January 2009

# Studying GEN

John J. McCarthy
*University of Massachusetts, Amherst*, jmccarthy@linguist.umass.edu

# Studying GEN[1]

John J. McCarthy

University of Massachusetts Amherst

**Abstract**

In Optimality Theory, phonological patterns are accounted for with output constraints ranked in a hierarchy. There is little explanatory role for a theory of operations, and hence little has been said about the Gen component. This situation has changed with the emergence of a derivational version of Optimality Theory called Harmonic Serialism.

One of the principal differences between Harmonic Serialism and standard Optimality Theory is that Harmonic Serialism's Gen is limited to doing one thing at a time. Harmonic Serialism's analyses and explanations depend on knowing what it means to "do one thing at a time", and that requires a precise theory of Gen.

This article explains Gen's role in Harmonic Serialism and describes two techniques for discovering properties of Gen.

## 1. Introduction

In rule-based phonology, the explanations for phonological patterns come from the theory of operations. Optimality Theory is different. In OT, phonological patterns are accounted for with a hierarchy of markedness and faithfulness constraints. There is little or no explanatory role for a theory of operations in OT.

OT does have an operational component, however. It is called GEN, and its role is to take an underlying representation and transform it into a set of candidate output forms that are evaluated by the constraints. GEN has received little attention in the OT literature precisely because its properties are unimportant in OT's explanations for phonological patterns.

This situation has changed with the emergence of a derivational version of OT called Harmonic Serialism (HS). One of the principal differences between HS and "classic" OT is that HS's GEN is limited to doing one thing at a time. HS's analyses and explanations depend on knowing what it means to "do one thing at a time", and that requires a precise theory of GEN.

Because HS places so much reliance on GEN, it is possible to draw interesting inferences about the properties of GEN in HS. This article describes techniques for studying GEN in HS. It begins (section 2) with a brief overview of the principles of HS, followed (section 3) by an explanation of how such inferences are made, focusing on two techniques that are illustrated with examples in sections 4 and 5.

**2. A brief introduction to Harmonic Serialism**

It is usually assumed that the mapping from underlying to surface forms happens in a single step in Optimality Theory (Prince and Smolensky 1993/2004). This assumption is questioned in recent work on a derivational version of OT called Harmonic Serialism. HS was briefly considered by Prince and Smolensky (1993/2004), but then set aside. Lately, I and others have begun to reexamine HS, finding that it has a number of attractive properties (see Elfner 2009; Jesney 2009; Kimper 2008; McCarthy 2000, 2002, 2007a, b, c, 2008b, c; Pater 2010; Pruitt 2008; Wolf 2008).

HS's differences from classic OT can be described very briefly. As I noted in the introduction, HS's GEN is limited to making one change at a time. Because there can be many changes in the mapping from underlying to surface representations, the output of each pass through GEN and EVAL is submitted as the input to another pass through GEN and EVAL. This *GEN → EVAL → GEN … loop* continues until no further changes are possible. This is the sense in which HS is a derivational version of OT.

For example, Arabic maps underlying /ktub/ to [ʔuktub] 'write!' by epenthesizing glottal stop and a vowel. Since HS's GEN is limited to doing one thing at a time, it cannot epenthesize both of these segments simultaneously. Instead, the derivation requires an intermediate step where the vowel has been epenthesized but the glottal stop has not: <ktub, uktub, ʔuktub>. The tableaux in (1) illustrate. On the first pass through GEN, the input is the underlying form /ktub/ and the candidate set includes faithful [ktub] plus all of the forms that are just one change distant from [ktub]: [uktub], [ktubu], [xtub], and so on. This candidate set is evaluated at step 1, and the winner is [uktub].[2] At step 2, [uktub] becomes the input to another pass through GEN, which produces the candidate set [uktub], [ʔuktub], [uxtub], and so on. The winner in this evaluation is [ʔuktub]. At step 3, [ʔuktub] is the input to GEN, and the candidate set is [ʔuktub], [ʔuktubu], etc. The grammar once again selects [ʔuktub] as the winner. At this point, the derivation has *converged* on a surface form, because no further changes are possible.[3]

(1) Derivation of /ktub/ → [ʔuktub]

Step 1: First pass through GEN → EVAL → GEN … loop

| ktub | *COMPLEX-ONSET | ONSET | DEP |
|---|---|---|---|
| a. → uktub | | 1 | 1 |
| b. ktub | 1 **W** | **L** | **L** |

Step 2: Second pass through GEN → EVAL → GEN … loop

| uktub | *COMPLEX-ONSET | ONSET | DEP |
|---|---|---|---|
| a.→ ʔuktub | | | 1 |
| b. uktub | | 1 **W** | **L** |

Step 3: Third pass through Gen → Eval → Gen … loop: Convergence!

| ʔuktub | *Complex-Onset | Onset | Dep |
|---|---|---|---|
| a. → ʔuktub | | | |
| b. ʔuktubu | | | 1 **W** |

Because Eval applies repeatedly, each step in the derivation <ktub, uktub, ʔuktub> must better satisfy the constraint hierarchy than its predecessor. This property of HS is called *harmonic improvement*. Harmonic improvement is always determined relative to a particular constraint hierarchy that is invariant across all iterations of the Gen → Eval → Gen … loop.

