Locality and Globality in Phonological Variation*

Wendell A. Kimper

Abstract

This paper explores the consequences of Harmonic Serialism (HS), a derivational variant of OT, for local and global phonological variation. Variation is local when each locus within a single form may vary independently, as in French schwa deletion. Variation is global when all loci within a single form must covary, as in labial (de)voicing in Warao. Within the framework of Optimality Theory (OT) with parallel evaluation, only global variation is predicted to exist. In this paper, I show how implementing a multiple-rankings theory of phonological variation within HS accounts for both local and global variation.

1 Introduction

Harmonic Serialism (HS) differs from parallel versions of Optimality Theory (OT) (Prince and Smolensky, 1993/2004) in that, instead of a single pass through GEN and EVAL, optimality is evaluated derivationally — changes are made one at a time, and a form undergoes a new pass through GEN and EVAL with each change.

This difference between parallel and serial versions of OT has implications for theories of phonological variation which make use of variable constraint ranking. Specifically, parallel and serial OT make different predictions about when variation should be local and when it should be global. In this paper, I propose Serial Variation — a combination of Harmonic Serialism and a multiple-rankings theory of variation — and show that its predictions account for the attested facts with respect to local and global variation.

Local variation is defined as in (1):

(1) **Local Variation**: for a form with multiple loci of an optional or variable process, the choice at each locus may be independent from the choices at other loci.

For example, schwa in French may optionally delete; in a form with multiple schwas, some may be preserved while others undergo deletion (Dell, 1973; Riggle and Wilson, 2005):

(2) *envie de te le demander* 'feel like asking you for it'

a. ɑ̃vi də tə lə dəmɑ̃de

b. ɑ̃vi d_ tə lə dəmɑ̃de

*This research was supported by grant BCS-0813829 from the National Science Foundation to the University of Massachusetts Amherst.

c. ãvi də t_ lə dəmãde

d. ãvi də tə l_ dəmãde

e. ãvi də tə lə d_mãde

f. ãvi d_ tə l_ dəmãde

g. ãvi də t_ lə d_mãde

The preference to delete schwa and the preference to preserve it are instantiated simultaneously in forms like (2b-e). That is, the choice of deletion or preservation is made for *each schwa*, not for the domain as a whole.

Global variation is defined as in (3):

(3) **Global Variation**: for a form with multiple loci of an optional or variable process, the choice must be the same at all loci.

For example, labial obstruents in Warao are variably voiced or voiceless, but all labial obstruents in a word must covary (Osborn, 1966; Riggle and Wilson, 2005):

(4) /paro + parera/ 'weak'

a. paroparera

b. barobarera

c. *parobarera

d. *baroparera

Variation in Warao is global, because the choice of voiced or voiceless obstruents is consistent throughout the domain, and cannot differ at separate loci.

In parallel OT, because all loci are evaluated simultaneously under the same conditions (the same constraint ranking), all variation is predicted to be global. Candidates with local variation are harmonically bounded (Vaux, 2003; Riggle and Wilson, 2005).

Serial Variation permits an account of local variation, as (Pater, 2007) notes. Because HS restricts Gen to producing only a single change, each locus of variation is evaluated in a separate step. Furthermore, because a new instantiation of Eval provides a new opportunity to select a ranking of variably ordered constraints, the conditions under which each locus is evaluated may differ. This means that a separate choice is made at each locus, giving rise to local variation.

Under certain conditions, Serial Variation also predicts global variation. A locally variable form must not only be chosen as optimal at a particular step, but the derivation must also be able to *converge* on that form. With certain types of constraint interaction, converging on locally variable forms is impossible, and globality is predicted in those cases.

The paper is organized as follows. The remainder of this introduction provides necessary background to Harmonic Serialism (Section 1.1), discusses the relevant properties of the theories of variation assumed (Section 1.2), and highlights the difference between serial and parallel evaluation in the context of variation (Section 1.3).

Section 2 shows how Serial Variation accounts for variation in Bengali Minor Phrase assignment, where strict inheritance of prosodic structure permits locality. Section 3 shows how a serial account predicts locality in prosodically conditioned deletion of French schwa. Section 4 illustrates the conditions under which Serial Variation predicts global variation, and provides an account of labial (de)voicing in Warao. Finally, Section 5 compares Serial Variation with alternative accounts of local variation.

1.1 Harmonic Serialism

Harmonic Serialism (HS) is a derivational variant of Optimality Theory briefly considered by Prince and Smolensky (1993/2004) and discussed more thoroughly in McCarthy (2000, 2007b, et seq.). The fundamental mechanics of the theory are the same as in classic OT: Gen operates on an input to produce a set of candidates, and the well-formedness of these candidates is evaluated by the language's constraint hierarchy at Eval, with the optimal candidate incurring the fewest violations of the highest-ranked constraints.

In parallel OT, a derivation consists of a single pass through Gen and Eval. The candidate set produced by Gen includes all possible changes to an input (and is therefore infinite), and the form chosen as optimal at Eval surfaces as the final output. In HS, however, a derivation involves multiple passes through Gen and Eval.

The derivation proceeds as follows. Gen, rather than producing an infinite candidate set, is restricted to producing candidates that differ from the input by one single change. Determining what may be considered a single change is a central question in the HS research program — in this paper, the crucial distinction is between a single *instance* of change and a single *type* of change. In Serial Variation, it is necessary that Gen be restricted to producing a single instance of change; for example, possible single changes include the construction of a single prosodic unit or a change in feature value on a single segment.

The resulting (finite) candidate set is evaluated by the language's constraint hierarchy at Eval and, like in parallel OT, an optimal candidate is chosen. However, instead of exiting as the surface form, the optimal candidate is sent back to Gen — this form serves as the new input. A new candidate set is generated, again with each candidate differing from the (new) input by one single change, and Eval chooses an optimum from this candidate set. This loop continues until the single changes produced by Gen are no longer harmonically improving. That is, the derivation *converges* when the input to the current step is chosen as the optimal candidate.

Proceeding in this fashion has several consequences for the behavior of phonological processes (McCarthy, 2007b). Because Gen is restricted to performing a single change, all processes must be *gradual*. Furthermore, because these single changes are evaluated by the language's constraint hierarchy, they must be *harmonically improving*. As a result of this gradual harmonic improvement, optimality in HS is *local* rather than *global*. Each single change is evaluated independently, and the optimum is the best possible change at that particular step in the derivation.

As (McCarthy, 2007b) points out, these properties of HS give it a number of independent typological advantages over parallel versions of OT. For more discussion of these, see Pruitt (2008) on foot parsing, Pater (2008b) on syllabification, Jesney (2009) on positional faithfulness, Elfner (2009) on epenthesis, McCarthy (2008a) on cluster simplification, McCarthy (2008b) on metrically conditioned syncope, McCarthy (2008a) on harmony, and McCarthy (2007a); Wolf (2008) on opacity (both phonology-phonology and phonology-morphology).

In the context of variation, HS has another desirable typological consequence: local optimality predicts local variation and, under certain circumstances (see Section 4), global variation. Serial evaluation, when combined with independently motivated theories of phonological variation, allows variants to be chosen separately at each locus within the form rather than globally for the form as a whole.

1.2 Variation

A number of models of phonological variation have been proposed within OT which derive variability through multiple available rankings of a language's constraint hierarchy. These include Anttila (1997)'s Partially Ordered Constraints, Boersma (1997)'s Stochastic OT, and Reynolds (1994); Nagy and Reynolds (1997)'s Floating Constraints.[1] Crucially, what these models have in common is the notion that a grammar is something other than a total order of constraints — instead, it is some partial order (or, in the case of Stochastic OT, a set of values along a numerical scale) that becomes a total order of constraints at EVAL. Variation arises from the fact that a given grammar allows the possibility of multiple total orders which produce different optimal candidates.

The present discussion does not crucially rely on choosing among these multiple-rankings models. For simplicity of exposition and representation, I will be assuming a Partially Ordered Constraints model — Stochastic OT and Floating Constraints would be equally compatible with the analysis given in this paper, and the interested reader can freely substitute their preferred theory.[2]

In the Partially Ordered Constraints model, a grammar consists of constraints and their rankings, but the rankings are incomplete. For example, in the grammar in (5), constraints B and C are not ranked with respect to each other:

(5) Constraints: A, B, C
Rankings: A $\gg$ B, A $\gg$ C

From this partial order, two total orders are possible: A $\gg$ B $\gg$ C and A $\gg$ C $\gg$ B. These total orders may choose different optimal candidates, as illustrated by the tableaux in (6) and (7):

(6)

input	A	B	C
a. ☞ $cand_1$	*		*
b. $cand_2$	*	*!	

(7)

input	A	C	B
a. $cand_1$	*	*!	
b. ☞ $cand_2$	*		*

[1]For an overview of theories of phonological variation, see Anttila (2007); Coetzee and Pater (2008).

[2]Models of variation based in Harmonic Grammar, such as Noisy HG (Boersma and Pater, 2008) or MaxEnt (Goldwater and Johnson, 2003) are also compatible with the proposal presented here.

Each time a candidate set is evaluated at EVAL, a total order consistent with the language's partial order is chosen. When possible total orders disagree about which candidate is optimal, variation arises — the choice of total order determines the choice of variant.

It should be noted that different models of variation make different predictions about the relative frequencies of attested variants, but all have some ability to model frequency effects in variation. Because the analysis in this paper makes use of these theories, the capacity to model frequency is inherited. How best to accomplish this is an important and unanswered question; however, little to no data on frequency is available for the particular cases of local variation under discussion. This paper is therefore concerned only with attested vs. unattested variants, and frequency is set aside.

1.3 Parallel vs. Serial Evaluation

In parallel OT, a derivation includes only one pass through EVAL, and all possible changes are evaluated concurrently. This means that there is only one opportunity to select a total ranking of partially ordered constraints, and every locus in the domain is subject to that ranking.

As Vaux (2003) points out, local variation poses a problem for parallel evaluation in OT; because harmony is evaluated in parallel for the entire domain, the optimal variant should be optimal throughout. If two constraints are in variation, the optimal candidate will always be the one that best satisfies the constraint that is higher ranked for any given evaluation.

Consider the following schematic example. We have the constraints *VOICE and IDENT(VOICE), partially ranked with respect to one another in the grammar. If *VOICE outranks IDENT(VOICE) in the total order chosen at EVAL, a candidate with no voicing will be selected as optimal.[3]

(8)

dada	*VOICE	IDENT(VOICE)
a. dada	W_2	L
b. data	W_1	L_1
c. tada	W_1	L_1
d. ☞ tata		$_2$

However, if IDENT(VOICE) outranks *VOICE in the total order chosen at EVAL, all voicing will be preserved:

[3]The format of the tableaux in this paper is a modification of Prince (2000)'s Comparative Tableaux. The subscripted numbers represent the number of violations a candidate receives on a given constraint; for each constraint column, a W or L marks cells with more or fewer violations than the winner (respectively).

(9)

dada	IDENT(VOICE)	*VOICE
a. ☞ dada		*2*
b. data	W*1*	L*1*
c. tada	W*1*	L*1*
d. tata	W*2*	L

No re-ranking of these constraints will result in a victory for candidates b or c, but they will not lose to the same candidate each time — they are *collectively harmonically bounded* (Samek-Lodovici and Prince, 2002). Changing the definitions of the constraints can alter which candidates win, but the problem remains: not all of the attested forms will be allowed to win. Expanding the constraint set can get a few more options to be possible, but the upper bound on possible loci exceeds the upper bound on reasonable additions to the hierarchy.

In HS, however, a derivation involves multiple passes through the GEN→EVAL loop. Implementing a Partially Ordered Constraints model of variation in HS means that each step of the derivation involves a new invocation of EVAL and hence a new selection of a total order from the grammar's partial order. There is no requirement that the same total order be selected at each step, meaning that the choice of variant at each step may differ.

The proposal set forth in this paper, Serial Variation, is simply the combination of Harmonic Serialism and a multiple-rankings theory of variation like Partially Ordered Constraints. A derivation proceeds in stepwise fashion through the GEN→EVAL loop, and at each instance of EVAL a total order is imposed on the language's constraint hierarchy. Because each locus of variation is evaluated in a separate step, and because the total order imposed at EVAL need not be the same at each step, the choice of variant at each locus is independent.

