Learning in Harmonic Serialism and the necessity of a richer base

Anne-Michelle Tessier and Karen Jesney

Phonology / Volume 31 / Issue 01 / May 2014, pp 155 - 178
DOI: 10.1017/S0952675714000062, Published online: 20 May 2014

Link to this article: http://journals.cambridge.org/abstract_S0952675714000062

How to cite this article: Anne-Michelle Tessier and Karen Jesney (2014). Learning in Harmonic Serialism and the necessity of a richer base. Phonology, 31, pp 155-178 doi:10.1017/S0952675714000062

Request Permissions: Click here
Learning in Harmonic Serialism and the necessity of a richer base*

Anne-Michelle Tessier
University of Alberta

Karen Jesney
University of Southern California

This paper reassesses the hypothesis that early phonotactic learning of constraint-based grammars relies on the Identity Map – i.e. it uses observed surface forms as the inputs which cause errors and drive learning via constraint reranking. We argue that this approach’s success is closely tied to Optimality Theory’s fully parallel grammatical evaluation. In the constraint-based derivational framework of Harmonic Serialism (HS; McCarthy 2000, 2007b), reliance on observed surface forms as inputs can block the discovery of ‘hidden rankings’ between markedness constraints, preventing the learner from discovering a restrictive grammar. This paper illustrates the problem, using a pattern of positional vowel restrictions in Punu (Kwenzi Mikala 1980), and considers the role of various learning assumptions. We conclude that hidden rankings are a fundamental obstacle to restrictive error-driven learning in any HS-like framework, and that learning in such frameworks inevitably requires consideration of some unattested surface forms as inputs, even at the earliest learning stages.

1 Introduction

It is a standard assumption in constraint-based phonology that learning involves choosing both a ranking of constraints and a correct set of underlying forms. While a language’s phonological properties can be uniquely described by its ranking, correct underlying forms for each lexical item are still crucial. Knowing the ranking \*VoicedObstruentCoda \( \gg \) Ident[voice] ensures that final /d/ maps to [t],

* E-mail: AMTESSIER@UALBERTA.CA, JESNEY@USC.EDU.

We wish to thank the three anonymous Phonology reviewers and the editors for their generous and challenging questions, and especially for their comments that compelled us to search for a real-world language example. For suggestions and feedback on this and related work, we also thank Michael Becker, John McCarthy, Claire Moore-Cantwell, Joe Pater, Robert Staubs, Matt Wolf and audiences at NELS 42, MIT, the University of Michigan and the University of Toronto.
but a speaker must also know the underlying voicing specification of every word-final obstruent to ensure that [rat] alternates correctly in suffixed forms, either as /rat+ə/ → [ratə] or as /rad+ə/ → [radə]. Learning underlying forms and rankings concurrently is particularly challenging, because assumptions about each can have consequences for the other; for relevant discussion, see Kager (1999), Boersma (2001), Tesar et al. (2003), Jarosz (2006), Apoussidou (2007), Tesar & Prince (2007), Pater et al. (2012), inter alia.

One prominent strategy for tackling this problem is to divide and conquer: first learning just the surface phonotactics, regulating what is and is not observed in surface forms without reference to morphology, and only later incorporating evidence from alternations. This partitioning of the learning task allows the learner to initially ignore underlying forms, by assuming that mappings are always faithful. At the first stage, learning a language with final obstruent devoicing will mean observing unanalysed words like [rat], [ratə] and [radə], and constructing a grammar that treats each observed surface form as a fully faithful parse of some underlying form: /rat/ → [rat], /ratə/ → [ratə], /radə/ → [radə]. We will call this approach LEARNING FROM THE IDENTITY MAP.

Learning from the Identity Map has been very useful in the study of OT grammars and their learnability for at least two reasons. First, this approach appears to mirror natural L1 development. As emphasised by Hayes (2004), perception studies have demonstrated convincingly that infants acquire extensive sensitivity to language-specific phonotactics before their first birthday (see, for example, the literature cited in Jusczyk 1997), while good evidence of alternations begins to emerge only at 14–17 months (see for example White et al. 2008). Very young learners must therefore acquire their knowledge of phonotactics without the paradigmatic evidence for unfaithful underlying forms – e.g. without knowing that [rat] and [radə] are related, and that they share the underlying representation /rad/.