**3. Inferences about Gen in Harmonic Serialism**

What is the theory of Gen? HS's one-change-at-a-time requirement offers an entrée to Gen's internal workings. Empirical arguments can show whether or not a particular mapping requires one or more than one change. This section sketches two ways of making such arguments, and the following sections exemplify them.

Suppose that some language maps underlying /A/ to surface [C]. Suppose too that our theory of Gen in HS does not allow the /A/ → [C] mapping to take place in a single step. Instead, the intermediate form [B] is required, so the full HS derivation is <A, B, C>. From the harmonic improvement property of HS, it follows that the constraint hierarchy of this language must evaluate [B] as more harmonic than [A] and less harmonic than [C]: A ≺ B ≺ C. In Arabic, this ordering is [ktub] ≺ [uktub] ≺ [ʔuktub].

It is sometimes not so obvious that the harmonic ordering A ≺ B ≺ C can be achieved. Perhaps other facts of the language require the harmonic ordering B ≺ A. Or perhaps no member of the universal constraint set Con favors B over A. The analyst might propose adding this constraint to CON, but any such proposal has to be rejected if it does not satisfy established criteria for membership in CON (for which see McCarthy 2008a: chapter 4).

If the harmonic ordering A ≺ B ≺ C cannot be achieved, either because it is incompatible with other facts of the language or because it would require an unsupportable constraint, then HS leaves us only one other possibility: our initial assumption about GEN was incorrect. Whereas we previously thought that GEN requires two steps to get from /A/ to [C], we have now been forced to conclude that this mapping is accomplished in a single step. This technique makes it possible to draw inferences about the properties of GEN in HS.

There is a second technique. Suppose no known language maps /A/ to [C], and we are confident that this gap is principled rather than accidental. If unfaithful [C] is more harmonic than faithful [A] under some ranking(s) of CON and no other candidate is more harmonic than [C], then we have what is known as a "too many repairs" (TMR) problem: violations of the markedness constraint that disfavors [A] are mysteriously never "repaired" by changing /A/ into [C]. This typological gap can be explained in HS, however, if

(i) GEN is assumed to require an intermediate step [B] between /A/ and [C], and

(ii) [B] is not more harmonic than [A] under any ranking of CON that maps [B] to [C].

In short, the /A/ → [C] mapping is unattested because it requires an intermediate step [B] that does not improve harmony.

To sum up, we can draw inferences about GEN by studying situations where a mapping /A/ → [C] might or might not require an intermediate step [B], depending on how GEN is defined. There are two conditions to consider:

(i) **Attested mapping without intermediate step.** If /A/ → [C] is attested in some language, and if there is no [B] that is less harmonic than [A] and more harmonic than [C], then GEN must be defined so that [B] is not a required intermediate step. In other words, GEN must include [C] in the candidate set from /A/, because there is no possible two-step path from /A/ to [C]. This mode of inference about GEN is the topic of section 4.

(ii) **Unattested mapping with intermediate step.** If /A/ → [C] is an unattested mapping and a principled typological gap, then an explanation is required. HS offers a novel explanation for such gaps: define GEN so that [C] is not in the candidate set from /A/, and define CON so that there is no [B] that is more harmonic than [A] and less harmonic than [C]. Examples of this type are the topic of section 5.

## 4. Attested mapping without intermediate step

If the mapping /A/ → [C] is attested, and if the only harmonically improving path from /A/ to [C] is the direct route with no intermediate steps, then GEN must be defined so that [C] is in /A/'s candidate set. In this way, attested mappings supply evidence about what it means for GEN to "do one thing at a time". I will discuss an example of syncope, using it to address the question of whether a syncope process and the resyllabification process that it triggers occur in different derivational steps.

In Cairene Arabic, syncope affects short high vowels in unstressed non-final open syllables in internal and external sandhi (see (2)). The syllable boundaries (indicated by periods/full stops) show that the erstwhile onset to the deleted vowel is resyllabified as coda to the preceding vowel.

(2) 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? Specifically, which of these is the correct derivation of the first example in (2): <…, ˈwi.ħi.ʃa, ˈwiħ.ʃa>, with simultaneous syncope and resyllabification, or <…, ˈwi.ħi.ʃa, ˈwi.ħ.ʃa, ˈwiħ.ʃa>, with sequential syncope and resyllabification? I will argue that syncope and resyllabification must be simultaneous by showing that the sequential derivation leads to a ranking paradox.

The evidence comes from the observation that syncope does not affect a vowel that is preceded by a consonant cluster: /ħagar kibiːr/ → [ˈħa.gar.k**i**.ˈbiːr] 'big stone'. Syncope is blocked in this situation because the resulting triconsonantal cluster would be unsyllabifiable. If syncope and resyllabification require separate derivational steps, then the only way to block syncope in this example is to rule out all of the possible dispositions of the [k] other than adjoining it to a nearby syllable. Those dispositions are listed in (3). All of them are prohibited as the immediate output of syncope, so the constraints against them — Parse-Segment, Headedness(σ), and *Nuc/Cons — have to dominate the constraint that favors syncope.