Consider again the above schematic example. Our input is /dada/, and the possible surface forms include [dada] (the faithful mapping), [tata] (the unmarked mapping), and the intermediate variants [data] and [tada] (which combine faithfulness and markedness reduction). We can derive the intermediate variants — the locally variable forms — by varying the total order of constraints at each step of the derivation.

On the first step, a total order with *VOICE ≫ IDENT(VOICE) compels us to devoice one of the consonants. Because we can only perform a single instance of change, GEN cannot produce a candidate that fully satisfies *VOICE — [tata] is not a possible competitor here.

(10) **Step 1**

dada	*VOICE	IDENT(VOICE)
a. dada	W*2*	L
b. ☞ data	*1*	*1*
c. ☞ tada	*1*	*1*

On the first step, we choose one of the candidates that performs better, though still not perfectly, on *Voice.[4] On the second step, our input is the output of the previous step, [tada].[5] This time, a total order of Ident(Voice) ≫ *Voice is chosen at Eval. Rather than continuing down the devoicing path to satisfy markedness, the optimal candidate remains faithful to the input.[6]

(11) **Step 2**: *Convergence*

tada	Ident(Voice)	*Voice
a. dada	W*1*	W*2*
b. ☞ tada		*1*
c. tata	W*1*	L

Because the input and output of this step are identical, the derivation converges and [tada] is selected as optimal. As this schematic example demonstrates, Serial Variation is able to predict locality in variation as a consequence of the theoretical architecture of Harmonic Serialism. The following sections show the proposal in action in Bengali and French, and Section 4 will show how particular types of constraint interaction can — and do — produce globality in variation in Serial Variation.

2 Bengali Minor Phrases

In Bengali, a Minor Phrase — which is marked by a L*H% pitch contour and final-syllable lengthening, and serves as the domain for several assimilation processes — may be either a single word or some XP (Hayes and Lahiri, 1991). For example, an adjective-noun sequence may be parsed as a single MiP or as two distinct MiPs:

(12) *patla šari* 'thin sari'

a. (patla šari)

b. (patla)(šari)

In larger XP structures, this optionality is local rather than global. Within a single phrasing domain, it is possible to have a combination of single-word and larger MiPs:

(13) *k*^{h}*ub ṭɔk gur-er jonno* 'of very bad molasses'

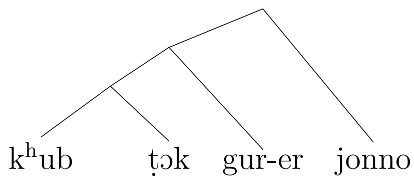
[4]With this limited constraint set, we have tied candidates — I assume that lower-ranked constraints will prefer one over the other (Emergence of the Unmarked). How actual ties are handled in HS is an interesting and outstanding question, but beyond the scope of this paper.

[5]Throughout the paper, I will use square brackets for any non-underlying form, including failed candidates and outputs of intermediate derivational steps. It should be noted, however, that in Harmonic Serialism — unlike in parallel versions of OT — there is a meaningful difference between outputs and surface forms.

[6]Unlike in OT-CC (McCarthy, 2007a), faithfulness is reckoned with respect to the input to the current step, not the ultimate input — going back to *dada* violates faithfulness, rather than satisfying it. This has been assumed in recent literature on HS; see Jesney (2009) for some desirable typological consequences of this, and see Section 2.3 for the consequences for the present discussion.

k^{h}ub ṭɔk gur-er jonno
very bad molasses-GEN of

'of very bad molasses'



a. (k^{h}ub ṭɔk gur-er jonno)
b. (k^{h}ub ṭɔk gur-er)(jonno)
c. (k^{h}ub ṭɔk)(gur-er)(jonno)
d. (k^{h}ub)(ṭɔk)(gur-er)(jonno)

In this section, I will present a serial analysis of Bengali Minor Phrasing. Section 2.1 lays out the basic analysis of optionality in phrasing. Section 2.2 extends the basic analysis to cases of local variation, demonstrating that serial construction of MiPs produces the desired locality effects; Section 2.2.1 and Section 2.2.2 derive the attested variants for a MiP domain with three words, and 2.2.3 shows how this extends to domains of increasing (and potentially unbounded) length. The analysis of locality given here is based on the same basic principle as Truckenbrodt (2002)'s cyclic account of Bengali MiP assignment — see Section 5.3 for a comparison of HS with cyclic versions of OT.

2.1 The Basics of MiP Assignment

In a HS framework, Minor Phrases are built one at a time — the single change performed by GEN is the construction of a single Minor Phrase. Following Pruitt (2008), I will attribute the following properties to the serial construction of prosodic structure:

i. Head and dependents are assigned in the same step. That is, MiPs must be constructed one at a time, but we can construct a *whole* MiP at each step (cf. Elfner 2008).

ii. Prosodic structure is monotonic. This can be accomplished either language-specifically through high-ranked faithfulness to prosodic units, or via a universal property of GEN (e.g. Prince 1985's Free Element Condition) — the result is that a prosodic unit constructed at one step in the derivation cannot be modified on subsequent steps (see Section 2.3 for more discussion of this point).

Precursors to (ii) can also be found in discussions in the literature on cyclic structure preservation; see for example Itô (1986); Kiparsky (1985); Selkirk (1980); Steriade (1988).

Pruitt (2008) points out that, since only one prosodic unit is constructed on each step, constraints preferring exhaustive parsing — for example, PARSE-σ (Prince and Smolensky, 1993/2004) or EXHAUSTIVITY (Selkirk, 1995) — also prefer larger prosodic units. Larger MiPs allow more material to be parsed on a given step, and so will perform better on a constraint ilke EXHAUSTIVITY(MIP):

(14) Exhaustivity(MiP): Assign one violation mark for each Lexical Word that is not parsed as part of some Minor Phrase.[7]

However, single-word MiPs are preferred by a constraint that prohibits words from appearing in dependent position:

(15) *WeakWord: Assign one violation mark for each Lexical Word that is parsed as a dependent in a Minor Phrase.

This constraint follows the tradition of constraints like Weight-to-Stress (Prince, 1991; Prince and Smolensky, 1993/2004), which want prominent units to be parsed into prominent positions. See also de Lacy (2006) for further discussion of constraints of this type — extending the generalizations about weak and prominent positions to sentence prosody is not without complication (do constraints against prominent units in weak positions extend all the way up the prosodic hierarchy?) but it is reasonable to suggest that the same basic principles are at play.

*WeakWord differs from constraints like Weight-to-Stress in one crucial way — it is not violated by unparsed material. It is thus able to distinguish between words that have been committed to a weak position and words that have simply not yet been committed to any position.

In Bengali, Exhaustivity(MiP) and *WeakWord are unranked with respect to each other in the grammar's partial order. At each instantiation of Eval, some ranking of these two constraints is chosen.

If Exhaustivity(MiP) ≫ *WeakWord at the first step in the derivation (the first pass through the Gen→Eval loop), the optimal choice is to build a single MiP that encompasses both words. At the second step in the derivation, it is no longer possible to build any further MiPs,[8] so the input is selected as the optimal output, and the derivation converges.

(16) **Step 1**

patla šari	Exh(MiP)	*WeakWord
a. ☞ (patla šari)		*1*
b. (patla) šari	W*1*	L
c. patla (šari)	W*1*	L
d. patla šari	W*2*	L

Step 2: *Convergence*

[7]Whether this constraint refers to the Prosodic Word or Lexical Word is an open question, and hinges on assumptions about top-down versus bottom-up parsing and the interactions of word-level and sentence-level phonology. The choice between these assumptions is non-trivial, but tangential to the primary goals of this paper. I'll assume a top-down model of parsing, but the discussion here could be brought in line with bottom-up parsing with minor adjustments.

[8]I will assume that either Gen is unable to produce recursive prosodic structure, or that there is some constraint undominated in Bengali which militates against it.

(patla šari)	EXH(MIP)	*WEAKWORD
a. ☞ (patla šari)		*1*

On the other hand, if a ranking of *WEAKWORD ≫ EXHAUSTIVITY(MIP) is selected on the first step, the optimal choice will be to build a single-word MiP. With the constraints we have seen so far, either word will be equally acceptable — see Section 2.2.2 for further discussion. On the second step, regardless of which ranking is chosen, the optimal choice will be to parse the remaining word into its own MiP. The derivation will converge on the third step, since it is no longer possible to build additional MiPs.

(17) **Step 1**

patla šari	*WEAKWORD	EXH(MIP)
a. (patla šari)	W*1*	L
b. ☞ (patla) šari		*1*
c. ☞ patla (šari)		*1*
d. patla šari		W*2*

Step 2

patla (šari)	*WEAKWORD	EXH(MIP)
a. ☞ (patla) (šari)		
b. patla (šari)		W*1*

Step 3: *Convergence*

(patla) (šari)	*WEAKWORD	EXH(MIP)
a. ☞ (patla) (šari)		

With simple cases, where there is only one locus of variation within a form, Harmonic Serialism and Parallel OT make identical predictions about which variants should be possible — there is no opportunity here for variation to be local. However, it is possible that a given form may contain multiple loci where an optional or variable process may apply. In these cases, as we will see in the next section, HS is able to straightforwardly account for the fact that choice of a variant at each locus may differ from the choice made at other loci.

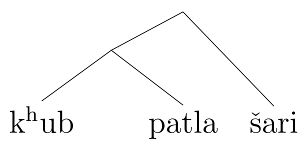
2.2 Local Variation

The domain for assignment of a Minor Phrase is within a Major Phrase — according to Kratzer and Selkirk (2007), a Major Phrase is assigned to the highest phrase within the spell-out domain of a phase. What this means for the present discussion of Bengali prosody is that the domain of Minor Phrase assignment can be a constituent like a PP or DP, and these may contain unboundedly many lexical items (due to e.g. the iteration of adjectives, etc.).

The distinction between Major and Minor phrases is not without controversy. However, see Khan (2007) for empirical evidence supporting an intermediate level of phrasing between the Minor Phrase (Hayes and Lahiri's p-phrase) and the Intonational Phrase — Minor Phrases are marked by a L*H% pitch contour and serve as the domain for several segmental assimilation processes, and Major Phrases are marked with a pitch boost and additional final-syllable lengthening at the right edge.

In cases where the domain of MiP assignment is larger than a simple pair of words, multiple options are available and attested. For example, when the domain contains three words, as in (18), the options available for parsing are as in (19a-c):

(18) k^h*ub patla šari* 'very thin sari'



(19) a. (k^hub patla šari)
b. (k^hub patla)(šari)
c. (k^hub)(patla)(šari)
d. *(k^hub)(patla šari)

The range of possible variants includes global variants: a parse where the entire domain is grouped into a single Minor Phrase (19a) and a parse where each word receives its own Minor Phrase (19c). However, (19b) and (19d) represent mixed variants — we see both a single-word MiP and a multiple-word MiP within the same Major Phrase. The parse in (19b) is a permissible combination of single-word and larger MiPs, but (19d) is unattested — see Section 2.2.2 for a discussion of how that option is ruled out.[9]

An account of local variation in Bengali MiPs must of course be able to produce the globally optional variants in (19a) and (19c). It must also be able to account for the mixed variants, including distinguishing between (19b) and (19d). Finally, it must extend beyond the three-word cases to domains of potentially unbounded length.

This section will deal with each of these issues in turn. Section 2.2.1 will present derivations arriving at the global variants in precisely the same manner as the examples seen in Section 2.1. Section 2.2.2 demonstrates that multiple passes through Eval in HS permit derivations arriving at the mixed variant in (19b), and furthermore shows that directional parsing (via alignment constraints) allows us to distinguish between (19b) and (19d). Section 2.2.3 extends this analysis to larger Minor Phrasing domains.

[9] Hayes and Lahiri (1991, p.84-85) also discuss some degree of prosodic variation in higher-level syntactic structures — the generalizations for possible and impossible parses in constructions like (18) do not hold true of those higher-level syntactic structures. Those cases differ from the ones under consideration here not only in syntactic structure (it's likely that those cases are restricted by constraints on MaP assignment) but also in the conditioning factors for variation — the options in (19) are available regardless of speech rate or discourse context, while the other cases discussed by Hayes and Lahiri are variable only in fast speech or under given-marking.

2.2.1 Global Variants

The derivation resulting in the form in (19a) works in precisely the same manner as (16) — on the first step, Exhaustivity(MiP) ranks above *WeakWord, and an exhaustive MiP is constructed. On the second step, since it is no longer possible to build additional MiPs, the derivation converges.