Second, learning via the Identity Map in OT has proven quite successful in allowing a language’s surface data to be captured restrictively. Error-driven learners can acquire phonotactics within this framework by examining a current grammar’s errors, and reranking constraints via an algorithm which ensures that errors become suboptimal compared to the observed targets (§3 will provide careful illustration of this process). Some aspects of implementing this learning process have not yet been fully resolved (as observed by Prince & Tesar 2004 and Tessier 2007; see also Alderete & Tesar 2002), but at least the correct ranking of markedness constraints in OT is easy enough to determine from errors when the Identity Map is assumed. As such, learning based on the Identity Map might well be the most effective simplifying assumption for establishing

---

1 Tesar (2008) attributes this term, or rather the learning use of the Identity Map in OT, to Prince & Tesar (2004); the general approach has been in use since at least as far back as Smolensky (1996).
the crucial rankings that govern a language’s phonotactics. (See Tesar 2008, 2011 for discussion of the informativeness of different types of mappings in OT learning.)

This paper demonstrates, however, that ease of learning is not stable across constraint-based frameworks. In Harmonic Serialism (starting with McCarthy 2000, 2007b), the Identity Map is in fact not at all suited to reasoning from observed errors to correct rankings – even when only the ranking of markedness constraints is involved.

2 Capturing inventory restrictions in constraint-based grammars

Various theoretical mechanisms exist for encoding phonotactic restrictions, but in Optimality Theory and related approaches this job falls solely to the workings of constraints. The constraint ranking is responsible for mapping the rich base of all possible underlying forms onto a language-specific set of licit surface forms, without any additional restrictions on the sounds or sequences posited underlingly. For example, the claim that creaky front rounded vowels like [œ] are categorically absent from English is modelled by ranking the relevant markedness constraints above conflicting faithfulness constraints, thereby ensuring that any input /œ/ will map to a grammatical surface segment – perhaps a modal-voiced unrounded vowel, as in (1).

(1) Illicit input forms map to grammatical surface forms in parallel OT

<table>
<thead>
<tr>
<th>/œ/</th>
<th>*FRONTRdV;*CREAKYVOICE;IDENT[LR];IDENT[rd]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [œ]</td>
<td>*!</td>
</tr>
<tr>
<td>b. [o]</td>
<td>*!</td>
</tr>
<tr>
<td>c. [e]</td>
<td>*!</td>
</tr>
<tr>
<td>d. [e]</td>
<td>*!</td>
</tr>
</tbody>
</table>

Under this ranking, input /œ/ undergoes two featural changes, losing its original specifications for both creak and rounding – and OT’s parallel method of evaluation achieves these multiple changes in a single mapping. The fact that unrounding and loss of creak apply simultaneously is by now very familiar to OT practitioners, but it is still a peculiarity of the massively parallel candidate evaluation introduced by Prince & Smolensky (1993). In any framework where changes must be gradual, the ordering of those changes can have non-trivial consequences.

2 Throughout this paper we use the term ‘input’ in the OT sense, to mean a string that is fed to the grammar to return an optimal output. In the acquisition literature, ‘input’ can also refer to whatever data or evidence the learner receives about the target grammar, but we use the term in its OT sense throughout this paper in order to minimise confusion.
2.1 Capturing phonotactics in Harmonic Serialism

Harmonic Serialism (HS) is an alternative to parallel OT which was initially suggested by Prince & Smolensky (1993: 79), and which has been pursued in earnest in recent years, yielding some notable improvements in typological adequacy (see McCarthy 2008a, b, Pruitt 2010, 2012, Jesney 2011, Kimper 2011, Elfner to appear, Pater to appear; and also §4.3). This paper cannot hope to argue comprehensively for Harmonic Serialism as a grammatical framework; instead, the focus here is on the properties of HS that affect the learnability of its grammars, and particularly on its implications for learning from the Identity Map.

At the heart of HS is a standard Eval function: the grammar takes an input and selects an optimal output, based on the relative ranking of markedness and faithfulness constraints. Crucially, and unlike parallel OT, HS is a fundamentally derivational model, which typically requires multiple loops through Eval to get from underlying to surface form. While a language’s constraint ranking remains the same throughout the derivation, only the fully faithful form and forms that differ from the current input by a single change are considered as candidates at each step. Once an optimum is selected, it becomes the input to the next step, where a similarly limited set of candidates is considered. In contrast to parallel OT, the transition from input to output in HS is therefore gradual.

In a Harmonic Serialist grammar, phonotactic restrictions are under grammatical control, just as they are in parallel OT. Due to the gradual nature of the model, however, any input that is more than one change away from a licit surface form will necessarily pass through Eval more than once in order to reach its optimal output state. In the example above, input /θ/ would need to pass through the derivation /θ/ → o → [e] or the derivation /θ/ → e → [e], where each form differs from the previous by a single change.