(3) Forbidden results of syncope after a cluster

*ˈ$(\text{ħa})_\sigma$ $(\text{gar})_\sigma$ k $(\text{ˈbiːr})_\sigma$ [k] is unsyllabified.

*ˈ$(\text{ħa})_\sigma$ $(\text{gar})_\sigma$ $(\text{k\_})_\sigma$ $(\text{ˈbiːr})_\sigma$ [k] is the onset of a degenerate syllable.

*ˈ$(\text{ħa})_\sigma$ $(\text{gar})_\sigma$ $(\text{k̩})_\sigma$ $(\text{ˈbiːr})_\sigma$ [k] is syllabic.

We will now attempt to extend this analysis to the examples in (2) where syncope is not blocked because only one consonant precedes. When syncope affects [ˈwi.ħi.ʃa], the immediate output prior to resyllabification has to be one of the forms in (4). One of these forms has to be permitted, because syncope does occur in this case. At the next step, [ħ] resyllabifies as coda of the preceding syllable.

(4) Permitted result of syncope after a single consonant

ˈ$(\text{wi})_\sigma$ ħ $(\text{ʃa})_\sigma$ [ʃ] is unsyllabified, *or*

ˈ$(\text{wi})_\sigma$ $(\text{ħ \_})_\sigma$ $(\text{ʃa})_\sigma$ [ʃ] is the onset of a degenerate syllable, *or*

ˈ$(\text{wi})_\sigma$ $(\text{ħ̩})_\sigma$ $(\text{ʃa})_\sigma$ [ʃ] is syllabic.

We have now reached a contradiction. If GEN requires separate steps to effect syncope and resyllabification, then the immediate result of deleting the vowel in a CV syllable is a C that is unsyllabified, syllabified in a degenerate syllable, or syllabified as a nucleus. The evidence in (3) requires that all of these configurations be prohibited by markedness constraints ranked higher than the constraint that favors syncope. The evidence in (4), however, shows that one of these configurations has to be allowed, so one of those markedness constraints has to be ranked below the constraint that favors syncope.

This contradiction can be seen formally by comparing tableaux (5) and (6). Because this language has syncope, the faithfulness constraint MAX-V has to be dominated by some syncope-favoring markedness constraint — here, that constraint is WEAK<*i* (McCarthy 2007a: 169-174).[4] And because there is no syncope after a cluster, WEAK<*i* must itself be dominated by the markedness constraints against unsyllabified segments, headless syllables, and syllabic consonants. This ranking is shown in (5). In (6), I made the arbitrary decision that [ˈ(wi)$_{\sigma}$ (ħ _)$_{\sigma}$ (ʃa)$_{\sigma}$] is the winner, but the argument could also be made with one of the other forms in (4). For this form to win, the constraint against headless syllables has to be ranked *below* WEAK<*i*. That is the contradiction.

(5) No syncope in <…, ˈħa.gar.ki.ˈbiːr>

| | ˈ(ħa)$_{\sigma}$ (gar)$_{\sigma}$ (ki)$_{\sigma}$ ˈ(biːr)$_{\sigma}$ | PRS-SEG | HEAD(σ) | *NUC/CONS | WEAK<*i* | MAX-V |
|---|---|---|---|---|---|---|
| a.→ | ˈ(ħa)$_{\sigma}$ (gar)$_{\sigma}$ (ki)$_{\sigma}$ ˈ(biːr)$_{\sigma}$ | | | | 1 | |
| b. | ˈ(ħa)$_{\sigma}$ (gar)$_{\sigma}$ k ˈ(biːr)$_{\sigma}$ | 1 **W** | | | L | 1 **W** |
| c. | ˈ(ħa)$_{\sigma}$ (gar)$_{\sigma}$ (k_)$_{\sigma}$ ˈ(biːr)$_{\sigma}$ | | 1 **W** | | L | 1 **W** |
| d. | ˈ(ħa)$_{\sigma}$ (gar)$_{\sigma}$ (k̩)$_{\sigma}$ ˈ(biːr)$_{\sigma}$ | | | 1 **W** | L | 1 **W** |

(6) Syncope in <…, ˈwi.ħi.ʃa, ˈwi.ħ.ʃa, …>

| | ˈ(wi)$_{\sigma}$ (ħi)$_{\sigma}$ (ʃa)$_{\sigma}$ | PRS-SEG | *NUC/CONS | WEAK<*i* | HEAD(σ) | MAX-V |
|---|---|---|---|---|---|---|
| a.→ | ˈ(wi)$_{\sigma}$ (ħ _)$_{\sigma}$ (ʃa)$_{\sigma}$ | | | | 1 | 1 |
| b. | ˈ(wi)$_{\sigma}$ (ħi)$_{\sigma}$ (ʃa)$_{\sigma}$ | | | 1 **W** | **L** | **L** |
| c. | ˈ(wi)$_{\sigma}$ ħ (ʃa)$_{\sigma}$ | 1 **W** | | | | 1 |
| d. | ˈ(wi)$_{\sigma}$ (ħ̩)$_{\sigma}$ (ʃa)$_{\sigma}$ | | 1 **W** | | | 1 |

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, when there is no way of knowing whether it will eventually be possible to attach the stray consonant in a nearby syllable, as actually occurs in [ˈ(wiħ)$_{\sigma}$ (ʃa)$_{\sigma}$].