(20) **Step 1**

k^hub patla šari	Exhaustivity(MiP)	*WeakWord
a. ☞ (k^hub patla šari)		*2*
b. (k^hub patla) šari	W*1*	L*1*
c. k^hub (patla šari)	W*1*	L*1*
d. (k^hub) patla šari	W*2*	L
e. k^hub (patla) šari	W*2*	L
f. k^hub patla (šari)	W*2*	L
g. k^hub patla šari	W*3*	L

Step 2: *Convergence*

(k^hub patla šari)	Exhaustivity(MiP)	*WeakWord
a. ☞ (k^hub patla šari)		*2*

Likewise, the derivation arriving at (19c) is similar to the derivation in (17). On the first step, a ranking of *WeakWord ≫ Exhaustivity(MiP) is chosen, and the optimal choice is to build a single-word MiP.[10] On the second step, a ranking of *WeakWord ≫ Exhaustivity(MiP) is chosen again, and another single-word MiP is built. On the third step, regardless of ranking, the only remaining option for prosodic parsing is to build another single-word MiP, and on the fourth step it is no longer possible to build additional MiPs and the derivation converges.

(21) **Step 1**

[10]See Section 2.2.2 for why the rightmost word is chosen.

k^hub patla šari		*WeakWord	Exhaustivity(MiP)
a.	(k^hub patla šari)	W*2*	L
b.	(k^hub patla) šari	W*1*	L*1*
c.	k^hub (patla šari)	W*1*	L*1*
d. ☞	(k^hub) patla šari		*2*
e. ☞	k^hub (patla) šari		*2*
f. ☞	k^hub patla (šari)		*2*
g.	k^hub patla šari		W*3*

Step 2

k^hub patla (šari)		*WeakWord	Exhaustivity(MiP)
a.	(k^hub patla)(šari)	W*1*	L
b. ☞	(k^hub) patla (šari)		*1*
c. ☞	k^hub (patla)(šari)		*1*
d.	k^hub patla (šari)		W*2*

Step 3

k^hub (patla)(šari)		*WeakWord	Exhaustivity(MiP)
a. ☞	(k^hub)(patla)(šari)		
b.	k^hub (patla)(šari)		W*1*

Step 4: *Convergence*

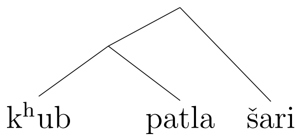
(k^hub)(patla)(šari)		*WeakWord	Exhaustivity(MiP)
a. ☞	(k^hub)(patla)(šari)		

The variants shown here represent *global* options, where one preference is expressed consistently throughout the entire form. It is important to note that local variation does not preclude the possibility of these forms — it is still possible to arrive at these global variants serially. Next, we turn to the variants which manifest mixed preferences.

2.2.2 Mixed Variants

In (19) above, only one of the two possible mixed variants was attested:

(22) *k^hub patla šari* 'very thin sari'



a. (k^hub patla šari)

b. (k^hub patla)(šari)

c. (k^hub)(patla)(šari)

d. *(k^hub)(patla šari)

In this section, I'll show that serial evaluation allows derivations that produce mixed variants, and directional parsing (via alignment constraints) restricts the possible derivations to only those which produce the attested forms.

On the first step in the derivation in (21), a ranking of *WeakWord ≫ Exhaustivity(MiP) forces the parsing of a single-word MiP. This raises the question: which word is parsed on this step? Given the constraints established so far, parsing *any* of the three words in the input into a MiP is an equally harmonic choice — the candidates are tied with respect to the hierarchy as it stands.

Ties between candidates are unreliable; lower-ranked constraints will be left to decide among the candidates (McCarthy and Prince, 1994). The tie will be broken, and in order for the attested range of variants to surface, *k^hub patla (šari)* must be optimal at the first step. If *k^hub (patla) šari* is chosen at the first step, the result will always be a domain parsed into single-word MiPs. If *(k^hub) patla šari* is chosen at the first step, the result will either be a domain parsed into single-word MiPs or the unattested mixed parse in (22).

There are two possible ways of ensuring that *k^hub patla (šari)* will be optimal at the first step. We could constrain the order in which constituents are parsed with constraints referring to the depth of syntactic embedding; less deeply embedded constituents are parsed before more deeply embedded ones. Alternatively, we could achieve a similar effect via directionality of parsing.

Determining which of these possible approaches is better requires more information about the typology of sentence prosody than is currently available.[11] The latter approach is chosen here because it makes use of already-familiar alignment constraints:

(23) AlignR: Assign one violation mark for each Minor Phrase that is not aligned with the right edge of a Major Phrase.

(24) AlignL: Assign one violation mark for each Minor Phrase that is not aligned with the left edge of a Major Phrase.

To ensure complete parsing, Exhaustivity(MiP) must dominate both alignment constraints. In a serial framework, ranking of alignment constraints will determine the direction of parsing; if AlignR ≫ AlignL, the first single-word MiP will be at the right edge. In

[11]Specifically, an alignment-based account predicts a possible language Bengali′ in which parsing is directional at the expense of constituency — if this prediction proves false, something like like Kratzer and Selkirk (2007)'s Highest Phrase Condition could be employed here in place of alignment.

(21), directionality didn't matter — repeated rankings of *WeakWord ≫ Exhaustivity(MiP) would produce an output with all single-word MiPs regardless of the direction of parsing.

However, in the mixed variants, selecting a different ranking of *WeakWord and Exhaustivity(MiP) at each instantiation of Eval results in forms with a combination of single-word MiPs and larger MiPs. In these derivations, directionality becomes important. A ranking of AlignR ≫ AlignL allows variation between *WeakWord and Exhaustivity(MiP) to successfully produce the form in (22b) and rule out the form in (22d).

On the first step, *WeakWord ≫ Exhaustivity(MiP) and a single-word MiP is built; the ranking of our alignment constraints forces that MiP to be aligned with the right edge. On the second step, Exhaustivity(MiP) ≫ *WeakWord and the optimal MiP encompasses all remaining unparsed material. On the third step it's no longer possible to build additional MiPs, and the derivation converges.

(25) **Step 1**

k^hub patla šari	*WeakWord	Exh(MiP)	AlignR	AlignL
a. (k^hub patla šari)	W_2	L		L
b. (k^hub patla) šari	W_1	L_1	W_1	L
c. k^hub (patla šari)	W_1	L_1		L_1
d. (k^hub) patla šari		2	W_2	L
e. k^hub (patla) šari		2	W_1	L_1
f. ☞ k^hub patla (šari)		2		2
g. k^hub patla šari		W_3		L

Step 2

k^hub patla (šari)	Exh(MiP)	*WeakWord	AlignR	AlignL
a. ☞ (k^hub patla)(šari)		1	1	2
b. k^hub (patla)(šari)	W_1	L	1	W_3
c. k^hub patla (šari)	W_2	L	L	2

Step 3 *Convergence*

(k^hub patla)(šari)	Exh(MiP)	*WeakWord	AlignR	AlignL
a. ☞ (k^hub patla)(šari)		1	1	2

Arriving at the unattested parse in (22d) with this ranking of alignment constraints is impossible. Since ranking AlignR over AlignL compels parsing of MiPs to begin at the right edge, we would need a derivation like the one in (26):

(26) a. **Step 1** k^hub (patla šari)

b. **Step 2** *(k^hub)(patla šari)

The problem here is that the parse in (26a) satisfies neither EXHAUSTIVITY(MIP) nor *WEAKWORD. Because it is harmonically bounded, it will never be chosen as the initial MiP, and (22d) is correctly predicted to be an impossible variant.

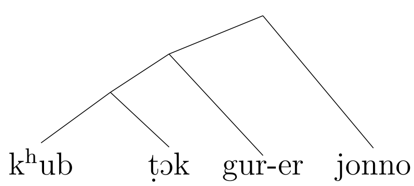
Choosing different rankings at each instantiation of EVAL is what allows derivations resulting in forms which combine preferences regarding the size of MiPs. By controlling directionality of parsing, a ranking of ALIGNR ≫ ALIGNL correctly predicts which mixed variants are permitted.

2.2.3 XPs of Unbounded Length

The analysis established thus far continues to derive all the attested variants, even as the size of the domain for Minor Phrasing grows longer, with no need for additional constraints. The choice for each step is to parse, from right to left, either a single-word MiP or the remainder of the domain; as the length of the domain increases, this continues to be sufficient to account for all the locally variable possibilities.

For example, the relevant domain in (27) is four words long, and all the forms in (28) are attested parsings:

(27) k^hub ṭɔk gur-er jonno
very bad molasses-GEN of
'of very bad molasses'



(28) a. (k^hub ṭɔk gur-er jonno)
b. (k^hub ṭɔk gur-er)(jonno)
c. (k^hub ṭɔk)(gur-er)(jonno)
d. (k^hub)(ṭɔk)(gur-er)(jonno)

Consistent rankings at each step will produce the global variants, in exactly the same manner as the two- and three-word domains. A ranking of EXHAUSTIVITY(MIP) ≫ *WEAKWORD on the first step will result in a single MiP that encompasses the entire domain, as in (28a). A ranking of *WEAKWORD ≫ EXHAUSTIVITY(MIP) at each step will result in each MiP consisting of a single word, as in (28d).

To derive (28b), the derivation begins with a total order of *WEAKWORD ≫ EXHAUSTIVITY(MIP) on the first step. This will force the first MiP created to be a single word, aligned to the right edge. On the second step, a total order where EXHAUSTIVITY(MIP) ≫ *WEAKWORD will force all the remaining material to be parsed into a single MiP. On the third step, the derivation converges — it is no longer harmonically improving to build MiPs

(29) **Step 1**

k^{h}ub tɔk gur-er jonno		*WkWd	Exh(MiP)	AlignR	AlignL
a.	(k^{h}ub tɔk gur-er jonno)	W_3	L		L
b.	(k^{h}ub tɔk gur-er) jonno	W_2	L_1	W_1	L
c.	(k^{h}ub tɔk) gur-er jonno	W_1	L_2	W_2	L
d.	(k^{h}ub) tɔk gur-er jonno		3	W_3	L
e.	k^{h}ub (tɔk) gur-er jonno		3	W_2	L_1
f.	k^{h}ub tɔk (gur-er) jonno		3	W_1	L_2
g. ☞	k^{h}ub tɔk gur-er (jonno)		3		3
h.	k^{h}ub tɔk gur-er jonno		W_4		L

Step 2

k^{h}ub tɔk gur-er (jonno)		Exh(MiP)	*WkWd	AlignR	AlignL
a. ☞	(k^{h}ub tɔk gur-er) (jonno)		2	1	3
b.	(k^{h}ub tɔk) gur-er (jonno)	W_1	L_1	W_2	3
c.	(k^{h}ub) tɔk gur-er (jonno)	W_2	L	W_3	3
d.	k^{h}ub (tɔk) gur-er (jonno)	W_2	L	W_2	W_4
e.	k^{h}ub tɔk (gur-er)(jonno)	W_2	L	1	W_5
f.	k^{h}ub tɔk gur-er (jonno)	W_3	L	L	3

Step 3: *Convergence*

(k^{h}ub tɔk gur-er) (jonno)		Exh(MiP)	*WkWd	AlignR	AlignL
a. ☞	(k^{h}ub tɔk gur-er) (jonno)		2	1	3

To derive (28c), the derivation begins with a total order of *WkWd ≫ Exhaustivity(MiP) on both the first and second steps. This will result in parsing of right-aligned single-word MiPs for two out of the four words in the form. On the third step, a total order of Exhaustivity(MiP) ≫ *WkWd will parse the remaining two words together into a single MiP. At the fourth step, it is no longer harmonically improving to build additional MiPs, and the derivation converges.