Crucially, such multi-step derivations require additional rankings of the markedness constraints that are unnecessary in parallel OT. One possible HS ranking that can map /θ/ to [e] is given in (2).

\[(2) \ \text{*FrontRoundV} \gg \text{*CreakyVoice} \gg \text{Ident[Lar]}, \text{Ident[round]}\]

Given the ranking in (2), the derivation proceeds as follows. First, the HS Gen is fed an input, here /θ/, and returns a finite set of output candidates, each of which differs from the input by a single change (characterising the set of ‘single changes’ is a large part of working on the theory – see McCarthy 2010). In this example, Gen’s operations will be changing the

---

3 HS differs in this way from Stratal OT and similar frameworks used in Bermúdez-Otero (1999, 2003), Kiparsky (2000) and Rubach (2000). For an alternative implementation of HS that allows for limited changes to the constraint ranking between steps, see Kimper (2011).
vowel’s rounding and laryngeal specifications, so the candidate set for input /ø/ is (3).

\[ (3) \] /ø/ → [ø] 

- fully faithful [ø] violation of Ident[lar]
- violation of Ident[round]

Once the candidate set has been enumerated, Eval applies in its normal fashion to winnow the candidate set down to a single optimum. As the tableau in (4) shows, at this first step, the ranking in (2) prefers mapping /ø/ to [e]. The licit output segment [e] is not a member of the candidate set at this stage, because it requires two featural changes.

(4) Step 1: input /ø/ is mapped to the intermediate form [e]

<table>
<thead>
<tr>
<th></th>
<th>*FRONTRDV</th>
<th>*CREAKYVOICE</th>
<th>Ident[lar]; Ident[rd]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [ø]</td>
<td>![ ]</td>
<td>![ ]</td>
<td>![ ]</td>
</tr>
<tr>
<td>b. [ø]</td>
<td>![ ]</td>
<td>![ ]</td>
<td>![ ]</td>
</tr>
<tr>
<td>c. [e]</td>
<td>![ ]</td>
<td>![ ]</td>
<td>![ ]</td>
</tr>
</tbody>
</table>

Because the input and the output are non-identical, the output [e] is taken as the input to the following step. Fed to Gen, this produces a new candidate set; the tableau in (5) shows the result of this second iteration.

(5) Step 2: input /e/ is mapped to the intermediate form [e]

<table>
<thead>
<tr>
<th></th>
<th>*FRONTRDV</th>
<th>*CREAKYVOICE</th>
<th>Ident[lar]; Ident[rd]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [e]</td>
<td>![ ]</td>
<td>![ ]</td>
<td>![ ]</td>
</tr>
<tr>
<td>b. [ø]</td>
<td>![ ]</td>
<td>![ ]</td>
<td>![ ]</td>
</tr>
<tr>
<td>c. [e]</td>
<td>![ ]</td>
<td>![ ]</td>
<td>![ ]</td>
</tr>
</tbody>
</table>

Since /e/ → [e] only involves a change in laryngeal specification, [e] is an available output candidate at this step and, given the constraint ranking, is chosen as optimal. This form becomes the input for a third iteration, where the fully faithful form is selected. At this point the derivation is complete, and the third step’s output is the final surface form.

(6) Step 3: input /e/ is mapped faithfully to [e]: convergence

<table>
<thead>
<tr>
<th></th>
<th>*FRONTRDV</th>
<th>*CREAKYVOICE</th>
<th>Ident[lar]; Ident[rd]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [e]</td>
<td>![ ]</td>
<td>![ ]</td>
<td>![ ]</td>
</tr>
<tr>
<td>b. [e]</td>
<td>![ ]</td>
<td>![ ]</td>
<td>![ ]</td>
</tr>
<tr>
<td>c. [ø]</td>
<td>![ ]</td>
<td>![ ]</td>
<td>![ ]</td>
</tr>
</tbody>
</table>

In fact, many other candidates that differ by a single change will also be included in the candidate set – e.g. [ø̃], where the vowel is nasalised, and [o], where the vowel’s [back] specification is altered. For simplicity we will consider only those candidates on the shortest path to the licit output [e].
For the purposes of this paper, the most important element of the derivation above is that the ranking of the two markedness constraints dictates the order in which processes apply. It is because *FRONTROUNDV is highest-ranked that the first step in (4) involves unrounding of the vowel. If *CREAKYVOICE were highest-ranked, the intermediate form /ʊ/ would be selected at the first step instead, and the alternative path /ʊ/ → ə → [ɛ] would prove optimal. In this case, both paths map input /ʊ/ to output [ɛ], and so the ranking of the constraints affects only the internal workings of the grammar, not the surface vowel inventory.

