from a language that requires this constraint even in a standard OT analysis.

That language is Sudanese Arabic. Sudanese has a process of syncope that affects short vowels in unstressed non-final open syllables. This process can create surface violations of *COMP-SYLL, such as *ji.káːd.bu*, but it cannot create surface violations of *COMPLEX-CODA, such as **jágd.bu*. Unless these two constraints are distinguished, there is no way of analyzing Sudanese in standard OT or HS.

(12) Syncope in Sudanese Arabic (Hamid 1984:82ff.)

a.	/fihim-u/	fíh.mu	'understood (m. pl.)'	
	/jikaːtib-u/	ji.káːd.bu	'correspond (m. pl.)'	
	/maːsik-a/	máːs.ka	'holding (f. sg.)'	
b.	/jaktub-u/	ják.ti.bu	'write (m. pl.)'	*jágd.bu
	/jitarɟim-u/	ji.tár.ɟi.mu	'translate (m. pl.)'	*ji.tárɟ.mu

There are reasons to think that HS might be a better theory of phonology than standard OT. The evidence comes from language typology and is discussed below on pages 19ff., 23ff., and 30ff.

FOR FURTHER READING

HS was first mentioned in OT's *locus classicus*, Prince and Smolensky (1993/2004), but it was not pursued there and was in fact rejected in favor of the standard parallel version of OT.

The case for HS was reopened in McCarthy (2000, 2002:159-163, 2007b), where some general consequences of this theory were identified and discussed.

To date, most arguments for HS fall into two main categories. Some arguments are based on the fact that HS, unlike standard OT, has forms that are intermediate in the derivation, neither underlying nor surface. The claim is that some generalizations can only be stated correctly on these intermediate forms. Examples of this kind of argument include stress-syncope interaction (McCarthy 2008b), stress-epenthesis interaction (Elfner 2009) (see below pages 26ff.), positional faithfulness (Jesney to appear) (see below pages 19ff.), local variation (Kimper to appear), and phonetics-phonology interaction (McCarthy 2011). Other arguments are based on language typology, specifically the claim that standard OT predicts unattested phonological patterns but HS does not (see below pages 23ff.). Examples of this kind of argument include autosegmental spreading (McCarthy 2007b, to appear), apocope and metathesis (McCarthy 2007b), consonant cluster simplification (McCarthy 2008a), stress (Pruitt 2008), and reduplication (McCarthy, Kimper and Mullin 2010).

HS is distinct from, though related to, OT with candidate chains (OT-CC), in which a HS-like system is used to construct derivations that then compete against one another. Research on OT-CC has particularly focused on phonological opacity (McCarthy 2007a) and phonology-morphology interactions (Wolf 2008).

QUESTION

1. The argument in (7) that (re)syllabification cannot require a separate step is based on an unstated assumption: syllables without nuclei are never produced by GEN. Suppose this assumption is changed, so GEN is able to produce a candidate like *Δf.ʕal*, where *Δ* denotes an empty nucleus that can later be filled by epenthesizing a vowel. Does changing this assumption affect the argument?

EXERCISES

2. Construct a harmonic improvement tableau for the derivation in (1).
3. Construct a complete HS analysis of Palauan.
4. Review the Tibetan exercise on page 51 of *Doing OT*. Are there any differences between the standard OT and HS analyses of Tibetan? Explain your answer.
5. Imagine a language that is identical to Yawelmani except that /xat-k^{ʔ}a/ surfaces as *xa.tik^{ʔ}*, with apocope and epenthesis. Construct a HS analysis of this language.

2 Supplement to section 2.6 (page 82) *Harmonic bounding in HS*

An interesting aspect of HS is that candidates that are harmonically bounded in standard OT can be intermediate winners in HS. This situation arises whenever a process is equally applicable in more than location. Because HS's GEN is gradual, the process can only apply once in each step, though eventually it will apply everywhere. This aspect of HS already came up in the discussion of Palauan on page 6; now we'll look at it in more detail.

For example, Cairene Arabic has a process that shortens long vowels in unstressed syllables:

(13) Unstressed syllable shortening in Cairene

/itnaːʔiʃ-na/	ʔitna̲ʔíʃ-na	'we discussed'
/ʃaːf-it-ak/	ʃa̲fítak	'she saw you (m. sg.)'
/ma-ʃaːf-uː-niː-ʃ/	maʃa̲fu̲níːʃ	'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 IDENT(long) in Cairene:

(14) WSP > > IDENT(long) (standard OT analysis)

/ma-ʃaːf-uː-niː-ʃ/	WSP	IDENT(long)
a. → maʃafuníːʃ		**
b. maʃaːfuːníːʃ	**W	L
c. maʃaːfuíníːʃ	*W	*L
d. maʃafuːníːʃ	*W	*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 (yet) 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:

(15) Step 2 of HS analysis

maʃaːfuːníːʃ	WSP	IDENT(long)
a. → maʃafuːníːʃ	*	*
b. → maʃaːfuníːʃ	*	*
c. maʃaːfuːníːʃ	**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:

(16) Step 3 of HS analysis

	maʃafu:ní:ʃ	WSP	IDENT(long)
a. →	maʃafuní:ʃ		*
b. →	maʃafu:ní:ʃ	*W	L

Both of the step 2 winners in the HS analysis are harmonically bounded in standard OT. As tableau (17) shows, they are harmonically bounded by the candidates that shorten neither vowel or both. Because this tableau is intended to show harmonic bounding rather than ranking, it omits all of the appurtenances of ranking, the Ws, Ls, and solid vertical lines.