(30) **Step 1**

k^hub tɔk gur-er jonno	*WkWd	Exh(MiP)	AlignR	AlignL
a. (k^hub tɔk gur-er jonno)	W_3	L		L
b. (k^hub tɔk gur-er) jonno	W_2	L_1	W_1	L
c. (k^hub tɔk) gur-er jonno	W_1	L_2	W_2	L
d. (k^hub) tɔk gur-er jonno		*3*	W_3	L
e. k^hub (tɔk) gur-er jonno		*3*	W_2	L_1
f. k^hub tɔk (gur-er) jonno		*3*	W_1	L_2
g. ☞ k^hub tɔk gur-er (jonno)		*3*		*3*
h. k^hub tɔk gur-er jonno		W_4		L

Step 2

k^hub tɔk gur-er (jonno)	*WkWd	Exh(MiP)	AlignR	AlignL
a. (k^hub tɔk gur-er)(jonno)	W_2	L	*1*	L_3
b. (k^hub tɔk) gur-er (jonno)	W_1	L_1	W_2	L_3
c. (k^hub) tɔk gur-er (jonno)		*2*	W_3	L_3
d. k^hub (tɔk) gur-er (jonno)		*2*	W_2	L_4
e. ☞ k^hub tɔk (gur-er)(jonno)		*2*	*1*	*5*
f. k^hub tɔk gur-er (jonno)		W_3	L	L_3

Step 3

k^hub tɔk (gur-er)(jonno)	Exh(MiP)	*WkWd	AlignR	AlignL
a. ☞ (k^hub tɔk)(gur-er)(jonno)		*1*	*3*	*5*
b. (k^hub) tɔk (gur-er)(jonno)	W_1	L	W_4	*5*
c. k^hub (tɔk)(gur-er)(jonno)	W_1	L	*3*	W_6
d. k^hub tɔk (gur-er)(jonno)	W_2	L	L_1	*5*

Step 4: *Convergence*

(k^hub tɔk) (gur-er)(jonno)	Exh(MiP)	*WkWd	AlignR	AlignL
a. ☞ (k^hub tɔk)(gur-er)(jonno)		*1*	*3*	*5*

The account in this paper successfully extends to MiP domains four words long, and as the XP grows larger the ability of a serial derivation to produce the attested phrasing options will remain constant.

Accommodating these facts in parallel OT would require expanding the constraint set. Adding a constraint demanding binary MiPs (Selkirk and Tateishi, 1988) could produce [(k^hub ṭɔk)(gur-er)(jonno)] as a possible winner. However, it's unclear what independently justified constraint will prefer [(k^hub ṭɔk gur-er)(jonno)] over its competitors. With five and six word examples, it becomes fairly clear that no existing constraints will be able to produce all the attested variants.

The number of plausible constraints preferring attested variants is limited, while the number of words within a MiP parsing domain (and hence the number of variants for MiP assignment) is unbounded. This is largely independent of the particular set of constraints being used — there will always be an upper bound to the number of independently justified constraints preferring local variants.

2.3 Monotonicity and Convergence

In order for the derivations discussed above to produce the attested variants, it is necessary to treat the construction of prosodic structure as monotonic.

Without monotonicity, we can undo or change MiPs that we constructed on a previous step. In Bengali, this means that the derivation will be unable to converge on some of the locally variable forms. For example, if the candidate set for a fully-parsed input [(k^hub patla)(šari)] includes a candidate which has parsed a previously-parsed word into its own MiP, this candidate will win:

(31) **Step 3**

(k^hub patla)(šari)	Exh(MiP)	*WeakWord
a. (k^hub patla)(šari)		*1*
b. ☞ (k^hub)(patla)(šari)		L

Both candidates satisfy Exhaustivity(MiP), but the candidate with all single-word MiPs also satisfies *WeakWord. The parse [(k^hub patla)(šari)] will never map faithfully, and so will never be converged upon.

However, if prosodic structure is protected by faithfulness constraints, ranking those faithfulness constraints sufficiently high commits us to the structure we've already built. A change in preference regarding that structure at a later step will have no effect on what we've done so far.

The parse [(k^hub)(patla)(šari)] is unfaithful to the prosodic structure in the input [(k^hub patla)(šari)], and incurs violations of high-ranked faithfulness. The derivation will converge on the faithful mapping:

(32) **Step 3**

(k^hub patla)(šari)	Faith(MiP)	Exh(MiP)	*WeakWord
a. ☞ (k^hub patla)(šari)			*1*
b. (k^hub)(patla)(šari)	W*1*		L

Faithfulness is able to force monotonicity here because we are faithful to the input to the current step, rather than the ultimate input. The derivation is committed to preserving the structure that has just been built, rather than being committed to preserving whatever structure may or may not be present underlyingly.

However, high-ranked faithfulness is only one way to get monotonicity. Pruitt (2008) proposes a universal requirement of strict inheritance for iterative foot parsing:

(33) **Strict inheritance**: "Any foot that is built in the course of deriving stress is inherited by any member of the candidate set for subsequent iterations." (Pruitt, 2008, 6)

This idea is not new — in much of the rule-based literature on prosodic structure, Prince (1985)'s Free Element Condition (FEC) ("rules of primary metrical analysis apply only to free elements — those that do not stand in the metrical relationship being established") was either implicitly or explicitly assumed.

The question now is whether a restriction like strict inheritance or the FEC is universal (a restriction on Gen) or violable (a faithfulness constraint or constraints). Pruitt argues for strict inheritance as a universal requirement on the basis of the typology of stress systems — certain non-local interactions pathologically predicted in Parallel OT are eliminated in HS if foot building is iterative and necessarily monotonic.

If strict inheritance is violable, it should be possible to find a language like Bengali′, with Exh(MiP) and *WeakWord ranked above faithfulness. In this language, prosodic structure building would be subject to the same convergence requirements as other markedness/markedness interactions (see Section 4) and variation would be predicted to be global. In the absence of any evidence bearing on this question, I will follow Pruitt in treating strict inheritance as a universal property of Gen, but either approach is equally compatible with the analyses presented in this paper.

Regardless of its source, monotonicity gives us locality in variation because a committment to what we have already done means that a subsequent change in ranking will necessarily result in a form that mixes preferences. In other words, we can do it differently next time, but we can't change the way we've just done it.

2.4 Summary

In summary, a serial analysis of MiP assignment in Bengali predicts locality of variation. The monotonicity of prosodic structure building operations ensures that we will be able to converge on locally variable forms, and alignment constraints force directional parsing, which eliminates the unattested locally variable forms.

This analysis extends beyond the very simple cases, and remains unchanged even with potentially unbounded increases in the number of words in the domain of parsing.

3 French Schwa Deletion

Another example of a locally optional process is schwa deletion in French. This is a much-studied phenomenon (Anderson, 1982; Andreasson, 2004; Angoujard, 2006; Basbøll, 1981; Côté and Morrison, 2007; Dell, 1973; Selkirk, 1978; Walker, 1993; Eychenne, 2006), and an exhaustive account of the process is beyond the scope of this paper. In order to focus sharply on the problem of local variation, I analyze a subset of the French data from Dell (1973). In Section 3.3, I introduce some of the complications and suggest how they can be accommodated in the analysis presented here. In this section, I will provide an overview of schwa deletion, and I will show how the HS analysis of variation developed in this paper extends to this case.

Optional deletion of schwa in French can be seen in the following example:[12]

(34) vəne isi ∼ v_ne isi *(venez ici)* 'come here'

In forms with more than one schwa, optionality is local. In (35) and (36), every schwa may surface faithfully (35a, 36a), any single schwa may delete (35b-e, 36b-d), or as many schwas as possible may delete (35f-g, 36e):

(35) *envie de te le demander* 'feel like asking you for it'

a. ãvi də tə lə dəmãde
b. ãvi d_ tə lə dəmãde
c. ãvi də t_ lə dəmãde
d. ãvi də tə l_ dəmãde
e. ãvi də tə lə d_mãde
f. ãvi d_ tə l_ dəmãde
g. ãvi də t_ lə d_mãde

(36) *la queue de ce renard* 'this fox's tail'

a. la kø də sə rənar
b. la kø d_ sə rənar
c. la kø də s_ rənar
d. la kø də sə r_ nar
e. la kø d_ sə r_ nar

The generalization is summarized in (37):

(37) Schwa may optionally delete unless an adjacent schwa has also deleted.

If optionality were global rather than local, we would expect only (35a) and either (35f) or (35g).[13] The decision to delete or not is made for *each* individual schwa, rather than for the domain as a whole.

In the remainder of this section, I will show how Serial Variation permits an account of schwa deletion that captures the locality of this decision. In the analysis presented here, syllables headed by schwa are optionally parsed into dependent positions of feet, and all

[12] The symbol _ is used here to mark the location of a deleted schwa.

[13] The two cases of maximal deletion differ with respect to the properties of the sequences of consonants created as a result — presumably, some constraint (on e.g. sonority sequencing) will prefer one over the other.

dependent schwas subsequently delete. The section is organized as follows. Section 3.1 lays out an analysis of French foot prosody; Section 3.1.1 provides an account of the role schwa plays in foot parsing, and Section 3.1.2 presents an analysis of local variation in cases with multiple schwas. Section 3.2 shows how obligatory deletion of dependent schwas results in the deletion patterns shown in (36) above. Finally, Section 3.3 discusses some of the additional complications in the French data and suggests possible ways to account for those complications within the present framework.

3.1 The French Foot

Selkirk (1978) presents a prosodic account of schwa deletion in French. Her analysis works as follows. Feet in French are by default monosyllabic. Syllables containing schwa may be optionally adjoined as a dependent to a foot to their left. All schwas in dependent position in a foot delete, provided they are not preceded by a sequence of two or more consonants.

Selkirk's account captures the generalization that French has no secondary stress — because each syllable is, on the surface, the head of its own foot, all syllables that do not bear main stress (are not the head foot) are equally (non-)prominent. As Angoujard (2006) points out, Selkirk's account also captures the generalization that two schwas in a row may not both delete, because two schwas in a row will not both be parsed as dependents in feet, so long as we restrict French to a single foot type.

The analysis of foot parsing developed in this paper closely follows Pruitt (2008)'s theory of iterative footing. Gen is restricted to creating maximally disyllabic feet, head and dependent are parsed in a single step, and prosodic structure is strictly inherited.

The discussion here begins with a brief analysis of how monosyllabic feet are constructed in words with no schwas. To motivate foot parsing, we need a constraint that punishes unparsed syllables:

(38) Parse-σ: Assign one violation mark for every syllable that is not parsed as part of some foot.

As we saw above, constraints of this type have a somewhat different effect in a serial analysis than in a parallel one; a larger foot will parse more material, and therefore will better satisfy Parse-σ.

To force monosyllabic feet in non-schwa contexts, Parse-σ must be dominated by a constraint that punishes full vowels from appearing in weak positions:

(39) *WeakV(>ə): Assign one violation mark for a vowel more sonorous than schwa[14] in the weak position of a foot.

*WeakV(>ə) is equivalent to $*{-}\Delta_{Ft} > \{ə\}$ in the schema from de Lacy (2006, 227). Like *WeakWord in the preceding section, this constraint is based on the observation that prominent things — in this case, more sonorant vowels — are dispreferred in non-prominent positions.

[14]It should be noted that schwa in French is not, in its phonetic realization, the same as schwa in other languages. Defining what separates French schwa from other vowels is controversial; should sonority be a poor criterion, an alternative constraint could refer to e.g. place features.

The interaction of PARSE-σ and *WEAKV(>ə) produces monosyllabic feet in words with no schwa. On the first step, the optimal candidate is a monosyllabic foot: disyllabic feet are eliminated by *WEAKV(>ə), and doing nothing is ruled out by PARSE-σ. On the second step, the optimal candidate parses another monosyllabic foot. Finally, on the third step the remaining syllable is parsed into a monosyllabic foot. On the fourth step, no more feet can be parsed.

(40) **Step 1**

si.ka.tris	*WEAKV(>ə)	PARSE-σ
a. si.ka.tris		W*3*
b. ☞ si.ka(ˈtris)		*2*
c. ☞ si(ˈka)tris		*2*
d. ☞ (ˈsi)ka.tris		*2*
e. si(ka.ˈtris)	W*1*	L*1*
f. (si.ˈka)tris	W*1*	L*1*

Step 2

si.ka(ˈtris)	*WEAKV(>ə)	PARSE-σ
a. si.ka(ˈtris)		W*2*
b. ☞ si(ˌka)(ˈtris)		*1*
c. ☞ (ˌsi)ka(ˈtris)		*1*
d. (si.ˌka)(ˈtris)	W*1*	L

Step 3

si(ka)(ˈtris)	*WEAKV(>ə)	PARSE-σ
a. si(ˌka)(ˈtris)		W*1*
b. ☞ (ˌsi)(ˌka)(ˈtris)		

Step 4: *Convergence*

The winning candidate here violates a constraint like FTBIN, which requires binary feet; we can assume that this constraint is crucially dominated. Additionally, multiple candidates parsing monosyllabic feet are equally harmonic — unlike in Bengali, directionality of parsing is not crucial here. I'll assume right to left parsing, under the assumption that the head foot is parsed first (French main stress falls on the rightmost syllable of the prosodic word) but this is not a necessary assumption for the present analysis.