There are many other circumstances, however, where the order of serial processes is indeed crucial in order for surface phonotactics to be adequately captured. The rest of this paper presents a specific example from one language, and demonstrates the challenge that such patterns present for any HS learner that acquires phonotactics via the Identity Map.\(^5\)

3 Learning a positional inventory restriction in Harmonic Serialism

3.1 The data: [+low] vowels in Punu

Our illustrative example comes from the distribution of low vowels in the Bantu language Punu, spoken in Gabon and the Republic of Congo. As described by Kwenzi Mikala (1980) and Hyman (2002, 2008), Punu has eight surface vowels – [i u e o ɛ ɔ a ə] – whose distribution is restricted by phonological and morphological factors. The focus here is on the distribution of the vowels [ɛ e ɔ o], which Hyman (2002, 2008) analyses together with [a] as phonologically [+low] (or ‘open’ in his terminology).

The vowels [ɔ] and [e] participate in a productive [ATR] harmony process. As illustrated in (7a) for the back pair, these vowels surface as [+ATR] [ɔ] when followed by [i], and as [−ATR] [ɔ] when followed by [a]. The vowels [ɔ] and [ə] are in free variation before [u], as shown in (7b).\(^6\)

(7) Punu ATR harmony (Kwenzi Mikala 1980)

a. [posi] ‘freshness’ [mbəmə] ‘python’
   [nosi] ‘bee’ [ubəkəsəna] ‘to kill each other’

b. [ɡɔyʊ] ~ [ɡɔyʊ] ‘arms’

The [+low] vowels are also subject to a positional restriction that limits them to radical (stem-initial) position. In (7) all instances of [ɔ] and [ə] are

\(^5\) The ordering of phonological processes has been extensively studied for decades – primarily, though not exclusively, in rule-based frameworks. The crucial ordering discussed in this paper is a feeding relation (as in Kiparsky 1968, 1973; see more recently Baković 2007), but the complicated relationship between the overall hidden ranking problem and types of rule-ordering relationships remains a topic for future research.

\(^6\) Tone is suppressed throughout.
stem-initial – either word-initial or following the [u-] infinitival prefix; Kwenzi Mikala (1980) provides similar data for [e] and [ɛ].

The [+low] vowel [a] respects the same basic distributional pattern, but can also appear in prefixes, and in certain additional morphological contexts that are set aside here (Kwenzi Mikala 1980: 10–11). It is this basic positional restriction, and specifically its effect on the [−ATR] vowels [ɔ] and [ɛ], that is the primary focus of this section.

The restriction of [ɔ] and [ɛ] to stem-initial position can be captured using two interacting markedness constraints. The first, *NON-INITIAL[+low], militates against [ɛ ɔ e o a] in positions outside the initial syllable of the stem. The second, *[−low, −ATR], militates against non-central [−low, −ATR] vowels, including [i] and [o], which are not found in the surface inventory of the language. Prefixal quirks aside, these two markedness constraints are always obeyed in Punu, and so must outrank the conflicting faithfulness constraints IDENT[low] and IDENT[ATR] in the final grammar. Any general constraints on vowel-height features that are violated in Punu, like *[+low], will need to be dominated by IDENT[low]. The HS ranking of this constraint set for Punu is given in (8).

\[
(8) \ast_{\text{NON-INITIAL}[+low]} \gg \ast_{[−low, −ATR]} \gg \text{IDENT}[low] \gg \\
[+low] \gg \text{IDENT}[ATR]
\]

The ranking of *NON-INITIAL[+low] above *[−low, −ATR] is crucial to the treatment of input [ɔ] and [ɛ] vowels in non-initial positions. Given a hypothetical input like /pasɔ/, the grammar must ultimately map [ɔ] onto one of the vowels allowed in non-initial position, and any such mapping requires changes to multiple features of the input vowel. Mapping non-initial /ɔ/ to [u], for instance, requires changes to both the features [low] and [ATR]. The ranking of these two markedness constraints ensures that /ɔ/ is able to map to [u] by way of intermediate [u].

Given input /pasɔ/, GEN provides the fully faithful candidate and the two unfaithful candidates [paso] and [pasu] as possible optima at the first step, but not yet the eventual winner [pasu]. With the ranking in (8), [pasu] is selected as optimal (9a). At the following step, the attested output form [pasu] is one change away from the new input and so is provided as a candidate by GEN and selected as optimal. With no further harmonically improving changes possible, the derivation converges at step 3 (not shown).

7 In absolute word-initial position, the [+ATR] vowels [e] and [o] can also arise through the coalescence of [a+i] and [a+u] (Kwenzi Mikala 1980: 9).