The situation is different if GEN can perform syncope and resyllabification in a single step. To see why, we will first look at the single-consonant case and then at the cluster case. Under this revised definition of GEN, the candidate set from [ˈwi.ħi.ʃa] includes all of the candidates in (6) plus [ˈ(wiħ)$_\sigma$ (ʃa)$_\sigma$], with resyllabification. This is the candidate that wins, as shown in tableau (7). It wins because all of the possible results of syncope other than resyllabification are ruled out by undominated consonants.

(7) Syncope in <…, ˈwi.ħi.ʃa, ˈwiħ.ʃa>

| | ˈ(wi)$_\sigma$ (ħi)$_\sigma$ (ʃa)$_\sigma$ | PRS-SEG | HEAD(σ) | *NUC/CONS | WEAK<*i* | MAX-V |
|---|---|---|---|---|---|---|
| a.→ | ˈ(wiħ)$_\sigma$ (ʃa)$_\sigma$ | | | | | 1 |
| b. | ˈ(wi)$_\sigma$ (ħi)$_\sigma$ (ʃa)$_\sigma$ | | | | 1 **W** | **L** |
| c. | ˈ(wi)$_\sigma$ ħ (ʃa)$_\sigma$ | 1 **W** | | | | 1 |
| d. | ˈ(wi)$_\sigma$ (ħ _)$_\sigma$ (ʃa)$_\sigma$ | | 1 **W** | | | 1 |
| e. | ˈ(wi)$_\sigma$ (ħ̩)$_\sigma$ (ʃa)$_\sigma$ | | | 1 **W** | | 1 |

The same undominated constraints that rule out options other than resyllabification are also at work in evaluating the candidates derived from [ˈħa.gar.ki.ˈbiːr]. In this case, though, resyllabification also has to be ruled out as an option, and that can be done by including NO-COMPLEX-ONSET and NO-COMPLEX-CODA among the undominated constraints, as shown in tableau (8).

(8) No syncope in < …, ˈħa.gar.ki.ˈbiːr >

| | ˈ(ħa)$_\sigma$ (gar)$_\sigma$ (ki)$_\sigma$ (ˈbiːr)$_\sigma$ | PRS-SEG | HEAD (σ) | *NUC/CONS | NO-COMP-COD | NO-COMP-ONS | WEAK < *i* | MAX-V |
|---|---|---|---|---|---|---|---|---|
| a.→ | ˈ(ħa)$_\sigma$ (gar)$_\sigma$ (ki)$_\sigma$ (ˈbiːr)$_\sigma$ | | | | | | 1 | |
| b. | ˈ(ħa)$_\sigma$ (gar)$_\sigma$ k (ˈbiːr)$_\sigma$ | 1 **W** | | | | | **L** | 1 **W** |
| c. | ˈ(ħa)$_\sigma$ (gar)$_\sigma$ (k_)$_\sigma$ (ˈbiːr)$_\sigma$ | | 1 **W** | | | | **L** | 1 **W** |
| d. | ˈ(ħa)$_\sigma$ (gar)$_\sigma$ (k̩)$_\sigma$ (ˈbiːr)$_\sigma$ | | | 1 **W** | | | **L** | 1 **W** |
| e. | ˈ(ħa)$_\sigma$ (gark)$_\sigma$ (ˈbiːr)$_\sigma$ | | | | 1 **W** | | **L** | 1 **W** |
| f. | ˈ(ħa)$_\sigma$ (gar)$_\sigma$ (ˈkbiːr)$_\sigma$ | | | | | 1 **W** | **L** | |

In this revised analysis, there is no look-ahead problem because the candidate set provided by GEN already "looks ahead" to the consequences of resyllabification, because syncope and resyllabification can co-occur in a single derivational step.

This example illustrates a general method for discovering the properties of GEN in HS. 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 at hand, 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 a headless syllable after a closed syllable, then it would be possible to segregate the syncope and resyllabification steps. Such a constraint would allow, say, [ˈ(wi)$_\sigma$ (ħ_)$_\sigma$ (ʃa)$_\sigma$] but not its counterpart [ˈ(ħa)$_\sigma$ (gar)$_\sigma$ (k_)$_\sigma$ (ˈbiːr)$_\sigma$]. The argument above relies on the assumption that no such constraints exist — an assumption that can be tested empirically using techniques like those described in McCarthy (2008a: chapter 4).

Although my focus here is on explaining a technique for studying Gen rather than drawing conclusions about Gen itself, I would be remiss if I did not consider 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 resyllabification is a faithful operation. It is therefore possible to combine syncope and resyllabification into a single derivational step.[5]

This hypothesis about Gen can be tested by applying the techniques in this paper 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.