The following section will extend this analysis to words with schwa.

3.1.1 Schwa and Foot Parsing

In (41) below, the optionality of schwa deletion is a result of the fact that schwa may optionally be parsed into the weak position of a disyllabic foot, and schwas in weak position then delete:

(41) vəne isi ∼ v_ne isi *(venez ici)* 'come here'

Selkirk (1978)'s analysis results in trochaic feet — this assumption is not explicitly justified, but seems to be based on the generalization that word-final schwas always delete. However, parsing trochaic feet means that word-initial schwas will never be subject to deletion; word-initial syllables will always be parsed into the strong position of a foot. Selkirk derives deletion of inital schwas by assuming that they are in the weak position of a foot whose head is in a preceding word — however, this relies on the assumption that the domain of foot parsing is larger than a prosodic word, and fails to account for deletion of word-initial schwas following a pause (for example, in utterance-initial contexts).

To ensure word-initial weak schwas with reasonable assumptions about prosodic domains, it is necessary to adopt iambic rather than trochaic feet.[15] I will assume that, in French, Iamb ≫ Trochee. Candidates with trochaic feet will not be considered in the tableaux that follow.

*WeakV(>ə) applies only to full vowels — schwa is freely parsed into weak position. In order to get schwa footed on its own (and hence not deleted) we need a constraint prohibiting all vowels from appearing in weak position.

(42) *WeakV: assign one violation mark for any vowel parsed into the weak position of a foot.

This constraint (which is equivalent to de Lacy's $*\text{–}\Delta_{Ft} \geq \{ə\}$) is unranked with respect to Parse-σ in the partial order of the grammar. If a total order with Parse-σ ≫ *WeakV is chosen at Eval, syllables headed by schwa will be parsed in dependent position in a foot.

On the first step, Parse-σ prefers a disyllabic foot, and *WeakV is not ranked highly enough to interfere. On the second step, since we have exhaustively parsed the domain into feet, the derivation converges.

(43) **Step 1**

vəne	Parse-σ	*WeakV
a. və.ne	W$_2$	L
b. və(ˈne)	W$_1$	L
c. (ˈvə)ne	W$_1$	L
d. ☞ (və.ˈne)		1

Step 2: *Convergence*

[15]If word-final schwas are in fact underlying rather than epenthetic, their deletion can be handled in the following way: alignment is violated to avoid a primary-stressed schwa, and then in the following step the schwa deletes to restore alignment. This can be accomplished with a ranking of *ə́ ≫ AllFtR ≫ Max-ə.

The notion of convergence in HS is crucially dependent on the constraints under consideration — a derivation converges when the constraint set selects the input as the output. The derivation in (43) converges with respect to the limited constraint set under examination, only PARSE-σ and *WEAKV. See Section 3.2 for derivations that reflect a more complete constraint set for French schwa, which will not in fact converge on [(və.ˈne)].

If instead a total order of *WEAKV ≫ PARSE-σ is chosen, syllables headed by schwa will be parsed as their own monosyllabic feet. On the first step, *WEAKV prefers to leave the schwa unparsed rather than parsing it as part of a disyllabic foot. On the second step, parsing the schwa-headed syllable as its own foot satisfies both *WEAKV and PARSE-σ. On the third step, the derivation converges.

(44) **Step 1**

və.ne	PARSE-σ	*WEAKV
a. və.ne	W_2	
b. ☞ və(ˈne)	*1*	
c. (ˈvə)ne	*1*	
d. (və.ˈne)	L	W_1

Step 2

və(ˈne)	PARSE-σ	*WEAKV
a. və(ˈne)	W_1	
b. ☞ (ˌvə)(ˈne)		

Step 3: *Convergence*

In a word with one schwa, these two possible rankings of PARSE-σ and *WEAKV result in the possible foot parsings in (45):

(45) a. (ˌvə)(ˈne)

b. (və.ˈne)

The possible feet parsed in French — a monosyllabic foot or an iamb with a schwa as a dependent — should be unsurprising. Given standard assumptions about the the typology of foot parsing (see e.g. Hayes 1995) and constraints restricting weak elements to weak positions (de Lacy, 2006; Prince, 1991, see also Kenstowicz 1996 on quality-sensitive parsing), a language like French is expected to be possible.

In the following section, we will see how variation between these two footings is local in words with multiple schwas.

3.1.2 Local Variation

When there is more than one schwa in a given form, deletion is locally optional. For example, any of the deletion patterns in (46a-e) are possible. As we will see in Section 3.2, this results from the parses in (47a-e).[16]

(46) *la queue de ce renard* 'this fox's tail'

a. la kø də sə rənar

b. la kø d_ sə rənar

c. la kø də s_ rənar

d. la kø də sə r_ nar

e. la kø d_ sə r_ nar

(47) a. (ˌdə)(ˌsə)(ˌrə)(ˈnar)

b. (d_ ˌsə)(ˌrə)(ˈnar)

c. (ˌdə)(s_ ˌrə)(ˈnar)

d. (ˌdə)(ˌsə)(r_ ˈnar)

e. (d_ ˌsə)(r_ ˈnar)

We see that all schwas may be parsed in head position of their own feet (47a), any single schwa may be parsed as a dependent (47b-d), or multiple schwas may be parsed in dependent position (47e). In this section, I will show derivations resulting in each of the above combinations of monosyllabic and disyllabic feet.[17]

Since footing is iterative, and since each pass through EVAL gives us the opportunity for a new total ranking of partially ordered constraints, each schwa may differ with respect to whether or not it is parsed into the weak position of a foot.

For a derivation to result in all monosyllabic feet (47a), *WEAKV must outrank PARSE-σ at each step. On each step, a monosyllabic foot is chosen over a foot that would leave schwa in dependent position. At the fourth step, this is the only possible parse, regardless of the ranking of PARSE-σ and *WEAKV, and on the fifth step there is no material available to parse and the derivation converges.

(48) Monosyllabic feet: (47a), [(də)(sə)(rə)(nar)]

Step 1

də.sə.rə.nar	*WEAKV	PARSE-σ
a. də.sə.rə.nar		W_4
b. ☞ də.sə.rə(ˈnar)		$_3$
c. də.sə(rə.ˈnar)	W_1	L_2

[16]The examples with multiple schwas often involve clitics. The constituent containing a lexical word and clitics to its left has a single right-aligned stress, so I will assume that this constituent is a Prosodic Word, and that it serves as a domain for foot assignment (Selkirk, 1995).

[17]In (47) and elsewhere, I abstract away from resyllabification of stray consonants created by schwa deletion. For a brief discussion of this, see Section 3.3.

Step 2

də.sə.rə(ˈnar)	*WEAKV	PARSE-σ
a. də.sə.rə.(ˈnar)		W*3*
b. ☞ də.sə(ˌrə)(ˈnar)		*2*
c. də(sə.ˌrə)(ˈnar)	W*1*	L*1*

Step 3

də.sə(ˌrə)(ˈnar)	*WEAKV	PARSE-σ
a. də.sə(ˌrə)(ˈnar)		W*2*
b. ☞ də(ˌsə)(ˌrə)(ˈnar)		*1*
c. (də.ˌsə)(ˌrə)(ˈnar)	W*1*	L

Step 4

də(ˌsə)(ˌrə)(ˈnar)	*WEAKV	PARSE-σ
a. də(ˌsə)(ˌrə)(ˈnar)		W*1*
b. ☞ (ˌdə)(ˌsə)(ˌrə)(ˈnar)		

Step 5: *Convergence*

A derivation resulting in multiple disyllabic feet, on the other hand, will have PARSE-σ ranked above *WEAKV at each step. On the first step, a disyllabic foot will be parsed with a schwa in weak position, since it is more important to parse as much material as possible than it is to keep schwa from being a dependent. On the second step, we have the ranking PARSE-σ ≫ *WEAKV again, and we parse another disyllabic foot. On the third step, no more material remains to be parsed, so we converge.

(49) Disyllabic feet: (47e), [(də sə)(rənar)]

Step 1

də.sə.rə.nar	PARSE-σ	*WEAKV
a. də.sə.rə.nar	W*4*	L
b. də.sə.rə(ˈnar)	W*3*	L
c. ☞ də.sə(rə.ˈnar)	*2*	*1*

Step 2

də.sə(rə.ˈnar)	PARSE-σ	*WEAKV
a. də.sə(rə.ˈnar)	W*2*	L*1*
b. də(ˌsə)(rə.ˈnar)	W*1*	L*1*
c. ☞ (də.ˌsə)(rə.ˈnar)		*2*

Step 3: *Convergence*

The intermediate candidates — (47b-d), which show a combination of monosyllabic and disyllabic feet — result from derivations where the ranking of the variable constraints varies at each step.

To get (47b), the derivation proceeds as follows. On the first step, a ranking of *WeakV ≫ Parse-σ is chosen at Eval, and the optimal choice is to parse a monosyllabic foot. On the second step, *WeakV once again outranks Parse-σ, and another monosyllabic foot is built. On the third step, Parse-σ outranks *WeakV, and this time it is optimal to parse a disyllabic foot with schwa in weak position. On the fifth step, the derivation converges.

(50) Intermediate footing: (47b) [(də sə)(rə)(nar)]

Step 1

də.sə.rə.nar	*WeakV	Parse-σ
a. də.sə.rə.nar		W*4*
b. ☞ də.sə.rə(ˈnar)		*3*
c. də.sə(rə.ˈnar)	W*1*	L*2*

Step 2

də.sə.rə(ˈnar)	*WeakV	Parse-σ
a. də.sə.rə.(ˈnar)		W*3*
b. ☞ də.sə(ˌrə)(ˈnar)		*2*
c. də(sə.ˌrə)(ˈnar)	W*1*	L*1*

Step 3

də.sə(ˌrə)(ˈnar)	Parse-σ	*WeakV
a. də.sə(ˌrə)(ˈnar)	W*2*	L
b. də(ˌsə)(ˌrə)(ˈnar)	*1*	L
c. ☞ (də.ˌsə)(ˌrə)(ˈnar)		*1*

Step 4: *Convergence*

A derivation resulting in (46c) also begins with *WeakV ≫ Parse-σ on the first step, parsing a monosyllabic foot. On the second step, however, Parse-σ ranks above *WeakV, and a disyllabic foot is parsed with schwa in dependent position. At the third step, the only remaining possibility is to parse a monosyllabic foot. At the fourth step, the derivation converges.

(51) Intermediate footing: (47c) [(də)(sərə)(nar)]

Step 1

də.sə.rə.nar	*WeakV	Parse-σ
a. də.sə.rə.nar		W *4*
b. ☞ də.sə.rə(ˈnar)		*3*
c. də.sə(rə.ˈnar)	W *1*	L *2*

Step 2

də.sə.rə(ˈnar)	Parse-σ	*WeakV
a. də.sə.rə(ˈnar)	W *3*	L
b. də.sə(ˌrə)(ˈnar)	W *2*	L
c. ☞ də(sə.ˌrə)(ˈnar)	*1*	*1*

Step 3

də(sə.ˌrə)(ˈnar)	Parse-σ	*WeakV
a. də(sə.ˌrə)(ˈnar)	W *1*	L
b. ☞ (də)(sə.ˌrə)(ˈnar)		*1*

Step 4: *Convergence*

For a derivation resulting in (46d), the first step is a ranking of Parse-σ ≫ *WeakV, and a disyllabic foot with schwa in weak position is parsed. On the second step, we have a ranking of *WeakV ≫ Parse-σ, and a monosyllabic foot is parsed. On the third step, regardless of the ranking, the only thing left to do is to parse the remaining syllable into its own foot. On the fourth step the derivation converges.