8 The vowel [ɔ], which appears in a range of positions, is unaffected by this constraint. We assume here that this is a consequence of its reduced quality; in Hyman’s (2002, 2008) feature system it is underspecified for all features.

9 Similarly, in Hyman’s (2002, 2008) feature system, mapping /ɔ/ to [ɔ] would require changes to the features [low] and [round], and mapping /ɛ/ to [ɔ] would require changes to the features [low] and [front].
(9) a. Step 1: input /paso/ is mapped to the intermediate form [pasu]

<table>
<thead>
<tr>
<th>/paso/</th>
<th>*Non-init[+low]</th>
<th>*[−low, −ATR]</th>
<th>Ident[low]</th>
<th>Ident[ATR]</th>
</tr>
</thead>
<tbody>
<tr>
<td>i.</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ii.</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>iii.</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

b. Step 2: input /pasu/ is mapped to the surface form [pasu]

<table>
<thead>
<tr>
<th>/pasu/</th>
<th>*Non-init[+low]</th>
<th>*[−low, −ATR]</th>
<th>Ident[low]</th>
<th>Ident[ATR]</th>
</tr>
</thead>
<tbody>
<tr>
<td>i.</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ii.</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>iii.</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The ranking of [+low]/Non-initial above *[−low, −ATR] is key to this result; with the opposite ranking of these two markedness constraints, the initial mapping of /paso/ to [pasu] would be blocked and /paso/ would surface faithfully. With these facts in mind, we now turn to the Punu learning challenge of discovering these necessary rankings within HS using the Identity Map.

3.2 Learning Punu vowel distribution via the Identity Map

Following Smolensky (1996) and much subsequent work, we assume an initial state for learning in which all markedness constraints outrank all faithfulness constraints (\{M\} \supset \{F\}). We assume further that this partial ordering is resolved into a total ordering every time the grammar is used (as argued for in Boersma 2009), so that all markedness constraints outrank all faithfulness constraints in the initial state, but vacillate in their ranking with respect to each other on every iteration of EVAL.\(^{10}\)

(10) Initial state for Punu phonotactic learning

\[
\text{*Non-initial[+low], *[−low, −ATR], [+low] \supset Ident[low], Ident[ATR]}
\]

From this \{M\} \supset \{F\} starting point, the Identity Map learning approach attempts to discover the language’s phonotactics by aiming to

\(^{10}\) For the purposes of this paper we simply assume that a full ranking of the constraints which is consistent with the established partial ordering is selected at random on each iteration of EVAL (cf. Anttila 1997, 2002). Alternatively, applying a small amount of noise to numerical constraint values, as in Stochastic OT (Boersma 1998, Boersma & Hayes 2001), can provide the same effect. As shown by Boersma (2009), what is crucial is that a full ranking of the constraints be employed, rather than pooling the marks of constraints within the same stratum. We return briefly to the effects of noise in §4.1.
reproduce observed forms. When these attempts fail, the resulting errors are used to inform learning – i.e. the reranking of the implicated constraints. The learner of Punu will thus begin by observing surface forms like those in (7), and attempting to parse them as inputs using the grammar in (10). The ranking in (11) below represents one possible full ranking of the constraints that is consistent with the initial \{M\} \succ \{F\} ranking.

Since [+low] vowels like [o] are never observed outside of stem-initial position, and [−low, −ATR] vowels like [i] and [u] are never observed at all, the constraints \*Non-initial[+low] and \*[-low, −ATR] are not violated by any form inferred via the Identity Map. The only markedness constraint violated by identity mappings is thus general \* [+low]. This is illustrated in (11), taking the observed surface form [posi] as input to the learner.

(11) Punu learner’s error given the initial-state ranking

<table>
<thead>
<tr>
<th>input</th>
<th>*[-low, −ATR]</th>
<th>*Non-init [+low]</th>
<th>*[+low]</th>
<th>IDENT</th>
<th>IDENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>/posi/</td>
<td>a. [posi] (winner)</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>b. [posi]</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>c. [pusi] (loser)</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

The error here is straightforward: while the target language allows the surface form [posi], the learner in the initial state instead raises the vowel, selecting the surface form [pusi]. Because [pusi] is the only candidate that violates none of the markedness constraints, it is the only optimum possible under any of the total rankings consistent with the \{M\} \succ \{F\} initial state. Following Tesar & Smolensky (1998, 2000) and much subsequent work, we refer to the observed target form [posi] as the ‘winner’ and the grammar’s current optimum [pusi] as the ‘loser’. This terminology reflects the fact that under the ranking of the target grammar in (8) these forms are the ‘winner’ and ‘loser’ respectively. Given the Identity Map assumption, the ‘winner’ is consistently the fully faithful form, while the ‘loser’ will be one of the other candidates.