(17) Harmonic bounding of single shortening in standard OT

/ma-ʃa:f-u:-ni:-ʃ/	WSP	IDENT(long)
a. maʃafuní:ʃ		**
b. maʃa:fu:ní:ʃ	**	
c. maʃa:funí:ʃ	*	*
d. maʃafu:ní:ʃ	*	*

The singly-shortened candidates (17)c and (17)d will lose to (17)a if WSP is ranked higher, and they will lose to (17)b if IDENT(long) is ranked higher. Since they cannot win under any ranking, they are harmonically bounded.[2]

The fact that HS can have intermediate winners that are harmonically bounded in standard OT does not mean that HS overturns the whole notion of harmonic bounding, however. These intermediate winners are also harmonically bounded in the sense that they can never be the final output of the grammar. For one of the vowels to shorten, WSP has to dominate IDENT(long). But once the derivation is launched under that ranking, it will not converge until all unstressed vowels have shortened. In HS, harmonic bounding is a property of derivations rather than forms.

EXERCISE

6. Make up an example where a HS derivation requires an intermediate form that is harmonically bounded in standard OT. The example does not have to be real, but it should be as plausible as possible, and it should not involve vowel shortening.

[2] This is *collective harmonic bounding* (Samek-Lodovici and Prince 2005). The discussion of harmonic bounding in *Doing OT* is limited to simple harmonic bounding.

3 Supplement to section 2.11 (page 124) *Using OT-Help to assist analysis in HS*

OT-Help (Staubs et al. 2010) is a software package for doing analysis and research in HS, available at http://web.linguist.umass.edu/~OTHelp/. OT-Help takes advantage of the differences between HS and standard OT to offer capabilities that go far beyond those of OT-Soft. While OT-Soft requires the user to supply the candidates and their constraint violations, OT-Help generates its own candidates and applies the constraints to them. It does this by applying user-defined GEN and CON. The documentation and sample files, available at the same URL as the software, explain in detail how to use this program. Here I will just present a simple example of OT-Help in use.

Suppose we want to use OT-Help to study the processes of apocope, epenthesis, and vowel shortening in Yawelmani. We need to construct three files: inputs, GEN, and CON. The input file is just a specially structured list of underlying representations:

(18) OT-Help input file: *Yawelmani.txt*

```
[typology]
[begin tableaux]
taxAka              input
xatka               input
?ilkhin             input

[end of tableaux]
```

If we wanted to include additional underlying representations, we could just add them, one per line following the same format. We'll call this file *Yawelmani.txt*. (The *.txt* extension is required.)

I've taken some shortcuts in (18). One is to use capitalization for long vowels. Because OT-Help operates on strings of symbols rather than full linguistic structures, it is often helpful to represent a difference in structure by using a different symbol. If I were to use something like "`a:`" to represent *aː*, I would have the problem of "`:`" being treated like a separate segment by my operations and constraints. Another shortcut, for a similar reason, is elimination of the glottalization in the suffix -*k*ˀ*a*. It is irrelevant to the processes I'm interested in here, and it would just complicate the statement of my operations and constraints.

A GEN file consists of a list of operations, each of which constitutes the single change that HS's GEN is capable of. If the input file is called *Yawelmani.txt*, then the GEN file has to be called *Yawelmani.txt_OPERATIONS*. Here it is:

(19) OT-Help GEN file: *Yawelmani.txt_OPERATIONS*

```
[operation]
[long name]         Epenthesize-i
[active]            yes
[definition]                 i
[violated faith]    Dep

[operation]
[long name]         Delete-V
[active]            yes
[definition]        [aeiou]
[violated faith]    Max
```

```
[operation]
[long name]          Shorten
[active]             yes
[definition]         A        a
[definition]         E        e
[definition]         I        i
[definition]         O        o
[definition]         U        u
[violated faith]     Id(long)

[end operations]
```

The syntax of operation files is explained in detail in the OT-Help documentation, and I won't try to recapitulate that material here. Instead, I'll just try to give a sense of what these particular operations do. The first, `Epenthesize-i`, simply inserts *i* anywhere. The second, `Delete-V`, deletes a short vowel anywhere. The third, `Shorten`, shortens a long vowel anywhere. Applied to the input `taxAka`, these operations will give the following candidate set:

(20) Candidate set from `taxAka` obtained from (19)

```
itaxAka, tiaxAka, taixAka, …
txAka, taxAk
taxaka
```

A constraints file consists of a list of markedness constraints and place-holders for any faithfulness constraints that appear in the `[violated faith]` lines of the operations file. For *Yawelmani.txt*, the constraints file has to be called *Yawelmani.txt_CONSTRAINTS.* Here it is:

(21) OT-Help CON file: *Yawelmani.txt_CONSTRAINTS*

```
[constraint]
[long name]          *Comp-Syll
[active]             yes
[type]               markedness
[definition]         [^aeiou][^aeiouAEIOU]([^aeiouAEIOU]|$)

[constraint]
[long name]          *Comp-Coda
[active]             yes
[type]               markedness
[definition]         [^aeiouAEIOU][^aeiouAEIOU]([^aeiouAEIOU]|$)

[constraint]
[long name]          *V#
[active]             yes
[type]               markedness
[definition]         [aeiouAEIOU]$

[constraint]
[long name]          Dep
[active]             yes
[type]               faithfulness
```

```
[constraint]
[long name]          Max
[active]             yes
[type]               faithfulness

[constraint]
[long name]          Id(long)
[active]             yes
[type]               faithfulness

[end constraints]
```

To simplify the statement of the operations, I elected to define the markedness constraints in terms of strings of segments rather than syllables. `*Comp-Syll` is violated by a sequence of a consonant or long vowel followed by a consonant followed by another consonant or word boundary. `*Comp-Coda` is similar, except that a long vowel will not match the first term of the structural description. And `*V#` is violated by any word-final vowel, short or long.