**5. Unattested mapping with intermediate step**

Suppose /A/ never maps to [C] in any language, and we have reason to believe that this typological gap is principled rather than accidental. In classic OT, any explanation for this observation relies on harmonic bounding (Samek-Lodovici and Prince 2005): output [C] is not the optimum for input /A/ under any ranking of CON. The “too many repairs” problem mentioned in section 3 is the existence of cases where /A/ → [C] is unattested but [C] is not harmonically bounded under an otherwise trustworthy theory of CON.

HS adds another possible explanation for typological gaps: the nonexistence of a harmonically improving path from /A/ to [C] (McCarthy 2007a, b, c, 2008b). This mode of explanation has been shown to solve many too-many-repairs problems. In the course of solving those problems, specific assumptions about GEN are necessary, so studying such problems is a way of studying GEN.

For example, nothing said so far answers the question of whether GEN permits multiple applications of the *same* operation in a single step. In the hypothetical Arabic word /kat**i**bat**i**nu/, are both of the highlighted [i] vowels deleted on a single pass through GEN, or do they require two separate passes? A particular typological gap leads to an argument that multiple passes are required, and so GEN must be limited to a *single application* of an operation.

This typological gap involves observed limits on the situations where an unbounded string of segments can be deleted. These limits follow, I will argue, if GEN can delete only one segment at a time. An unbounded string can therefore delete only if each individual segmental deletion improves harmony. The non-existent cases are those where harmony improves only if a multisegmental string is deleted in a single step — an impossibility under this theory of GEN, though certainly possible in classic OT.

As I just noted, deletion of unbounded segmental strings is possible only if each individual deletion improves harmony. Many existing cases of this type fall under the rubric of *licensing*. A segment is licensed through attachment to higher-level prosodic structure, typically a syllable. Since the markedness constraint PARSE is violated once by each unsyllabified segment, deleting an unsyllabified segment improves performance on this constraint.

In Diola Fogny, for example, some consonants cannot be syllabified because of conditions on licit codas. There are examples where one or two consonants are deleted (see (9)), but presumably three or more consonants could also be deleted in the same context.

(9) Cluster simplification in Diola Fogny (Ito 1986; Kiparsky 1973; Sapir 1965)

| Underlying | Surface | |
|---|---|---|
| uʤuk-ʤa | u.ʤu.ʤa | 'if you see' |
| let-ku-ʤaw | le.ku.ʤaw | 'they won't go' |
| e-rent-rent | e.re.rent | 'it is light' |

Under the assumption that GEN can execute only one instance of an operation at a time, the derivation of the last item in (9) would have to proceed by deleting the unsyllabified [nt] cluster one segment at a time: <$(e)_\sigma$ $(re)_\sigma$ nt $(rent)_\sigma$, $(e)_\sigma$ $(re)_\sigma$ n $(rent)_\sigma$, $(e)_\sigma$ $(re)_\sigma$ $(rent)_\sigma$>.[6] Each step improves harmony over its predecessor because it better satisfies PARSE, as (10) shows.

(10) Derivation of /erentrent/ → [e.re.rent]

Step 1: Consonant deletion

| | /erentrent/ | DEP | PARSE | MAX |
|---|---|---|---|---|
| a.→ | $(e)_\sigma$ $(re)_\sigma$ n $(rent)_\sigma$ | | 1 | 1 |
| b. | $(e)_\sigma$ $(re)_\sigma$ nt $(rent)_\sigma$ | | 2 **W** | **L** |
| c. | $(e)_\sigma$ $(re)_\sigma$ $(nit)_\sigma$ $(rent)_\sigma$ | 1 **W** | **L** | **L** |

Step 2: Consonant deletion

| | $(e)_\sigma$ $(re)_\sigma$ n $(rent)_\sigma$ | DEP | PARSE | MAX |
|---|---|---|---|---|
| a.→ | $(e)_\sigma$ $(re)_\sigma$ $(rent)_\sigma$ | | | 1 |
| b. | $(e)_\sigma$ $(re)_\sigma$ n $(rent)_\sigma$ | | 1 **W** | **L** |
| c. | $(e)_\sigma$ $(re)_\sigma$ $(ni)_\sigma$ $(rent)_\sigma$ | 1 **W** | **L** | **L** |

Step 3: Convergence

| | $(e)_\sigma$ $(re)_\sigma$ $(rent)_\sigma$ | DEP | PARSE | MAX |
|---|---|---|---|---|
| a.→ | $(e)_\sigma$ $(re)_\sigma$ $(rent)_\sigma$ | | | |
| b. | $(e)_\sigma$ $(re)_\sigma$ $(ren)_\sigma$ | | | 1 **W** |

Deletion of unbounded segmental strings for licensing reasons is also common in prosodic morphology. Japanese hypocoristics, exemplified in (11), are the classic example. Segments are licensed by association with a bimoraic foot template. Segments that are not associated with the template are deleted because they violate a template-specific version of PARSE. Therefore, this deletion also improves harmony as it deletes segments one at a time: <midoɾi, $(mido)_{ft}$ ɾi, $(mido)_{ft}$ i, $(mido)_{ft}$>.