The ranking knowledge to be extracted from the error in (11) is elucidated in (12) using an Elementary Ranking Condition (ERC) vector (Prince 2002). In this format, every column tracks whether a constraint prefers the winner (W), the loser (L) or neither (e).

(12) ERC vector for the Punu learner’s initial-state error

<table>
<thead>
<tr>
<th>input</th>
<th>winner</th>
<th>loser</th>
<th>*[-low, −ATR]</th>
<th>*Non-init [+low]</th>
<th>*[+low]</th>
<th>IDENT</th>
<th>IDENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>/posi/</td>
<td>posi</td>
<td>pusi</td>
<td>e</td>
<td>e</td>
<td>L</td>
<td>W</td>
<td>e</td>
</tr>
</tbody>
</table>
The Constraint Demotion Lemma (originating in Prince & Smolensky 1993: 148, and reworded with the terminology used in this paper in Prince & Tesar 2004: 255) characterises how an ERC vector’s winner can be guaranteed to beat its loser – namely, for each error one of the constraints that prefers the winner must outrank all the constraints that prefer the loser. Achieving such a ranking is the first goal of the learner. Additionally, the learner must ensure that the grammar generates as few unobserved forms as possible – i.e. that it remains restrictive, e.g. Tesar (1996). This goal of restrictiveness can be implemented as in Biased Constraint Demotion (Prince & Tesar 2004), with a persistent bias that favours all the initial state rankings, starting with \{M\} \gg \{F\}, until evidence is given to the contrary (see also Hayes 2004).

For this biased learner, the ERC vector in (12) demonstrates that one \(M \gg F\) ranking must be subverted: the only way to resolve the error is to rank the winner-preferring faithfulness constraint IDENT[low] above the loser-preferring markedness constraint *[+low]. This ERC vector also demonstrates, however, that the remaining two markedness constraints do not distinguish between winner and loser, leaving them free to remain at the top of the hierarchy. The learner’s new ranking after learning from the error in (12) is shown in (13).

\[(13) \text{Grammar after learning from (12)}\]
\[
\begin{align*}
\ast \text{NON-INITIAL}[+\text{low}], \ast [-\text{low}, -\text{ATR}] \gg & \text{IDENT}[\text{low}] \gg \ast [+\text{low}] \gg \\
& \text{IDENT}[\text{ATR}]
\end{align*}
\]

With the Identity Map approach, the ranking in (13) is the end state of Punu phonotactic learning. Neither of the two top-ranked markedness constraints is violated on the surface, and so they never assign any Ls within the learner’s ERC vectors. As a result, they remain undominated and unranked with respect to each other; no amount of additional Punu data can establish a ranking between them.

\[(14) \text{Punu learner’s grammar in parallel OT}\]

a. Faithful mapping of initial [+low] vowels after learning

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>i. ([mb\text{om}]) \ast</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ii. ([mb\text{om}]) \ast</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>iii. ([mb\text{um}]) \ast</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>iv. ([mb\text{um}]) \ast</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
b. Repair of non-initial [+low] vowels after learning

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>i. [pasɔ]</td>
<td>*!</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ii. [paso]</td>
<td>*!</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>iii. [pasu]</td>
<td>*!</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>iv. [pasu]</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

This situation is unproblematic for the parallel OT learner. As the tableaux in (14) show, when a full range of candidates is considered, this ranking is sufficient to ensure that [+low] vowels are tolerated in stem-initial syllables (14a), even as they are mapped unfaithfully in non-initial environments (14b). (For simplicity, only violations for the relevant vowel are shown.)

The end state of phonotactic learning in (13) does not, however, match the target HS grammar seen previously in (8). HS requires an additional ranking between the top two markedness constraints in order to restrict [+low] vowels to initial position. In the absence of a determined ranking, the learner must simply impose an ordering upon the unranked constraints. If, as in (15) below, the ranking *[-low, ATR] >> *Non-init [+low] is imposed, mappings like /pasɔ/ → pasɔ → [pasu] are blocked. The fully faithful candidate [pasɔ] is selected at the first step, and the derivation converges upon an unattested surface form.11

(15) Punu learner’s grammar in Harmonic Serialism

Incorrect ranking of markedness constraints leads to immediate convergence

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [pasɔ]</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b. [paso]</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>c. [pasu]</td>
<td>*!</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