Before we go on to look at what OT-Help can do with these files, I need to take a moment to clarify what we have and have not accomplished here. We have taken a theory-based understanding of GEN and CON and translated it into a format that OT-Help can understand. That format is not itself a theory of anything; in fact, it's full of compromises and shortcuts that are most emphatically not true of the real theory. Despite these caveats, what we've done is also extremely useful, because it has forced us to be very explicit in our hypotheses about what GEN and CON do, and as we'll now see it also allows us to test these hypotheses.

When given the three *Yawelmani* files, OT-Help computes a typology. The concept of a typology is explained in chapter 5 of *Doing OT* and in the HS supplement to that chapter. For now, we're only going to use OT-Help to check our analysis of Yawelmani, just as we used OT-Soft. The output from OT-Help looks like this:

(22) OT-Help output from *Yawelmani* files

Inputs	taxAka	xatka	?ilkhin
1	taxAka	xatka	?ilkhin
2	taxAka	xatka	?ilikhin, ?ilkihin
3	taxAk	xatka	?ilkhin
4	taxAk	xatka	?ilikhin, ?ilkihin
5	taxAk	xatk	?ilkhin
6	taxAk	xatik	?ilikhin, ?ilkihin
7	taxAik	xatka	?ilikhin, ?ilkihin
8	taxAik	xatik	?ilikhin, ?ilkihin
9	taxak	xatka	?ilkhin
10	taxak	xatka	?ilikhin, ?ilkihin
11	taxak	xatk	?ilkhin
12	taxak	xatik	?ilikhin, ?ilkihin

```
Grammar 1: Dep, Max, Id(long) >> *Comp-Syll, *Comp-Coda, *V#
Grammar 2: *Comp-Syll, *Comp-Coda, Max, Id(long) >> *V#, Dep
Grammar 3: Dep, Id(long) >> *Comp-Coda >> *V# >> *Comp-Syll, Max
Grammar 4: *Comp-Coda, Id(long) >> *V#, Dep >> *Comp-Syll, Max
Grammar 5: *V#, Dep, Id(long) >> *Comp-Syll, *Comp-Coda, Max
Grammar 6: *V#, Id(long) >> *Comp-Coda, Max >> Dep >> *Comp-Syll
Grammar 7: *Comp-Coda, Id(long) >> *V# >> *Comp-Syll, Max >> Dep
Grammar 8: *V#, Id(long) >> *Comp-Syll, *Comp-Coda, Max >> Dep
Grammar 9: Dep >> *Comp-Coda >> *V# >> *Comp-Syll, Max >> Id(long)
Grammar 10: *Comp-Coda >> *V#, Dep >> *Comp-Syll, Max >> Id(long)
Grammar 11: *V#, Dep >> *Comp-Syll, *Comp-Coda, Max >> Id(long)
Grammar 12: *V# >> *Comp-Syll, *Comp-Coda, Max >> Dep >> Id(long)
```

Each row of the table, except the first, is a distinct language — a distinct combination of surface forms obtainable from the given inputs under some ranking of the constraints. Each grammar is the ranking that produces that language.

Because our goal at this point is only to confirm the correctness of our analysis of Yawelmani, all we do is look through the table for a language with the surface forms `taxak`, `xatka`, and `?ilikhin`. Grammar 10 matches that description (almost). The grammar of 10 is consistent with our analysis of Yawelmani, so the correctness of that analysis is confirmed.

I say that grammar 10 **almost** matches Yawelmani because 10 actually has two surface forms for /?ilkhin/, `?ilikhin` and `?ilkihin`. Clicking on them in OT-Help brings up a tableau that shows why there are two surface forms: they tie on all of the constraints. So OT-Help has done more than confirm the basic correctness of the ranking. It has also disclosed an inadequacy in the analysis: another constraint is required to decide where the epenthetic vowel goes in a cluster of three consonants. Discoveries like this prove the value of OT-Help's ability to generate and evaluate candidates automatically.

EXERCISES

7. Use OT-Help to check your solution to the Palauan problem in exercise 3.
8. Read the discussion of ties in section 1.2 of the OT-Help manual (available at http://web.linguist.umass.edu/~OTHelp/OTHelp2man.pdf). Then implement in OT-Help the HS analysis of the Cairene data in (13). How is the intermediate tie treated in OT-Help?

4 Supplement to section 3.3 (page 152) *Tips for writing up an analysis in HS*

Writing up an analysis is a bit more challenging in HS than in standard OT. The source of this added difficulty is the question of where to introduce the derivations. The derivations have to be harmonically improving, but harmonic improvement depends on the ranking, and the ranking arguments depend on the derivations. Where do you start to unravel this twisted skein?

The solution is to start with the data where the derivation is shortest but non-trivial. These are the examples where there is a single change before convergence. Thus, one would not want to start the discussion of Yawelmani with the example /taxaː-kˀa/ → *ta.xakˀ*, which has both apocope and vowel shortening before convergence. Begin instead with /laːn-hin/ → *lanhin*, which has only shortening. Once this has been analyzed, the foundation has been laid for showing that the derivation *ta.xakˀ* has to proceed by way of the intermediate form *ta.xaːkˀ* rather than *ta.xa.kˀa*, because otherwise vowel shortening would not improve harmony.

Once the presentation of short derivations has gone as far as it can, some rankings will be known, but there will remain a residue of rankings that can only be shown by studying the longer derivations. An example of such a ranking is the one proven in tableau (10): *V# dominates *COMP-SYLL. What is the best way of explaining this to readers? I recommend giving the derivation first and only then justifying it, like this:

i. Say what the derivation is: "The derivation is /taxaː-kˀa/ → *ta.xaːkˀ* → *ta.xakˀ*, with apocope before shortening."
ii. Show the ranking needed to produce this derivation, using tableaux like (10) and (11).
iii. Explain why a different ordering of apocope and shortening will not work.