(52) Intermediate footing: [(47d) (də)(sə)(rənar)]

Step 1

də.sə.rə.nar	Parse-σ	*WeakV
a. də.sə.rə.nar	W *4*	L
b. də.sə.rə(ˈnar)	W *3*	L
c. ☞ də.sə(rə.ˈnar)	*2*	*1*

Step 2

də.sə(rə.ˈnar)	*WeakV	Parse-σ
a. də.sə(rə.ˈnar)		W *2*
b. ☞ də(ˌsə)(rə.ˈnar)		*1*
c. (də.ˌsə)(rə.ˈnar)	W *1*	L

Step 3

də(ˌsə)(rə.ˈnar)	*WeakV	Parse-σ
a. də(ˌsə)(rə.ˈnar)		W$_1$
b. ☞ (ˌdə)(ˌsə)(rə.ˈnar)		

Step 4: *Convergence*

Because we can select a different total order of constraints at each instantiation of Eval, there is a possible derivational path leading to each of the foot parses giving us the deletion patterns in (46a-e).

As with Bengali Minor Phrasing, there is no principled limit on the number of schwas that can appear in a given word. The analysis established previously extends without modification to forms with n+1 schwas. For each schwa, we have the option of parsing it as a head or as a dependent; this remains true regardless of how many schwas there are.

3.2 Weak Schwa Deletion

In this section, I'll show how the variation in parsing discussed above leads to variation in schwa deletion. Schwa deletes whenever it is parsed into a weak position in a foot. To get this effect, we must rank *WeakV over Max:

(53) Max: Assign one violation mark for every segment in the input that is not present in the output.

In a parallel framework, this ranking would result in each schwa deleting rather than being parsed into a weak position. However, in HS, a derivation cannot look ahead; there is no way of knowing that parsing a weak schwa will later motivate deletion. Deletion is *intrinsically ordered* after parsing (McCarthy, 2008b).

The candidate set includes candidates with deletion and candidates with foot parsing — each differs from the input by a single change — but we cannot consider a candidate that has both deleted *and* parsed. Because we must perform changes gradually, [v_(ne)] is not in the candidate set.

Regardless of the ranking of Parse-σ and *WeakV, it will always be better to parse prosodic structure than to delete.

(54)

və.ne	Parse-σ	*WeakV	Max
a. və.ne	W$_2$	L	
b. v_.ne	W$_1$	L	W$_1$
c. və(ˈne)	W$_1$	L	
d. ☞ (və.ˈne)		$_1$	

(55)

və.ne	*WeakV	Parse-σ	Max
a. və.ne		W*2*	
b. v_.ne		*1*	W*1*
c. ☞ və(ˈne)		*1*	
d. (və.ˈne)	W*1*	L	

Since we can't parse and delete at the same time, (v_.ne) is not a possible candidate at the first step. However, once we have already parsed a disyllabic foot with a weak schwa, we will delete:

(56)

(və.ˈne)	*WeakV	Max
a. (və.ˈne)	W*1*	L
b. ☞ (v_.ˈne)		*1*

However, if the schwa is in its own monosyllabic foot, *WeakV is never violated and deletion is harmonically bounded.

(57)

(ˌvə)(ˈne)	*WeakV	Max
a. ☞ (ˌvə)(ˈne)		
b. (ˌv_)(ˈne)		W*1*

Deletion is motivated if and only if a schwa has been parsed into weak position. This means that variation in foot parsing will lead to variation in deletion. In the locally variable forms, each schwa has the option of being parsed into a dependent position; all and only dependent schwas will delete:

(58) *la queue de ce renard* 'this fox's tail'
(ˌdə)(ˌsə)(ˌrə)(ˈnar) → (ˌdə)(ˌsə)(ˌrə)(ˈnar)
(də ˌsə)(ˌrə)(ˈnar) → (d_ ˌsə)(ˌrə)(ˈnar)
(ˌdə)(sə ˌrə)(ˈnar) → (ˌdə)(s_ ˌrə)(ˈnar)
(ˌdə)(ˌsə)(rəˈnar) → (ˌdə)(ˌsə)(r_ ˈnar)

Because all and only dependent schwas delete, the parses in (47) above lead to the attested options for deletion. In each case, the deletion of schwa is determined by its role in foot structure.

3.3 Summary

In summary, locality in optional schwa deletion in French arises through variation in prosodic structure. Feet are generally monosyllabic, but schwa may optionally be parsed as a dependent in a disyllabic foot. All dependent schwas delete.

A prosodic account of French schwa deletion allows for derivations that result in all and only the possible deletions — there is a possible foot parsing where each schwa that may delete is eligible for deletion. A possible alternative would be to model schwa deletion as simple markedness/faithfulness interaction (*ə vs. Max); however, candidates which delete only a single schwa would tie with each other, and TETU effects would be expected (for example, deletion of only the schwa whose flanking consonants make the least-marked sonority sequence).

The data discussed here represent a subset of the data; schwa deletion is a complex phenomenon. I will discuss some of the relevant issues here, and suggest potential solutions.

Deletion is blocked when schwa is preceded by a sequence of two or more consonants, but is not blocked when a similar sequence of consonants follows schwa. Noske (1993) suggests that stray onsets left over from schwa deletion are resyllabified as the coda of preceding syllables; *ComplexCoda can therefore block deletion,[18] and Lamontagne (1996)'s relativized Contiguity (domain-specific and juncture-specific versions of the constraint) can prevent resyllabification as part of the onset of the following syllable.

Steriade and Fougeron (1997) provide evidence that schwa deletion in French does not erase a lexical schwa-zero contrast; this is not unexpected, given the above discussion, since the resulting syllabic structure in the examples with deletion will be different from the examples where there was no schwa to delete. See McCarthy (2003) on Bedouin Arabic and Kager (1997) on Macushi Carib for other cases where deletion results in something resembling contrastive syllabification.

Additionally, in most varieties of French, schwa is phonetically identical to the vowel [œ] (Dell, 1973; Côté and Morrison, 2007; Walker, 1993) — in some sense, it is an abstract segment differentiated from [œ] only by its alternation with zero. I have assumed that schwa is an abstract, featureless vowel; one possible way to account for its resemblance to [œ] is to suggest that schwa is augmented in contexts where it cannot or does not delete.

Schwa is not permitted word-finally, but [œ] is — word-final alternations between schwa and zero exist, but it is unclear whether those cases represent deletion of an underlying schwa or whether instead they are better categorized as participants in various schwa epenthesis processes that occur in French.

Finally, frequency of deletion varies lexically (Pater, 2008a) and as a function of speech rate (Dell, 1973). Because this analysis makes use of existing theories of variation, extant accounts of lexical and speech rate variation situated within those theories are applicable here as well.

These are important issues, and any full account of French schwa deletion must contend with them. The purpose of this section has not been to provide such an account, but rather to illustrate that the local variation we see in the process can be accounted for in a serial framework.

[18]In order for syllabification to block deletion, it is necessary that resyllabification of a stray consonant may be done in the same step as other processes; see McCarthy (2008b).

4 Globality and Warao Labial (de)Voicing

In the previous sections, the monotonic nature of prosodic structure building was necessary for the attested locally variable forms to surface; without strict inheritance (or undominated faithfulness), the derivation would be unable to converge on candidates that mix and match violations of opposing markedness constraints.

The prediction, then, is that non-monotonic phenomena will vary globally rather than locally.[19] In a context where two opposing markedness constraints are in variation, and both dominate the relevant faithfulness constraint(s), we should see consistent preferences expressed throughout the entire domain.

Warao provides an example of just such a situation. Labial obstruents vary freely with respect to voicing, and Osborn (1966) notes that "In words ... with two or more occurrences of /p/, the allophones are consistently [b] or [p] for each utterance of the word. If the first occurrence of /p/ in the word is [b], the following occurrence(s) will be [b]. If the first occurrence is [p], the following occurrence(s) will be [p]." For example:

(59) /paro + parera/ 'weak'

a. **p**aro**p**arera

b. **b**aro**b**arera

c. ***p**aro**b**arera

d. ***b**aro**p**arera

The globally variable options — all voiceless or all voiced — are attested, but variants combining voiced and voiceless labials are not permitted. Section 4.1 will discuss the opposing markedness forces at work in Warao and demonstrate how Serial Variation applies in this context, and Section 4.2 will discuss in more detail how non-monotonicity in voicing alternations leads to global variation in Warao.

4.1 Markedness and Labial Voice

The Warao obstruent inventory includes /p/, /t/, /k/, and /k^{w}/ — of these, /p/ is the only one that displays a voiced allophone; all other obstruents always surface as voiceless. The general voicelessness of obstruents is unsurprising typologically; a markedness constraint demanding voicelessness in obstruents dominates a faithfulness constraint to voicing:

(60) *ObsVoice: Assign one violation mark for every voiced obstruent.

(61) Ident(Voice): Assign one violation mark for every segment that is [α voice] in the input and [-α voice] in the output.

This ensures that obstruents will surface as voiceless:

[19] Many thanks to Anne-Michelle Tessier for insightful discussion on this point.

(62)

...d...	*ObsVoice	Ident(Voice)
a. ...d...	W*1*	L
b. ☞ ...t...		*1*

Voicing in labials is brought about not by faithfulness but by an opposing markedness constraint, which prohibits voiceless labials:

(63) *p: Assign one violation mark for every voiceless labial.

Voiceless labials often pattern as marked cross-linguistically. Arabic forbids them entirely (but does have voiced labials) and Japanese allows them only in geminates or in loanwords.

If *p and *ObsVoice are in variation, we can get variation between voiced and voiceless labials — because faithfulness is dominated, other obstruents will remain voiceless, but the special status of labials provokes them to become voiced when *p ≫ *ObsVoice.

If the input is /paro+parera/, with voiceless labials,[20] a ranking of *ObsVoice ≫ *p on the first step will converge immediately and will give us voiceless labials in the output.

(64) All voiceless: (59a), [**p**aro**p**arera]
Step 1: *Convergence*

/paro+parera/	*ObsVoice	*p	Id(Voice)
a. ☞ **p**aro**p**arera		*2*	
b. **p**aro**b**arera	W*1*	L*1*	W*1*
c. **b**aro**p**arera	W*1*	L*1*	W*1*

We can get voiced labials with a consistent ranking of *p ≫ *ObsVoice. On the first step, *p compels us to voice one of the voiceless labials. According to this constraint set, [parobarera] and [baroparera] are tied — for the purposes of this analysis, it doesn't matter which one we voice first, since both will eventually become voiced. The choice between the two becomes a matter of Emergence of the Unmarked.

An anonymous reviewer points out that, while a constraint against intervocalic voiceless segments would prefer [parobarera], there seems to be no known constraint that would, in parallel OT, prefer a mapping of [paroparera] → [baroparera]. If this is true, it remains true in HS. Since the candidates tie on *p, the choice falls to lower ranked constraints; if no lower ranked constraints prefer [baroparera] for this input, then the derivational path from [paroparera] → [barobarera] will always pass through [parobarera] rather than [baroperera].

Again, because we can only perform one change at a time, we can't voice both at once — [barobarera] is not yet a competitor. On the second step, *p compels us to once again voice the remaining labial. Finally, on the third step, *p's ranking prevents us from satisfying *ObsVoice by devoicing one of the labials we have previously voiced. The derivation converges, because the faithful mapping best satisfies the higher-ranked markedness constraint.

[20]Because this is allophonic, we do not know what the input is, but because faithfulness is low-ranked, the input is of little importance.

(65) All voiced: (59a), [**b**aro**b**arera]

Step 1

/paro+parera/	*P	*ObsVoice	Id(Voice)
a. **p**aro**p**arera	W_2	L	L
b. ☞ **p**aro**b**arera	*1*	*1*	*1*
c. ☞ **b**aro**p**arera	*1*	*1*	*1*

Step 2

baro**p**arera	*P	*ObsVoice	Id(Voice)
a. **p**aro**p**arera	W_2	L	*1*
b. **b**aro**p**arera	W_1	L_1	L
c. ☞ **b**aro**b**arera		*2*	*1*

Step 3: *Convergence*

baro**b**arera	*P	*ObsVoice	Id(Voice)
a. **b**aro**p**arera	W_1	L_1	W_1
b. **p**aro**b**arera	W_1	L_1	W_1
c. ☞ **b**aro**b**arera		*2*	

Intermediate variants, which mix voiced and voiceless labials, can — and in fact must be — selected as the local optimum at intermediate steps in the derivation. There is a possible mapping from /paroparera/ to [parobarera]:

(66)

/paro+parera/	*P	*ObsVoice	Id(Voice)
a. **p**aro**p**arera	W_2		
b. ☞ **p**aro**b**arera	*1*	*1*	*1*
c. ☞ **b**aro**p**arera	*1*	*1*	*1*

However, in order for [parobarera] to surface as an attested output, we must be able to converge on it. With faithfulness crucially dominated, [parobarera] will never map to [parobarera]:

(67)

paro**b**arera	*P	*ObsVoice	Id(Voice)
a. ☞ **p**aro**p**arera	*2*		*1*
b. **p**aro**b**arera	*1*	*1*	
c. ☞ **b**aro**b**arera		*2*	*1*

The locally variable form [baroparera] must compete with [paroparera] and [barobarera], which harmonically bound it. Thus, locally variable forms will never surface as optimal in Warao, and variation is correctly predicted to be global.