(11) Japanese hypocoristics (Mester 1990; Poser 1984, 1990)

| | |
|---|---|
| midoɾi | mido-tʃan |
| kinsuke | kin-tʃan |
| wasabuɾoː | wasa-tʃan |

We have seen that the assumption about GEN under consideration — that it permits no more than one instance of an operation at a time — is compatible with deletion of unbounded strings under some circumstances. But there are also circumstances where deletion of unbounded strings should be impossible. These circumstances arise whenever the constraint(s) favoring deletion require deletion of more than one segment at a time to achieve harmonic improvement.

The following example involves two markedness constraints, FINAL-C and CODA/son, dominating MAX. I will first introduce these constraints and then compare their effects in classic OT and HS.

The constraint FINAL-C requires words to end in a consonant (Gafos 1998; Krämer 2003b; McCarthy 1993; McCarthy and Prince 1994; Orie and Bricker 2000: 299-300; Swets 2004; Wiese 2001; Wiltshire 2003). This constraint has various effects, such as favoring final consonant epenthesis in Yucatec Maya Spanish loans (12) or final vowel deletion in Yapese (13).

(12) Final consonant epenthesis in Yucatec Maya (Krämer 2003a, b)

| Spanish original | Yucatec loan | |
|---|---|---|
| *catarro* | kaˈtaːroh | ‘cold/cough’ |
| *escuela* | ʔesˈkweːlah | ‘school’ |

(13) Apocope in Yapese (Jensen 1977; Krämer 2003a; Piggott 1991)

| Underlying | Surface | |
|---|---|---|
| ʔadi | ˈʔæːð | ‘liver’ |
| ʔadi-gu | ʔaˈðiːg | ‘my liver’ |
| ʔadi-mu | ʔaˈðiːm | ‘your (sg.) liver’ |
| ʔadi-na | ʔaˈðiːn | ‘his liver’ |

The constraint CODA/son requires coda consonants, including word-final ones, to be sonorant. The need for this constraint is well established in syllable theory (Zec 1995). For example, in Yidiny it blocks apocope: /gindanu/ → [ginˈdaːn] ‘moon’ vs. /gudaga/ → [guˈdaːga] ‘dog’.

Together, FINAL-C and CODA/son make an unwelcome typological prediction in classic OT. If these constraints both dominate MAX, then classic OT predicts the existence of a language that matches the following generalization: All segments are deleted between the rightmost sonorant consonant and the right edge of the word; but if a word already ends in a sonorant consonant, or if it contains no sonorant consonants, then nothing is deleted. Tableau (14) illustrates with the hypothetical underlying form /sanata/. This behavior is neither attested nor plausible, yet the grammar that leads to it is not a bit contrived — well-established constraints are interacting in typical ways.

(14) An unwanted effect of FINAL-C and CODA/son in classic OT

| /sanata/ | FINAL-C | CODA/son | MAX |
|---|---|---|---|
| a. → san | | | 3 |
| b. sanata | 1 **W** | | **L** |
| c. sanat | | 1 **W** | 1 **L** |
| d. sana | 1 **W** | | 2 **L** |

On the other hand, if GEN is limited to one instance of an operation at a time, then these same constraints will not predict this unwanted result in HS. What they do predict in HS depends on how they are ranked. If FINAL-C is ranked higher, as in (15), then all final vowels delete, regardless of whether the preceding consonant is a sonorant, but then deletion proceeds no further. If CODA/son is ranked higher, as in (16), then final vowels delete only after sonorants; otherwise, nothing happens.[7]

(15) If FINAL-C dominates CODA/son

Step 1 — Apocope

| /sanata/ | FINAL-C | CODA/son | MAX |
|---|---|---|---|
| a. → sa.nat | | 1 | 1 |
| b. sa.na.ta | 1 **W** | **L** | **L** |

Step 2 — Convergence

| sa.nat | FINAL-C | CODA/son | MAX |
|---|---|---|---|
| a. → sa.nat | | 1 | |
| b. sa.na | 1 **W** | **L** | 1 **W** |

(16) If CODA/son dominates FINAL-C

Step 1 — Convergence

| /sanata/ | CODA/son | FINAL-C | MAX |
|---|---|---|---|
| a. → sa.na.ta | | 1 | |
| b. sa.nat | 1 **W** | **L** | 1 **W** |

In neither case does the presence of the sonorant consonant [n] earlier in the word have any effect on the outcome. If GEN is limited to deleting one segment at a time, then HS cannot produce the highly nonlocal and implausible pattern in (14). This is not only a liability of classic OT and an advantage of HS; it is also the basis for an inference about GEN. “Do one thing at a time” must refer to one token, not type, of an operation, and the operations may include deletion of a single segment but not deletion of a multisegmental string.