In short, proceed **de**ductively rather than **in**ductively. Lay the derivational cards on the table, and then explain why this is the winning hand.

One aspect of HS that might require special expositional treatment is the existence of unresolved and unresolvable intermediate ties, such as the tie illustrated in (15) and (16). The danger for the writer and reader is that this kind of tie will end up being a major distraction requiring alternative derivations with their accompanying tableaux. As we saw earlier, these intermediate ties are not very interesting because they always end up in the same place anyway, so they don't really deserve all that attention. A better expositional strategy is to mention the tie when it first appears and then declare an ad hoc rule for how it will be handled in subsequent discussion. For instance, something like this would do for the vowel shortening example in (15) and (16):

> The grammar does not determine which unstressed vowel shortens first in *maʃaːfuːníːʃ*. In subsequent discussion, I will arbitrarily assume that vowels shorten from left to right, though nothing hinges on this decision.

Another aspect of HS requiring special expositional treatment is the existence of indeterminate process orderings. When two processes are non-interacting, it doesn't matter which of them occurs first. For example, Axininca Campa (exercise 34 in chapter 2 of *Doing OT*) has processes of *t* epenthesis and *a* epenthesis. Epenthesis of *t* responds to violations of ONSET, while epenthesis of *a* responds to violations of CODA-COND.

/i-N-koma-i/	ONSET	DEP
a. → iŋ.ko.ma.ti	*	*
b. iŋ.ko.ma.i	**	

/no-N-ʧʰik-wai-i/ 'I will continue to cut'	CODA-COND	DEP
a. → noñ.ʧʰi.ka.wai.ti		*
b. noñ.ʧʰik.wai.ti	*	

In the derivation /no-n-tʃʰik-wai-i/ → *noñtʃʰikawaiti*, which epenthesis process occurs first? The answer depends on the ranking of ONSET and CODA-COND: the one that is ranked higher is attended to first. For the derivation of *noñtʃʰikawaiti*, it doesn't matter which constraint is ranked higher, because the same surface form results from either ranking. If other data offer evidence about the ranking of ONSET and CODA-COND, fine, but if not, follow my recommendation for dealing with ties: explain the situation when it is first encountered in the exposition, and then state and follow an ad hoc rule for dealing with it subsequently — e.g., "I will arbitrarily assume that ONSET dominates CODA-COND".

5 Supplement to subsection 4.5.3 (page 192) *Consequences of HS for scalar constraints*

In Berber, any segment can be a syllable nucleus, but whenever there is a choice a higher sonority nucleus is preferred. For example, /t-rgl-t/ is syllabified *tR.gLt*,, with syllabic liquids (indicated by capitalization), rather than **trGlt*, with a syllabic stop. But syllabic stops are possible when no consonant of greater sonority is handy: *tF.tKt*.

Chapter 2 of Prince & Smolensky (1993/2004) contains an analysis of Berber that, somewhat surprisingly, does not use the constraint systems in (22) or (23) of *Doing OT*.

(22) Fixed hierarchy of constraints on sonority of nuclei
*NUC/PLOSIVE >> *NUC/FRICATIVE >> *NUC/NASAL >> *NUC/LIQUID

These systems are related to the sonority scale indirectly, through harmonic alignment. Instead of (22), Prince & Smolensky use a constraint called HNUC that refers directly to the sonority hierarchy: “A higher sonority nucleus is more harmonic than one of lower sonority”. Although this constraint states a preference, we can easily translate it into a constraint that assigns violation marks while still preserving its direct connection with the sonority hierarchy:

(23) HNUC
Assign a nucleus one violation mark for each degree of sonority less than *a* (McCarthy 2003:82).

Constraints defined in this way are referred to as **scalar**.

Pater (to appear) has shown that HNUC will not work as intended in a parallel analysis of Berber. The problem arises when a candidate with two consonantal nuclei competes against a candidate with only one. Compare *K.šM* and **kŠm* as parses of /kšm/. HNUC assigns to *K.šM* the violation marks of *K* **plus** the violation marks of *M*. It assigns to **kŠm* **only** the violation marks of *Š*, which lies between *K* and *M* on the sonority scale. So HNUC is favoring **kŠm* — the wrong result. In the standard parallel version of OT, HNUC cannot simply compare *M* and *Š*, which is what we want it to do. Despite its attractively high sonority, *M* gets dragged down by *K*’s low sonority.

What we need is the ability to compare nuclei one at a time, and HS gives us that. If making a consonant syllabic occupies a step of a HS derivation, then nuclei can only be compared one at a time. At step 1 from /kšm/, the unfaithful candidates are *Kš.m*, *kŠm*, and *k.šM*. HNUC favors *k.šM*, which becomes the input to step 2. At step 2, *K.šM* wins because, although *Š* is a better nucleus than *K*, *š* is already doing duty as the (required) onset of the syllable headed by *M*.

This example illustrates a difference between standard OT and HS. Because it compares complete output candidates, standard OT can produce non-local interactions, such as *M* being dragged down by non-adjacent *K*. In HS, by contrast, the decision about which consonant is the best nucleus is strictly local, because single consonantal nuclei are being compared. As we will see, there is reason to think that HS’s locality limitations are a better match with what languages actually do.

6 New subsection after 4.6.3 (page 208) *Consequences for the theory of faithfulness*

Harmonic Serialism has a couple of major implications for the theory of faithfulness. One is that it casts doubt on the need for correspondence theory, and the other is that it solves a serious problem with positional faithfulness constraints.