This is not to say that a language Warao′, with locally variable voicing, is impossible — such a language would be the result, for example, if *ObsVoice were in variation with Id(Voice). What forces globality in Warao is the fact that both markedness constraints must dominate faithfulness — in Warao′, locality is possible because markedness and faithfulness are in variation. However, in real Warao, the only possible voiced obstruents are labials — *ObsVoice must dominate Id(Voice), and variation is the result of variable interaction between two markedness constraints.

4.2 Non-Monotonicity and Convergence

In Warao, unlike in Bengali or French, it is possible to undo what we've done on a previous step. In order to understand how this predicts globality, we need to take a closer look at what is required for a derivation to converge.

With monotonic prosodic structure building, convergence will occur when it is no longer harmonically improving to build additional prosodic structure. Re-ranking of markedness constraints does not affect this: candidates which undo structures that have since been deemed sub-optimal are not possible members of the candidate set. The only possible options are to continue parsing or to stop (converge).

In interactions between markedness and faithfulness constraints,[21] convergence will occur when unfaithful mappings have eliminated all markedness violations, or when a ranking of faithfulness over markedness compels a faithful mapping.

With non-monotonic interactions between two markedness constraints, however, the conditions for convergence are a bit more complex. One of the opposing markedness constraints must become fully satisfied, and then that constraint must be subsequently ranked above its competitors. Because locally variable forms do not fully satisfy any of the competing markedness constraints, and because the candidate set includes candidates that do, the full satisfaction prerequisite for convergence will only be met by globally variable forms.

The requirement that a constraint be once again ranked above its competitors after achieving satisfaction may at first seem troubling: it relies on ranking the same constraint above its competitors multiple times in a row, and there is no hard guarantee that this will happen.

There is, however, a probabalistic guarantee: we will eventually have the same ranking multiple times in a row, just like flipping a coin enough times will eventually result in heads twice in a row. After some number of steps, the probability of having converged will approach 1. With two constraints, each ranking of them equally likely, probability of failing to converge at each step in a form with one locus of variation is .5 — in this situation, each step of the derivation constitutes an independent trial, and the probability of having failed to converge by step n can be calculated with the following equation:[22]

[21] None of the cases discussed in this paper involve markedness/faithfulness interactions, but possible examples include ATR spreading in Vata (Kaye, 1982), vowel-glide sequences in Armenian (Khanjian, 2008), and vowel reduction in Shimakonde (Liphola, 2001).

[22] When each ranking of constraints is not equally likely, or when there are multiple constraints in variation

(68) p(fail to converge through step n) = p(fail to converge at each step)^n

Thus, probability of failing to converge at step 1 is .5, at step 2 is .25, at step 3 is .125, and so on. We reach $p < .01$ by step 7.

With two loci of variation, we must take into account the fact that on certain steps — the steps where the local variant is the input — the probability of failing to converge is 1. Calculations of probability are now sensitive to whether or not the input is a local variant, and whether the step we're on is even or odd, but it is still relatively straightforward (here the probability of failing to converge at each step is based on the steps where it is possible, which is still .5):

(69) p(fail to converge through step n), with global input (e.g. [paroparera])

i. If n is even
= p(fail to converge at each step)$^{\frac{n}{2}}$

ii. If n is odd
= p(fail to converge at each step)$^{\frac{n+1}{2}}$

p(fail to converge through step n), with local input (e.g. [parobarera])

i. If n is even
= p(fail to converge at each step)$^{\frac{n}{2}}$

ii. If n is odd (and $n > 1$)
= p(fail to converge at each step)$^{\frac{n+1}{2}}$

In general, the probability of having failed to converge in words with two loci will take approximately twice as many steps to reach $p < .01$ as words with one locus.

While this is not a hard guarantee of convergence, the probability of nonconvergence becomes vanishingly small rather quickly. In other words, convergence is probabalistically guaranteed.

4.3 Summary

To summarize, the non-monotonic interaction of opposing markedness constraints in Warao leads the variation in labial (de)voicing to be global rather than local.

Serial Variation predicts locality in variation with monotonic processes and with variable markedness/faithfulness interactions, but globality in cases where markedness/markedness interaction is not restricted by monotonicity.

5 Comparison with Alternative Accounts

In this section, I'll illustrate a number of possible alternative accounts of locality in variation, and I'll show that the proposal made in this paper provides superior empirical coverage.

such that each outcome is no longer equally likely, the trials (steps) are no longer independent. Calculating the exact probability of having failed to converge by step n is therefore more complex, but the overall generalization still holds: within a relatively small number of steps, the probability of having failed to converge becomes vanishingly small.

5.1 Alternative Theories of Variation

The theories of phonological variation which are most problematic for local variation in Parallel OT all rely on variable constraint ranking. Does using a different theory of variation resolve the problem?

Coetzee (2006) casts variation in terms of gradient distinctions among sub-optimal candidates. Candidates that violate constraints which are above a certain threshold in the hierarchy are never optimal, but below that threshold the ranking of constraints determines relative frequency rather than grammaticality. This allows the mixed variants discussed above to become winners, but only because it allows *any* harmonically bounded candidate to win some of the time (as long as it does not have violations above the threshold). Adjusting assumptions about GEN to eliminate this problem, as Coetzee suggests, also eliminates the theory's ability to predict local optionality.

A Maximum Entropy model of variation within OT (MaxEnt/OT) (Goldwater and Johnson, 2003; Jäger and Rosenbach, 2006) permits locally optional candidates to win some of the time (Jesney, 2007). This is because MaxEnt/OT assigns some probability to all candidates, harmonically bounded or not. The probabilities assigned to harmonically bounded candidates are extremely low, which prevents them from surfacing enough to be problematic, but this means that harmonically bounded candidates in local optionality should also be vanishingly rare.

Both these theories suffer from the same essential difficulty — there is no reliable way to distinguish the harmonically bounded candidates we want (the locally optional candidates) from the harmonically bounded candidates we would like to remain harmonically bounded. We are left with the choice between an inability to account for local optionality or a loss of a great deal of the typological restrictiveness of the theory.

Kaplan (2008) accounts for local optionality via markedness suppression. In his account, a markedness constraint flagged as optional in a constraint hierarchy may undergo erasure of any number of violation marks incurred by the candidates under its evaluation. This is reminiscent of Anderson (1974)'s proposal for implementing optionality in generative rules: "If the rule is optional, eliminate from consideration any or all of the segments which could undergo the rule, together with their associated environments."

However, marking a constraint as optional impacts the relationships between that constraint and every constraint it dominates — we lose the ability to limit variation to conflicts between a particular subset of constraints, and we rob the optional constraint of its ability to invariantly dominate other constraints. Furthermore, it is doubtful that markedness suppression would be able to serve as a viable theory of variation in general. Accounting for various frequency effects or positional faithfulness effects would require either significant additions to the theory or combination with the theories of variation currently capable of handling such effects.

5.2 Positionally Indexed Constraints

The most empirically successful extant theory of local variation in OT is found in Riggle and Wilson (2005). They propose a solution based on positionally indexed constraints; at EVAL, a constraint is split into versions indexed to each locus of violation. When two constraints are in variation, their positionally-indexed versions may freely permute, allowing violations at one location to be assessed separately from violations at another. This predicts local optionality. However, it has the disadvantage of requiring a special mechanism specific to the phenomenon. More importantly, this model fails to account for cases where variation can be global.

With positionally indexed constraints, feature harmony constitutes the only context where global variation is predicted to be possible; a high-ranked constraint demanding agreement at different loci could force the choice of variant at each locus to be the same. The prediction, then, is that globality should occur only in cases of harmony.

Riggle and Wilson acknowledge that Warao poses a problem for this prediction, because it cannot be easily analyzed as harmony. In order for a harmony constraint to force covariation in Warao, it would need to require that non-adjacent consonants agree in voicing.[23]

Rose and Walker (2004) and others have argued for the existence of long-distance voicing assimilation. However, of the examples that serve as the basis of this argument, only one language — Kera — represents a productive synchronic process (rather than e.g. a static co-occurrence restriction on roots). Pearce (2006) has argued that what appears to be spreading of voicing in Kera is in fact best accounted for as tone spreading — low tones spread leftward in Kera, and obstruent Voice Onset Time (VOT) is shortened as a result of lowered F0. This shortened VOT leads voiceless obstruents to be perceived as voiced.

In Warao, covariation of labial voicing cannot be similarly attributed to correlation with the spreading of some other feature. In the absence of any convincing evidence for synchronic long-distance consonant-consonant voicing agreement, it remains implausible to propose such a constraint simply to account for the facts in Warao.

Serial Variation predicts globality in Warao from the type of constraint interaction involved — markedness/markedness variation with faithfulness crucially dominated. Other potential examples of globality resulting from this kind of interaction include Bole (Schuh, 2002), where sibilants may be either [+anterior] or [-anterior] but must covary, and Catalan (Mascaró, 2008), where vowels with variable height must covary.

Serial Variation, which requires no additional phenomenon-specific mechanism, is based on the observation that the predicted impossibility of local optionality in OT is a direct consequence of parallel evaluation. Serial Variation is also able to predict global variation under particular circumstances.

5.3 Cyclicity

For the cases of local variation discussed above, implementing a multiple-rankings theory of variation in a cyclic version of OT (e.g. Kiparsky 2000) would capture the relevant facts; for example, Truckenbrodt (2002) provides a cyclic account of local variation in Bengali MiP assignment.

A cyclic account of local variation predicts that loci in different cyclic domains may vary locally, but within the same cyclic domain we should see a consistent variant chosen. However, some examples of local variation occur within a single cyclic domain. For example,

[23]Many thanks to Donca Steriade and Adam Albright for insightful discussion on this point.

Riggle and Wilson (2005) point out the case of palatalization in Miya, where variation is local even in monomorphemic contexts.

Riggle and Wilson also argue against a cyclic account on the basis of two properties of locally variable plural marking in Pima. In multi-word compounds, plural marking can be realized on any non-empty subset of the stems:[24]

(70) ʔus-kàlit-váinom ‘wagon-knife’ (lit. ‘tree-car-knife’)

a. * ʔus-kàlit-váinom

b. ʔu**ʔu**s-kàlit-váinom

c. ʔus-kà**k**lit-váinom

d. ʔus-kàlit-vá**pa**inom

e. ʔu**ʔu**s-kà**k**lit-váinom

f. ʔus-kà**k**lit-vá**pa**inom

g. ʔu**ʔu**s-kàlit-vá**pa**inom

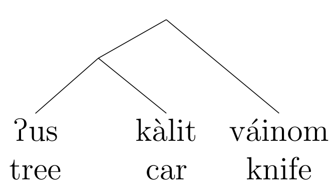
h. ʔu**ʔu**s-kà**k**lit-vá**pa**inom

The plural marker cannot fail to appear (70a), but it may appear on any one stem (70b-d), any two stems (70e-f), or all three stems (70g). All of the options are semantically equivalent, and the full range of options is available regardless of the internal head structure of the compound. Additionally, stems which permit allomorphic variation in plural marking outside of compounds must use the reduplicative plural marker within compounds.

Riggle and Wilson point out that the requirement that the plural marker must be realized at least once, and the restrictions on allomorphy, pose problems for cyclic versions of OT.

Cyclic theories commit to a particular starting-point based on morphological constituency, so forcing morpheme spellout on the first cycle would result in the plural marker appearing consistently on one particular stem. For example, with “wagon-knife” we can assume a structure as in (71):

(71) ‘wagon-knife’



We begin with the head of the innermost constituent, [kàlit]. We have, essentially, two options — we can realize the plural marker, or fail to realize it.