In the previous section, we saw how different definitions of Gen lead to different amounts of information being available to the constraint hierarchy at each derivational step. That way of looking at Gen is also relevant to this example. If Gen can delete only one segment at a time, then the information that /sanata/ contains an attractive word-final consonant, [n], is not available at the point in the derivation where the decision about further deletion has to be made. This is the crucial difference between classic OT and HS: in classic OT, one of the competing candidates has that attractive word-final consonant, but in HS that candidate is not in competition. The candidate set is the way that Gen makes information available to the constraint hierarchy, and the candidate set may be quite different in classic OT and HS.

Of course, the main point here is not so much to draw this particular inference about Gen, but rather to describe a general technique for drawing inferences about Gen in HS. This technique is based on typological observations of the following form: the mapping /A/ → [C] is a linguistically significant typological gap, but this mapping is not harmonically bounded. If Gen is defined so it interposes an intermediate step [B] between /A/ and [C], and if the derivation <A, B, C> is not harmonically improving under any ranking of Con, then this gap has a principled explanation in HS.

The technique described here for studying Gen and explaining typological gaps is broadly applicable. These are some other examples:

- A typological gap involving long-distance metathesis can be explained by assuming that Gen can move a segment across only one segment at a time (McCarthy 2007b).
- A typological gap involving long-distance autosegmental spreading can be explained by assuming that Gen can spread a feature to only one segment at a time (McCarthy 2007b).

- A typological gap involving the generalized trochee stress pattern can be explained by assuming that GEN assigns only one metrical foot at a time (Pruitt 2008).

Finally, it is interesting to note that these examples, as well as the one above, involve processes that would be described as iterative in rule-based phonology. In this way, HS connects with an important concern of phonological research in the early 1970s: whether a rule is applied simultaneously at every locus where its structural description is met (Anderson 1974; Chomsky and Halle 1968), or whether it applies in directional iterative fashion (Howard 1972; Johnson 1972; Kenstowicz and Kisseberth 1977; Lightner 1972).

**6. Conclusion**

My goal in this article has been to describe some general methods for reaching inferences about GEN in Harmonic Serialism. In classic OT, analyses and explanations are exquisitely dependent on the nature of CON. In HS, CON is equally important, but the nature of GEN is also crucial.

GEN is important in HS because it controls the availability of information to the constraint hierarchy at each step of the derivation. In section 4, I discussed a case where one way of defining GEN supplies too little information for a successful analysis of syncope. In section 5, I discussed a case where one way of defining GEN supplies too much information, since it permits an analysis of a non-existent pattern. Understanding how GEN regulates the flow of information in an HS derivation is the key to discovering its properties.

**References**

Albright, Adam (2001) "The emergence of the marked: Root-domain markedness in Lakhota," Unpublished paper. Linguistic Society of America Annual Meeting. [Available at http://web.mit.edu/albright/www/papers/Albright-EmergenceOfTheMarked-LSA04.pdf.]

Anderson, Stephen R. (1974) *The Organization of Phonology*. New York: Academic Press.

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

Chomsky, Noam, and Halle, Morris (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* (pp. 317-336). Dordrecht: Foris.

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

Gafos, Adamantios (1998) "Eliminating long-distance consonantal spreading." *Natural Language & Linguistic Theory* 16, 223-278.

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

Hock, Hans (1991) *Principles of Historical Linguistics (2nd ed.)*. Berlin: Mouton de Gruyter.

Howard, Irwin (1972) *A Directional Theory of Rule Application in Phonology*. Unpublished Doctoral dissertation, MIT, Cambridge, MA.

Ito, Junko (1986) *Syllable Theory in Prosodic Phonology*. Unpublished Doctoral dissertation, University of Massachusetts Amherst.

Jensen, John T. (1977) *Yapese Reference Grammar*. Honolulu: University Press of Hawaii.

Jesney, Karen (2009) "Positional faithfulness, non-locality, and the Harmonic Serialism solution." *Proceedings of NELS 39*. [Available on Rutgers Optimality Archive, ROA-1018.]

Johnson, C. Douglas (1972) *Formal Aspects of Phonological Description*. The Hague: Mouton.

Kenstowicz, Michael, and Kisseberth, Charles (1977) *Topics in Phonological Theory*. New York: Academic Press.

Kimper, Wendell (2008) "Local optionality and harmonic serialism," Unpublished paper. University of Massachusetts Amherst. [Available on Rutgers Optimality Archive, ROA-988.]

Kiparsky, Paul (1973) "'Elsewhere' in phonology". In Stephen R. Anderson, and Paul Kiparsky (eds.), *A Festschrift for Morris Halle* (pp. 93-106). New York: Holt, Rinehart and Winston.

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

Krämer, Martin (2003a) "The last consonant," Unpublished paper. University of Ulster. [Handout of talk presented at LAGB Spring Meeting, 14-16 April 2003. Available at http://www.hum.uit.no/a/kraemer/lagb03-ho-mk.pdf.]

Krämer, Martin (2003b) "What is wrong with the right side? Edge (a)symmetries in phonology and morphology," Unpublished paper. University of Ulster. [Available on Rutgers Optimality Archive, ROA-576.]

Lightner, Theodore (1972) *Problems in the Theory of Phonology*. Edmonton: Linguistic Research, Inc.