OT-Help implements HS without correspondence. It does this by linking the faithfulness violations with the application of operations in GEN. For example, DEP is violated in the mapping /ʔilk-hin/ → *ʔilikhin* not because an *i* in the output has no input correspondent, but because an insertion operation applied to produce this mapping. If we take the implementation of faithfulness in OT-Help as the outline of a theory of faithfulness, then we might say that GEN consists of a list of operations and the faithfulness constraints that they violate:

(24) Operational theory of faithfulness

Operation in GEN	Faithfulness constraint in CON
Insert	DEP
Delete	MAX
Change [αF] to [–αF]	IDENT(F)
Transpose	LINEARITY

Applying one of these operations in a particular context or domain may incur additional faithfulness violations. For example, deletion of a segment from the middle of a word violates not only MAX but also INPUT-CONTIGUITY (see page 197 of *Doing OT*).

The other way that HS impacts the theory of faithfulness is that it solves an outstanding problem with positional faithfulness constraints in standard OT.[3] Consider, for example, the positional faithfulness constraint $\text{IDENT}_{\text{stress}}$(nasal). It protects nasalization contrasts in stressed syllables. When ranked above $*\text{V}_{\text{nasal}}$, 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:

(25) Attested positional faithfulness effect (standard OT)

bãdõ	$\text{IDENT}_{\text{stress}}$(nasal)	PARSE-SYLL	TROCHEE	$*\text{V}_{\text{nasal}}$	IDENT(nasal)
a. → (bã́do)				*	*
b. (bádo)	*W			L	**W
c. (bã́dõ)				**W	L
d. (badṍ)			*W	*	*
e. bado		**W		L	**W

Because of $*\text{V}_{\text{nasal}}$, nasalized vowels are neutralized to oral in unstressed syllables, as in (25)a. But there is no neutralization in stressed syllables (cf. (25)b), because of $\text{IDENT}_{\text{stress}}$(nasal).

[3] The following text has been borrowed from McCarthy (2010a).

Parallel 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 (25): the surface reflex of /ã/ is stressed in both (25)a and (25)b. Candidates that are stressed differently or not at all, such as (25)d and (25)e, are ruled out by other constraints, so they do not depend on IDENT$_{stress}$(nasal) to exclude them.

Now consider what happens when stress differs among the viable candidates. In (26), 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, (26)a and (26)b, that differ in stress:

(26) Unattested positional faithfulness effect (parallel OT)

/pãko/	IDENT$_{stress}$(nasal)	PARSE-SYLL	*V$_{nasal}$	IDENT(nasal)	TROCHEE
a. → (pakó)				*	*
b. (páko)	*!			*	
c. (pã́ko)			*!		
d. pako		**!		*	

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 — **even though both vowels end up as oral at the surface**. No real language does anything remotely like this.

What is the source of this problem? Positional faithfulness constraints can be 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 standard OT, standard 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 (26) is unavoidable.

Jesney (to appear) shows that this 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 to its underlying correspondent". In HS, the input to GEN is not necessarily the underlying representation, so it can have structure that has been assigned by the grammar. Moreover, since the input is the same for all candidates being compared, problems like (26) 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 a positional faithfulness effect. If PARSE-SYLLABLE is ranked higher, as in tableau (27), then stress is assigned first. Stress (re)assignment and denasalization cannot co-occur because of gradualness.

(27) Step 1 from /pão/

/pãko/	IDENT$_{stress}$(nasal)	PARSE-SYLLABLE	*V$_{nasal}$	IDENT(nasal)	TROCHEE
a. → (pã́ko)			*		
b. pako		**W	L	*W	
c. (pãkó)			*		*W

The derivation then converges at step 2, shown in (28). Input *(pã́ko)* has a stressed nasalized vowel. Since this vowel is stressed in the input to step 2, redefined IDENT$_{stress}$(nasal) protects it from denasalization:

(28) Step 2 from /pãko/

(pã́ko)	IDENT$_{stress}$(nasal)	PARSE-SYLLABLE	*V$_{nasal}$	IDENT(nasal)	TROCHEE
a. → (pã́ko)			*		
b. (páko)	*W		L	*W	
c. (pãkó)			*		*W

The failure of the final-stressed candidate *(pãkó)* in (27)c and (28)c 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 parallel OT. In fact, it does not because 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 $n+1$. As we'll see when we look at typology in the HS supplement to chapter 5, the HS derivation gets stuck at a **local optimum** — and sometimes getting stuck is the right thing to do.

QUESTION

9. Coalescence is the merger of two segments into one. The resulting segment combines the feature values of its two parents. The treatment of coalescence in correspondence theory is discussed on page 197 of doing OT: coalescence itself violates UNIFORMITY, and there is also an IDENT violation for every mismatched feature value. For example, /ai/ → *ē* violates not only UNIFORMITY but also IDENT(+low) (because [+low] /a/ corresponds with [–low] to *ē*) and IDENT(+high) (because [+high] /i/ corresponds with [–high] *ē*).

Correspondence theory also recognizes the possibility of two input segments fusing into a single output segment. This phenomenon is called segmental coalescence, and Sanskrit (34) is an example. In this language, the sequences /ai/ and /au/ merge into [ē] and [ō], respectively. Both input segments are in correspondence with the single output segment. This relationship is indicated formally by giving the output segment two indices: /a$_1$i$_2$/ → [ē$_{1,2}$]. Coalescence violates the constraint UNIFORMITY, which is defined in (35). Furthermore, because the output segment is featurally distinct from both its input correspondents, coalescence processes also violate the IDENT(feature) or MAX(feature) constraints discussed below.