11 An anonymous reviewer rightly points out that the alternative mapping of /pasɔ/ → paso → [pasu] is in part blocked because there is no markedness constraint which prefers [o] to [ɔ]. If such a constraint were included, and its ranking were sufficiently high, this alternative path might prove optimal. The most obvious constraint preferring [o] over [ɔ] is *[-ATR]. Other data demonstrates, however, that [ATR] is contrastive within the language, indicating that *[-ATR] must be ranked below IDENT[ATR]. The key examples here come from forms like [e-lab-i] ‘he sees’ vs. [o-lab-o] ‘he will see’ (Kwenzi Mikala 1980: 9), where the [-ATR] vowel [a] appears in stem-initial position both before the [+ATR] vowel [i] and before the [-ATR] vowel [ɔ] (the vowels [e] and [o] in absolute initial position of these examples are the result of prefix coalescence). Along with the IDENT[low] >> *[+low] ranking in (14), we therefore assume that the learner establishes an IDENT[ATR] >> *[-ATR] ranking during initial phonotactic learning, eliminating the possibility of an alternative derivation that first maps /pasɔ/ to *[paso].
The end state in (13) is agnostic between the correct HS ranking in (8) and the incorrect one in (15). There is no way for the learner to detect the need for *NON-INITIAL[+low] to dominate *[−low, −ATR] based on learning from the Identity Map. As a result, the learner’s end-state grammar is insufficiently restrictive; faced with an input that violates the language’s surface distributional pattern, the learner may incorrectly choose to map it faithfully.

Crucially, identifying this hidden markedness ranking is a problem just for HS. In parallel OT there is no need to rank these two markedness constraints with respect to each other, since both are always obeyed, and forms that diverge from the input in multiple ways can be immediately considered and deemed optimal (as illustrated above in (14)).

3.3 Diagnosing the problem of Punu vowels: blaming the Identity Map

We have already suggested that the HS phonotactic learner’s failure to learn hidden rankings comes from its reliance on the Identity Map, rather than, for example, the constraint-demotion procedure itself. This section defends this analysis by reasoning from success: that is, by asking what kind of error could teach the learner that the ranking *NON-INITIAL[+low] ≫ *[−low, −ATR] is necessary in Punu.

In the case of Punu, the crucial error must cause the demotion of *[−low, −ATR], while leaving *NON-INITIAL[+low] untouched. The error must therefore be one where, of the two markedness constraints, only *[−low, −ATR] is loser-favouring. In other words, the optimum selected by the current grammar – i.e. the loser – must contain fewer [−low, −ATR] vowels than the winner. Given the very limited candidate space considered thus far, this places [pasɔ] as the loser and [pasʊ] as the winner in the necessary error. In effect, the required winner ~ loser pair is the first step of the derivation in (9) – i.e. /pasɔ/ → pasʊ → [pasʊ].

(16) Necessary ERC vector

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>/pasɔ/</td>
<td>pasʊ ~ pasɔ</td>
<td>W</td>
<td>L</td>
<td>W</td>
<td>L</td>
<td>e</td>
</tr>
</tbody>
</table>

This error provides the crucial evidence that *[−low, −ATR] must be demoted. Pairing it with the error in (12), as in (17) below, allows the learner to discover the hidden ranking.
Two crucial ERC vectors for learning the Punu hidden ranking

<table>
<thead>
<tr>
<th>input</th>
<th>winner ~ loser</th>
<th>*Non-init</th>
<th>*[−low, −ATR]</th>
<th>*[+low]</th>
<th>IDENT [low]</th>
<th>IDENT [ATR]</th>
</tr>
</thead>
<tbody>
<tr>
<td>/posi/</td>
<td>posi ~ pusi</td>
<td>e</td>
<td>e</td>
<td>L</td>
<td>W</td>
<td>e</td>
</tr>
<tr>
<td>/pasɔ/</td>
<td>pasɔ ~ pasɔ</td>
<td>W</td>
<td>L</td>
<td>W</td>
<td>L</td>
<td>e</td>
</tr>
</tbody>
</table>

An algorithm that can reason from (17) to the target ranking was described informally above, but its workings will now be illustrated step by step. This algorithm could be either Biased Constraint Demotion (Prince & Tesar 2004) or Low Faithfulness Constraint Demotion (Hayes 2004); with respect to these errors, they will find the same ranking. Both of these algorithms are biased to install markedness constraints as high as possible while building the new grammar stratum by stratum, beginning at the top. With the ERC vectors from (17) in hand, the first step is given in (18).

(18) **Step 1**

Install all markedness constraints that prefer no losers, i.e. that assign only W or e.