If we realize the plural marker, the possible end results will all include plural marking on [kàlit], and we won’t be able to get the forms in (70b), (70d), or (70g).

If we have the option of failing to realize the plural marker in [kàlit], there’s nothing to prevent us from exercising that option at subsequent cycles — we can’t look ahead to see

[24]The plural marker in Pima is reduplicative, and infixed — see Riggle (2006) for a discussion of the phonological properties of the reduplicant.

that we'll have another opportunity to spell out the morpheme — and we would predict the unattested form in (70a).

Furthermore, at this point in the derivation, information about whether or not the stem is part of a compound is inaccessible; there is no way of effectively restricting allomorph selection in these contexts.

Harmonic Serialism crucially differs from cyclic theories of OT in two relevant ways: first, the harmony of the entire form is evaluated at each step, even though only a single change can be made. Second, the starting-point is determined via constraint ranking, not cyclic relationships within morphological structures.

This means that the form's status as a compound may be available at the point of allomorph selection (see (Wolf, 2008) for a serial account of suppletive allomorphy). It also means that, at the first step, we have more options than simply realizing or failing to realize the plural marker in [kàlit]. Competitors include realizing the plural marker on [ʔus] or on [váinom] — it is no longer true that in order to get those options, we need to fail to realize the plural on the first step. The requirement that the plural must be realized at least once is not a problem for HS.

6 Conclusion

In this paper, I have proposed Serial Variation, a combination of Harmonic Serialism and a multiple-rankings theory of variation like Partially Ordered Constraints or Stochastic OT.

A HS derivation proceeds in step-wise fashion, undergoing multiple passes through a Gen→Eval loop, with a single change evaluated at each step. At each invocation of Eval, some total order is imposed on a language's incompletely ordered hierarchy, resulting in variation when constraints preferring opposing outcomes are variably ranked with respect to one another.

This proposal makes strong predictions about when variation should be local (allowing a different variant to be chosen at each locus of the process) and when it should be global (forcing all loci to choose the same variant).

Because we are evaluating only a single change at each step, and because the ranking according to which we are evaluating that change may differ from step to step, variation is predicted to be local. Because we must also be able to converge on a locally variable form, locality is predicted in cases where a process is monotonic (either through a universal requirement or undominated faithfulness) and when markedness and faithfulness are in variation. Globality, on the other hand, is predicted when opposing markedness constraints are variably ranked and both dominate faithfulness — in those cases, convergence on locally variable forms is impossible.

These predictions are borne out. Bengali Minor Phrase assignment and French schwa deletion are locally variable and involve monotonic structure building. Warao labial (de)voicing is globally variable, and involves markedness-markedness interaction.

In addition, the proposal advanced in this paper requires no additional theoretical mechanisms, because it results from the combination of existing independently motivated theories.

Acknowledgements

I am grateful for feedback from three anonymous reviewers and the associate editor, whose comments helped to improve this paper. This work has benefited enormously from feedback from John McCarthy and Joe Pater, who read and commented on numerous previous versions of this paper. This work has also benefited greatly from discussions with Adam Albright, Emily Elfner, Caren Rotello, Lisa Selkirk, Donca Steriade, Anne-Michelle Tessier, the UMass Phonology Group, and audiences at MUMM 2, UMMM 3, and OCP 6. All errors are, of course, my own.

References

Anderson S (1974) The organization of phonology. Academic Press, New York

Anderson S (1982) The analysis of French schwa. Language 58:121–138

Andreasson H (2004) Une contrainte de fidélité flottante pour le traitment du schwa et du la liason dans le canton de Vaud. In: Proceedings of PFC 3

Angoujard JP (2006) Phonologie déclarative. CNRS

Anttila A (1997) Deriving variation from grammar. In: Variation, change and phonological theory, Amsterdam: John Benjamins

Anttila A (2007) Variation and optionality, Cambridge University Press, chap 22, pp 519–536

Basbøll H (1981) Metrical theory and the French foot. In: Dressler W, Pfeiffer O (eds) Phonologica 1980, pp 12–36

Boersma P (1997) How we learn variation, optionality, and probability. In: Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam, vol 21, pp 43–58, ROA-221

Boersma P, Pater J (2008) Convergence properties of a gradual learning algorithm for Harmonic Grammar, ROA 970

Coetzee A (2006) Variation as accessing "non-optimal" candidates. Phonology 23(3):337–385

Coetzee A, Pater J (2008) The place of variation in phonological theory, ROA 946

Côté MH, Morrison GS (2007) The nature of the schwa-zero alternation in French clitics: experimental and non-experimental evidence. Journal of French Language Studies 17:159–186

Dell F (1973) Les règles et les sons: Introduction à la phonologie générative. Paris: Hermann

Elfner E (2008) Prosody circumscription in Harmonic Serialism. Paper presented at HUMDRUM, Rutgers University

Elfner E (2009) Syllabification and stress-epenthesis interactions in Harmonic Serialism, ROA 1047

Eychenne J (2006) Aspects de la phonologie du schwa dans la Francais contemporain: optimalite, visibilite prosodique, gradience. PhD thesis, Universite de Toulouse-Le Mirail

Goldwater S, Johnson M (2003) Learning OT constraint rankings using a maximum entropy model. In: Spenader J, Eriksson A, Östen Dahl (eds) Proceedings of the Workshop on Variation within Optimality theory, Stockholm University, pp 111–120

Hayes B (1995) Metrical stress theory: Principles and case studies. Chicago University Press

Hayes B, Lahiri A (1991) Bengali intonational phonology. Natural Language and Linguistic Theory 9(1):47–96

Itô J (1986) Syllable theory in prosodic phonology. PhD thesis, University of Massachusetts, Amherst

Jäger G, Rosenbach A (2006) Maximum entropy models and stochastic Optimality Theory. Linguistics 44(5):937–971

Jesney K (2007) The locus of variation in weighted constraint grammars. Presented at the Workshop on Variation, Gradience and Frequency in Phonology, Stanford, CA

Jesney K (2009) Positional faithfulness, non-locality, and the Harmonic Serialism solution. In: Lima S, Mullin K, Smith B (eds) Proceedings of the 39th Meeting of the North East Linguistics Society (NELS 39)

Kager R (1997) Rhythmic vowel deletion in Optimality Theory. In: Derivations and constraints in phonology, Oxford University Press, pp 463–499

Kaplan A (2008) Noniterativity is an emergent property of grammar. PhD thesis, University of California, Santa Cruz

Kaye J (1982) Harmony processes in Vata. In: Projet sur les langues Kru, UQAM, pp 60–151

Kenstowicz M (1996) Quality-sensitive stress. Revista di Linguistica 9(1):157–188

Khan SuD (2007) Phrasing and focus in Bengali, poster presented at the International Congress of Phonetic Sciences Satellite Workshop on Intonational Phonology: Understudied or Fieldwork Languages, Saarbrücken, 5 August.

Khanjian H (2008) Stress dependent vowel reduction in Armenian, paper presented at UMMM3

Kiparsky P (1985) Some consequences of Lexical Phonology. Phonology Yearbook 2

Kiparsky P (2000) Opacity and cyclicity. The Linguistic Review 17:351–367

Kratzer A, Selkirk E (2007) Phase theory and prosodic spellout: The case of verbs. The Linguistic Review 24:93–135 de Lacy P (2006) Markedness: Reduction and preservation in phonology. Cambridge University Press

Lamontagne G (1996) Relativized contiguity, part 1: Contiguity and syllable prosody, ROA150

Liphola MM (2001) Aspects of phonology and morphology of Shimakonde. PhD thesis, The Ohio State University

Mascaró J (2008) Phonological and morphological effects of asymmetrical DP concord. Colloquium presented at the University of Massachusetts, Amherst

McCarthy J (2000) Harmonic Serialism and parallelism. In: Hirotani M (ed) Proceedings of NELS 30, GLSA Publications, Amherst, MA, pp 501–24

McCarthy J (2003) Sympathy, cumulativity, and the Duke-of-York gambit. In: The syllable in Optimality Theory, Cambridge University Press, pp 23–76

McCarthy J (2007a) Hidden Generalizations: Phonological opacity in Optimality Theory. London: Equinox Publishing

McCarthy J (2007b) Restraint of analysis. In: S Blaho MK P Bye (ed) Freedom of Analysis, Mouton de Gruyter, Berlin and New York, pp 203–31

McCarthy J (2008a) The gradual path to cluster simplification. Phonology 25(2):271–319

McCarthy J (2008b) The serial interaction of stress and syncope. Natural Language and Linguistc Theory 26(3):499–546, phonology

McCarthy J, Prince A (1994) The emergence of the unmarked: Optimality in prosodic morphology. In: Gonzàles M (ed) Proceedings of the North East Linguistics Society, GLSA, vol 24, pp 333–379

Nagy N, Reynolds W (1997) Optimality Theory and variable word-final deletion in Faetar. Language variation and change 9(1):37–55

Noske R (1993) A theory of syllabification and segmental alternation. Max Niemeyer Verlag

Osborn H (1966) Warao I: Phonology and morphophonemics. International Journal of American Linguistics 32:108–123

Pater J (2007) Local Harmonic Serialism, handout: revised excerpts for workshop presentations at CASTL, Tromsø, March 27-28, 2007.

Pater J (2008a) Lexically conditioned variation in Harmonic Grammar, presented at OCP-5, Toulouse-Le Mirail, 01-26-08

Pater J (2008b) Serial Harmonic Grammar, handout from Univeristy of Southern California Colloqium

Pearce M (2006) Evidence against voicing spread in Kera. UCL Working Papers in Linguistics 18:345–355

Prince A (1985) Improving tree theory. In: Proceedings of the Berkeley Linguistics Society, vol 11, pp 471–490

Prince A (1991) Quantitative consequences of rhythmic organization, ms; Brandeis University

Prince A (2000) Comparative tableaux, new Brunswick, NJ: Rutgers University. ROA-376

Prince A, Smolensky P (1993/2004) Optimality Theory: Constraint interaction in generative phonology. Blackwell

Pruitt K (2008) Iterative foot optimization and locality in stress systems, university of Massachusetts, Amherst. ROA-999

Reynolds W (1994) Variation and phonological theory. PhD thesis, Univeristy of Pennsylvania

Riggle J (2006) Infixing reduplication in Pima and its theoretical consequences. Natural Language and Linguistic Theory 24:857–891

Riggle J, Wilson C (2005) Local optionality. In: Proceedings of NELS 35, Amherst, MA: GLSA

Rose S, Walker R (2004) A typology of consonant agreement as correspondence. Language 80(3):475–531

Samek-Lodovici V, Prince A (2002) Fundamental properties of harmonic bounding. Tech. rep., RuCCS-TR-71

Schuh R (2002) Palatalization in west chadic. Studies in African Linguistics 31:97–128

Selkirk E (1978) The French foot: On the status of "mute" e. Studies in French Linguistics 1:24–28

Selkirk E (1980) Prosodic domains in phonology, Anma Libri, Saratoga, CA, pp 107–129

Selkirk E (1995) The prosodic structure of function words. In: Beckman J, Walsh Dickey L, Urbanczyk S (eds) Papers in Optimality Theory, GLSA Publications, Amherst, MA, pp 439–70

Selkirk E, Tateishi K (1988) Constraints on minor phrase formation in Japanese. In: Proceedings of the 24th Annual Meeting of the Chicago Linguistics Society, Chicago Linguistics Society, Chicago, pp 316–339

Steriade D (1988) Reduplication and syllable transfer in Sanskrit and elsewhere. Phonology 5:73–155

Steriade D, Fougeron C (1997) Does deletion of French schwa lead to neutralization of lexical distinctions? In: Euro-Speech 1997, Proceedings of the 5th European Conference on Speech Communication and Technology, University of Patras, vol 7, pp 943–947

Truckenbrodt H (2002) Variation in p-phrasing in Bengali. In: Linguistic Variation Yearbook, vol 2, John Benjamins

Vaux B (2003) Why the phonological component must be serial and rule-based., paper presented at the 77th Meeting of the Linguistic Society of America.

Walker DC (1993) Schwa and /œ/ in French. Canadian Journal of Linguistics 38(3):43–64

Wolf M (2008) Optimal interleaving: Serial phonology-morphology interaction in a constraint-based model. PhD thesis, University of Massachusetts, Amherst