(34) Sanskrit vowel coalescence (Whitney 1889)

Underlying	*Surface*	
/tava$_1$ i$_2$ndra/	[tavē$_{1,2}$ndra]	'for you, Indra (vocative)'
/hita$_1$ u$_2$pda$_3$i$_4$ʃah/	[hitō$_{1,2}$padē$_{3,4}$ʃah]	'friendly advice'

(35) UNIFORMITY (UNIF) (No coalescence)
Let *input* = $i_1 i_2 i_3 \ldots i_n$ and *output* = $o_1 o_2 o_3 \ldots o_m$.
Assign one violation mark for every pair i_x and i_y
if $i_x \,\Re\, o_z$ and $i_y \,\Re\, o_z$.

The question is this: Can this view of coalescence be carried over to the operational theory of faithfulness? Why or why not? If not, are there other ways of analyzing these phenomena that are a better fit with the operational theory? (Hint: Look at section 4.6.2 of *Doing OT*.)

7 Supplement to section 4.8 (page 233)
HS and assimilation constraints

HS offers a new perspective on the problem of which markedness constraint favors deletion. See McCarthy (to appear).

8 New section after 5.6 (page 258) *Typological consequences of HS*

The typologies that HS allows are often more restrictive than standard OT typologies, even when the constraints are identical. I'll give an example, and then I'll explain why there's a difference.

The analysis of Yawelmani in chapter 2 of *Doing OT* showed the need for the constraint *V#, which is violated by vowel-final words. The Makassarese exercise in chapter 4 (pages 170-171) requires a constraint called CODA-COND that is violated by word-final consonants other than *ʔ* and *ŋ*. Now, suppose that we rank both of these constraints as well as DEP and IDENT above MAX. In standard OT, this grammar will truncate words after the rightmost *ʔ* or *ŋ*:

(29) A typological prediction of standard OT

/paŋasaka/	*V#	CODA-COND	DEP	IDENT(nasal)	MAX
a. → paŋ					*****
b. paŋasaka	*W				L
c. paŋasakaʔ			*W		L
d. paŋasak		*W			*L
e. paŋasaŋ				*W	*L
f. paŋasa	*W				**L
g. paŋas		*W			***L
h. paŋa	*W				****L

No language does this, and I believe most phonologists would agree that no language **could** do this. The constraints in (29) seem to be well supported, but under this ranking they are doing something that is not well supported at all.

HS does not make this prediction. What HS actually does predict depends on the ranking of *V# and CODA-COND. If *V# is ranked higher, then /paŋasaka/ → *paŋasak*, at which point the derivation converges:

(30) HS with *V# on top — Step 1

/paŋasaka/	*V#	DEP	IDENT(nasal)	CODA-COND	MAX
i. → paŋasak				*	*
ii. paŋasaka	*W			L	L
iii. paŋasakaʔ		*W		L	L

(31) Step 2 — Convergence

paŋasak	*V#	DEP	IDENT(nasal)	CODA-COND	MAX
a. → paŋasak				*	
b. paŋasa	*W			L	*W
c. paŋasaŋ			*W	L	

On the other hand, if CODA-COND is ranked higher, then /paŋasaka/ → *paŋasaka*, with immediate convergence:

(32) HS with CODA-COND on top — Step 1 — Convergence

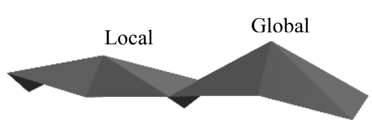
/paŋasaka/	CODA-COND	DEP	IDENT(nasal)	*V#	MAX
a. → paŋasaka				*	
b. paŋasak	*W			L	*W
c. paŋasakaʔ		*W		L	

In fact, **no** ranking of these constraints will produce the language that standard OT predicts. (See exercise 10.)

Why do standard OT and HS produce such different results with the same input and the same constraints? The most immediate answer is that standard OT's candidate set is much more extensive than HS's. In its one and only evaluation, standard OT offers a candidate, *paŋ*, that satisfies **both** *V# and CODA-COND. HS never gets to that candidate. The closest it gets at step 1 is *paŋasak*, which turns out not to be a viable pathway to *paŋ*. Although *paŋasak* wins at step 1 in (30), the candidate with further deletion *paŋasa* loses at step 2 in (31). Like Terry Malloy in *On the Waterfront*, although *paŋ* has the potential to beat all comers, it's never even a contender.

This example illustrates a fundamental difference between standard OT and HS. Standard OT grammars always find the **global optimum** of their large and diverse candidate sets. The global optimum is simply the candidate that best satisfies the constraint hierarchy — the optimal candidate, in short. HS derivations sometimes converge on a candidate that would only be a **local optimum** in the standard OT candidate set. A candidate is only a local optimum if it is more harmonic than some but not all other candidates. In effect, the HS derivation gets stuck: it reaches some local peak of harmony with no nearby candidates that are more harmonic. The situation is represented graphically in (33):

(33) Local and global optima



The goal of the HS research program is to test two empirical claims. One is that convergence on local optima is necessary. In other words, some aspects of language typology are better accounted for in HS than standard OT because HS grammars sometimes end up at local rather than global optima. The other claim is that convergence on local optima is sufficient. In other words, there are no aspects of language typology that would require a HS grammar to get to a global optimum that is inaccessible by any harmonically improving route.

The task of addressing these empirical claims is not trivial. It presents the same challenge as typological research in standard OT: typology depends on hypotheses about CON. To this it adds an additional challenge: in HS, typology also depends on hypotheses about GEN. For instance, standard OT and HS would not differ in their predictions for the /paŋasaka/ example if HS's GEN allowed multi-segment strings to delete in a single step. For fuller discussion of this point, see McCarthy (2010b).

Although this task is not trivial, it has been made easier by the availability of OT-Help (Staubs et al. 2010). The earlier discussion of OT-Help noted that it can be used to check analyses, but its main use is as a tool for computing typologies. The typological consequences of varying hypotheses about GEN and CON can be easily assessed over a range of inputs.