McCarthy, John J. (1993) "A case of surface constraint violation." *Canadian Journal of Linguistics* 38, 169-195.

McCarthy, John J. (2000) "Harmonic serialism and parallelism". In Masako Hirotani (ed.), *Proceedings of the North East Linguistics Society 30* (pp. 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* (pp. 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* (pp. 203-231). Berlin and New York: Mouton de Gruyter. [Available on Rutgers Optimality Archive, ROA-844.]

McCarthy, John J. (2007c) "Slouching towards optimality: Coda reduction in OT-CC". In Phonological Society of Japan (ed.), *Phonological Studies 10* (Vol. 10, pp. 89-104). Tokyo: Kaitakusha. [Available on Rutgers Optimality Archive, ROA-878.]

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., and Prince, Alan (1994) "The emergence of the unmarked: Optimality in prosodic morphology". In Mercè Gonzàlez (ed.), *Proceedings of the*

*North East Linguistic Society 24* (pp. 333-379). Amherst, MA: GLSA Publications. [Available on Rutgers Optimality Archive, ROA-13.]

Mester, Armin (1990) "Patterns of truncation." *Linguistic Inquiry* 21, 475-485.

Orie, Olanike Ola, and Bricker, Victoria R. (2000) "Placeless and historical laryngeals in Yucatec Maya." *International Journal of American Linguistics* 66, 283-317.

Pater, Joe (2010) "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.

Piggott, G. L. (1991) "Apocope and the licensing of empty-headed syllables." *The Linguistic Review* 8, 287-318.

Poser, William J. (1984) "Hypocoristic formation in Japanese". In M. Cobler, S. MacKaye, and M. Wescoat (eds.), *The Proceedings of the West Coast Conference on Formal Linguistics 3* (pp. 218-229). Stanford: Stanford Linguistic Association.

Poser, William J. (1990) "Evidence for foot structure in Japanese." *Language* 66, 78-105.

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* (pp. 269-304). Amherst, MA: GLSA. [Available on Rutgers Optimality Archive, ROA-562.]

Prince, Alan, and Smolensky, Paul (1993/2004) *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden, MA, and Oxford, UK: Blackwell. [Revision of 1993 technical report, Rutgers University Center for Cognitive Science. Available on Rutgers Optimality Archive, ROA-537.]

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

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

Sapir, J. D. (1965) *A Grammar of Diola-Fogny*. Cambridge: Cambridge University Press.

Sohn, Ho-min (1975) *Woleaian Reference Grammar*. Honolulu: University of Hawaii Press.

Swets, Francine (2004) *The Phonological Word in Tilburg Dutch.* Unpublished Doctoral dissertation, Universiteit van Amsterdam, Amsterdam. [Available at http://www.lotpublications.nl/publish/issues/Swets/index.html.]

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

Wiese, Richard (2001) "The structure of the German vocabulary: Edge marking of categories and functional considerations." *Linguistics* 39, 95-115.

Wiltshire, Caroline (2003) "Beyond codas: Word and phrase-final alignment". In Caroline Féry, and Ruben van de Vijver (eds.), *The Syllable in Optimality Theory* (pp. 254-268). Cambridge: Cambridge University Press.

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

Zec, Draga (1995) "Sonority constraints on syllable structure." *Phonology* 12, 85-129.

**Notes**

[1] This research was supported by grant BCS-0813829 from the National Science Foundation to the University of Massachusetts Amherst. For their comments and assistance, I am grateful to the other participants in that project: Emily Elfner, Karen Jesney, Wendell Kimper, Kevin Mullin, Joe Pater, Kathryn Pruitt, and Brian Smith.

[2] As usual in the OT literature, tableaux omit many losing candidates and the constraints that cause them to lose.

[3] Tableaux are in the comparative format introduced by Prince (2002). The integers just are the count of violation marks (asterisks). Ws and Ls appear only in loser rows, where they indicate how each loser performs relative to the winner on each constraint. A W means that the constraint favors the winner over the loser. Ls mark the opposite favoring relation. If a cell in a loser row contains neither W nor L, then the loser ties with the winner on that constraint. One advantage of comparative tableaux is that they present constraint ranking relations very transparently: in a properly ranked comparative tableau, every L has a W somewhere to its left across a solid line.

[4] WEAK<*i* assigns a violation mark for every weak syllable with a nucleus whose sonority is equal to or greater than that of a [+high] vowel. A syllable is "weak" if it has shorter overall duration because it meets the following criteria: it is open, its vowel is short, it is unstressed, and it is non-final. See McCarthy (2007a: 169-174).

[5] 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 lacks of faithfulness.

[6] The choice between deleting [n] or [t] first is arbitrary.

[7] There are actual languages with apocope patterns similar to those in (15) and (16). Derivation (15) is analogous to cases where apocope produces codas that the language otherwise forbids, as in Lakhota (Albright 2001), the history of modern Persian (Hock 1991: 116), and perhaps Woleaian, if final vowel devoicing is a type of apocope (Sohn 1975). The pattern in (16) is exemplified in Yidiny, mentioned earlier.