As of this writing, the extant readily accessible literature on typology in HS includes studies of autosegmental spreading (McCarthy 2007b, to appear), apocope and metathesis (McCarthy 2007b), consonant cluster simplification (McCarthy 2008a), stress (Pruitt 2008), and reduplication (McCarthy, Kimper and Mullin 2010). A particular focus of typological work in HS has been the too-many-solutions problem, which will be discussed below, pages 30ff.

EXERCISES

10. Use OT-Help to compute the factorial typology of the constraint set in (29)–(32). First use just the input /paŋasaka/, and then add the inputs /kapaŋa/ and /tapasaka/. Discuss your results.
11. Use OT-Help to confirm Jesney's (to appear) claim about the constraint set in (26)–(28): that in HS it won't produce a language where /pãko/ gets final stress but /bãdõ/, /sato/, and /kafõ/ get penult stress.

9 Supplement to section 6.4 (page 271) *HS and phonological opacity*

Although HS is derivational, it is not — or at least not yet — a general theory of phonological opacity. The rule-based derivation in (8) on page 270 of *Doing OT* exemplifies the type of opacity known as **counterbleeding**. This example (from Bedouin Arabic) is problematic in standard 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 *ħa:kmi:n*, with unpalatalized *k*. The actual winner, *ħa:k*^{j}*mi:n*, seems needlessly unfaithful because it has palatalized *k* even though the *i* that triggers palatalization is absent from the surface form.

(8) Opacity in rule-based phonology

Underlying	/ħa:kim-i:n/
Palatalization	ħa:k^{j}imi:n
Syncope	ħa:k^{j}mi:n
Surface	[ħa:k^{j}mi:n]

The initial impression is that HS fares no better. If *ħa:kmi:n* is a candidate at step 1, then it will win for the same reason that it wins in standard OT. The derivation will then converge on it at step 2. At step 1, *ħa:k*^{j}*imi:n* loses to *ħa:kmi:n* 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 *ħa:k*^{j}*mi:n* loses to *ħa:kmi:n* because it is unfaithful.

Nonetheless, there are cases of phonological opacity that can be analyzed in HS. Elfner (2009) presents a HS analysis of stress/epenthesis interactions in Levantine Arabic. In a rule-based analysis, the rule of stress assignment is ordered before the rule of epenthesis, as in the derivation (34)a. (The derivation (34)b is included to show the stress pattern in an otherwise identical word without any epenthetic vowels.)

(34) Rule-based analysis of Levantine stress/epenthesis interaction

	a. 'I wrote to him'	b. 'she wrote to him'
Underlying	/katab-t-l-u/	/katab-it-l-u/
Stress	katábtlu	katabítlu
Epenthesis	katábitlu	—

As we will now see, the HS analysis involves similar derivational steps.

At step 1 from underlying /katab-t-l-u/, the candidates include *katábtlu* with stress and *katabitlu* with epenthesis, but no candidates with both stress and epenthesis. The competition between *katábtlu* and *katabitlu* is a question of which markedness constraint is ranked higher, the constraint that wants stress (HEADEDNESS(word) — see *Doing OT* page 181) or the constraint against triconsonantal clusters (*COMPLEX-CODA). If PARSE-SYLLABLE is ranked higher, then *katábtlu* wins at step 1:

(35) Step 1 from /katab-t-l-u/

/katab-t-l-u/	HEADEDNESS(word)	*COMPLEX-CODA	DEP
a. → katábtlu		*	
b. katabtlu	*W	*	
c. katabitlu	*W	L	*W

At step 2, HEADEDNESS(word) has been satisfied, so *COMPLEX-CODA is next in line for attention. It is satisfied by epenthesis:

(36) Step 2 from /katab-t-l-u/

katábtlu	HEADEDNESS(word)	*COMPLEX-CODA	DEP
a. → katábitlu			*
b. katábtlu		*W	L

The derivation converges at step 3.

The key insight in this analysis is that the ranking of markedness constraints determines the order of the processes that are triggered by those markedness constraints. Ranking HEADEDNESS(word) higher ensures that stress is assigned first with epenthesis second, so the epenthetic vowel isn't there when stress is assigned. A similar solution could perhaps be developed for the *ħaːkʲmiːn* problem — see exercise 12. Whether this approach can be generalized to all cases of counterbleeding opacity remains to be seen.

Counterfeeding opacity is, if anything, more problematic in HS than classic OT. Recall from (13) 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:

(37) 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're derived from long vowels
   /ji-ʃiːl-uː-na/ jiʃi̱lúːna '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 whole point is that the processes **don't** chain together — underlying A does not map to surface C.

In standard 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 the derivation in (38). (To save space, this derivation begins after stress has already been assigned.)

(38) An unwanted HS derivation

jiʃiːlúːna
↓
GEN
↓
jiʃiːlúːna, jiʃilúːna, jiʃlúːna, …
↓
EVAL
↓
jiʃilúːna
↓
GEN
↓
jiʃilúːna, jiʃlúːna, …
↓
EVAL
↓
jiʃlúːna
↓
GEN
↓
jiʃlúːna, …
↓
EVAL
↓
jiʃlúːna

The input to the penultimate step of the derivation is *jiʃilúːna*, with a short *i* in the second syllable. As far as MAX(Vː) is concerned, this short *i* is no different 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 pages 19ff.).

Another solution 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. See exercise 13 for more on this point.

The most radical solution is to adopt something like OT-CC, which was mentioned on page 9. 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 McCarthy (2007a).

EXERCISES

12. Apply the approach to counterbleeding opacity in (35) and (36) to the *ħa:kʲmi:n* problem. For this to work, you will need to assume that syncope is a two-step process in which a vowel's mora is deleted before its features.
13. The Cairene data in (13) and (37) are in a broad transcription. In a narrower transcription (see (39)), a distinction is made between lax *ɪ*, which is the surface reflex of the underlying short vowel, and tense *i:* and *i*, which are the surface reflexes of the underlying long vowel, even when it has been shortened (Mitchell 1956:10-11, 112).[4] Use this detail to construct a HS analysis that does not require a faithfulness constraint that always looks back to the underlying representation.

(39) Cairene data (narrow transcription — cf. (37))

fíhɪm	'he understood'
fíhmu	'they understood'
jɪʃí:l	'he asks'
jɪʃilú:na	'they ask us'

[4] Final high vowels are also tense, because they are long in underlying representation (McCarthy 2005).

10 Supplement to section 6.6 (page 277)
HS and the too-many-solutions problem

On page 277, *Doing OT* brings up a too-many-solutions problem: why do deletion and assimilation in consonant clusters always target the first consonant? HS offers a new answer to this question (McCarthy 2008a). I will explain how HS deals with the assimilation problem and leave the deletion problem as an exercise for the reader.

The key idea is that place assimilation is a two-step process: first the CODA-COND-violating place feature is deleted, and only then the place feature of the following onset spreads to take its place. Thus, the derivation goes something like this: /pamta/ → *paNta* → *pan͡ta*. (*N* denotes a placeless nasal, and the ligature is used to mark a cluster that shares a single place autosegment.) A natural assumption is that HS's GEN cannot perform feature delinking and linking together in a single step. This assumption categorically rules out the one-step derivation /pamta/ → *pan͡ta*, which has to delink [labial] and spread [coronal].

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. Deleting place from either onset or coda introduces violations of the markedness constraint HAVE-PLACE and the faithfulness constraint IDENT(place), so they have to be ranked below CODA-COND:

(40) Step 1 of /pamta/ → *pan͡ta*

	/pamta/	CODA-COND	HAVE-PLACE	IDENT(place)
a. →	paNta		*	*
b.	pamta	*W	L	L
c.	pamʔa	*W	*	*
d.	pa^{m}n͡ta	*W	L	*

The final candidate is intended to represent the result of spreading [coronal] from *t* to the preceding nasal, while still preserving the nasal's [labial] feature. This candidate is harmonically bounded by the faithful candidate *pamta*.

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

(41) Step 2 of /pamta/ → *pan͡ta*

	paNta	CODA-COND	HAVE-PLACE	IDENT(place)
a. →	pan͡ta			*
b.	paNta		*W	L

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

In standard 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 standard 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 (40) shows. HS yields a more restrictive typology of place assimilation than standard OT does, all else being equal. This more restrictive typology better fits what we actually find in languages.

EXERCISE

14. Building on some of the ideas in the analysis just presented, develop an explanation in HS for why consonant clusters are simplified by deleting the first consonant and not the second: /patka/ → *paka*, **pata*.
15. Does positional faithfulness offer a competing resolution of this too-many-solutions problem? Explain your answer.

11 References

An overview of work on HS in tabular form can be viewed here:
http://works.bepress.com/john_j_mccarthy/102/

This and other material about HS can also be found at the website of the McCarthy-Pater NSF grant:
http://web.linguist.umass.edu/~mccarthy-pater-nsf/

Most of my work on HS is available for download from:
http://works.bepress.com/john_j_mccarthy/

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

Elfner, Emily. (2009) Harmonic serialism and stress-epenthesis interactions in Levantine Arabic. Unpublished paper, University of Massachusetts Amherst, Amherst, MA. Available at http://www.people.umass.edu/eelfner/elfner_arabic.pdf.

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

Hamid, Abdel Halim M. (1984) *A Descriptive Analysis of Sudanese Colloquial Arabic Phonology*. Doctoral dissertation, University of Illinois, Urbana, IL.

Jesney, Karen (to appear) 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*. Amherst, MA: GLSA Publications. Available on Rutgers Optimality Archive, ROA-1018.

Kimper, Wendell (to appear) Locality and globality in phonological variation. *Natural Language & Linguistic Theory*. Available on Rutgers Optimality Archive, ROA-988.

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) OT constraints are categorical. *Phonology* 20: 75-138. Available at http://people.umass.edu/jjmccart/categorical.pdf.

McCarthy, John J. (2005) The length of stem-final vowels in Colloquial Arabic. In Mohammad T. Alhawary and Elabbas Benmamoun (eds.), *Perspectives on Arabic Linguistics XVII-XVIII: Papers from the Seventeenth and Eighteenth Annual Symposia on Arabic Linguistics*, 1-26. Amsterdam: John Benjamins. Available on Rutgers Optimality Archive, ROA-616.

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) The gradual path to cluster simplification. *Phonology* 25: 271-319. doi:10.1017/S0952675708001486.

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

McCarthy, John J. (2010a) An introduction to Harmonic Serialism. *Language and Linguistics Compass* 3: 1-17.

McCarthy, John J. (2010b) Studying Gen. *Journal of the Phonetic Society of Japan* 13: 3-12.

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

McCarthy, John J., Wendell Kimper and Kevin Mullin. (2010) Reduplication in Harmonic Serialism. Unpublished paper, University of Massachusetts Amherst, Amherst.

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

Pater, Joe (to appear) 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.

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

Pruitt, Kathryn. (2008) Iterative foot optimization and locality in stress systems. Unpublished paper, University of Massachusetts Amherst, Amherst, MA. Available on Rutgers Optimality Archive, ROA-999.

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

Samek-Lodovici, Vieri and Alan Prince. (2005) Fundamental properties of harmonic bounding. Unpublished paper, University of London and Rutgers University, London and New Brunswick, NJ. Available on Rutgers Optimality Archive, ROA-785.

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/.

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.