*Result*

*+[low]/Non-initial ≫ all others

*After a stratum is built, remove any errors that have now been resolved.*

There is only one markedness constraint in (17) that is not loser-favouring on any comparison: *Non-initial+[low]. This constraint is installed in the top stratum, resolving the second error in (17). With only the first error left to resolve, the next stratum is built (shading indicates those constraints which have already been installed).

(19) **ERC vectors for creation of the second stratum**

<table>
<thead>
<tr>
<th>input</th>
<th>winner ~ loser</th>
<th>*Non-init</th>
<th>*[−low, −ATR]</th>
<th>*[+low]</th>
<th>IDENT [low]</th>
<th>IDENT [ATR]</th>
</tr>
</thead>
<tbody>
<tr>
<td>/posi/</td>
<td>posi ~ pusi</td>
<td>e</td>
<td>e</td>
<td>L</td>
<td>W</td>
<td>e</td>
</tr>
</tbody>
</table>

Step 1 then reappears.

(20) **Step 1 (repeated)**

Install all (uninstalled) markedness constraints that prefer no losers.

*Result*

*Non-initial+[low] ≫ *[−low, −ATR] ≫ all others

*After a stratum is built, remove any errors that have now been resolved.*
While *[-low, −ATR] does not prefer any losers, it also does not resolve the remaining error; its ‘e’ simply demonstrates to this learner that no evidence yet prevents the installation of *[-low, −ATR] in this relatively high-ranking stratum. In building the third stratum, the learner resolves the remaining error by installing the W-preferring faithfulness constraint.

(21) Step 1 (repeated)
Install all (uninstalled) markedness constraints that prefer no losers.

If all remaining markedness constraints prefer at least one loser, proceed to step 2.

Step 2
Install one faithfulness constraint that prefers only winners.\(^\text{12}\)

Result
*Non-initial[+low] \(\not\succ\) *[-low, −ATR] \(\not\succ\) Ident[low] \(\not\succ\) all others

After a stratum is built, remove any errors that have now been resolved.

With the installation of Ident[low], both errors in the ERC vector set have been resolved, and can be removed from the learning data. The ranking can then be completed purely on the basis of the learner’s general biases – i.e. the preference for markedness over faithfulness rankings.

(22) a. Step 1 (repeated)
Install all (uninstalled) markedness constraints that prefer no losers.

Result
*Non-initial[+low] \(\not\succ\) *[-low, −ATR] \(\not\succ\) Ident[low] \(\not\succ\)

*+[low] \(\not\succ\) all others

b. Step 1 (repeated)
Install all (uninstalled) markedness constraints that prefer no losers.

If all markedness constraints are installed, skip to step 3.

Step 3
Install all remaining faithfulness constraints and end.

Result
*Non-initial[+low] \(\not\succ\) *[-low, −ATR] \(\not\succ\) Ident[low] \(\not\succ\)

*+[low] \(\not\succ\) Ident[ATR]

\(^{12}\) This characterisation of step 2 is sufficient for our purposes, but overly simplistic in many cases. Both the Biased Constraint Demotion and Low Faithfulness Constraint Demotion algorithms spend much of their analytic energies determining exactly how faithfulness constraints should be installed when all remaining markedness constraints prefers at least one loser; see Hayes (2004) and Prince & Tesar (2004) for discussion.
The final result, in (22b), is a success, capturing the hidden ranking necessary in order to ensure that the HS grammar selects the correct Punu output forms.

This walk-through of the learner’s path based on the two ERC vectors in (17) demonstrates that error-driven HS learning can in fact find hidden rankings – but only with the correct data. Crucially, this requires an ERC vector like the one constructed in (16), where the winner [pasu] is not a possible surface form in the language being learned. Given the Identity Map assumption, this error should not be accessible to the learner.

Before moving on, we wish to emphasise that no observed surface form could substitute as the required winner in the ERC vector that drives learning of the hidden ranking *\text{NON-INITIAL}[^{+}\text{low}] \gg *[^{-}\text{low}, \text{ATR}]$. To create this ranking requires a candidate pair where *[^{-}\text{low}, \text{ATR}] is loser-favouring and *\text{NON-INITIAL}[^{+}\text{low}] is either winner-favouring or neutral. This means that the winner in this ERC vector must be [pasu], or some other non-observed form that includes a [^{-}\text{low}, \text{ATR}] vowel, while the loser must eliminate this vowel and so better satisfy *[^{-}\text{low}, \text{ATR}]$.

(23) Structure of the ERC vector required for learning the Punu hidden ranking

<table>
<thead>
<tr>
<th>input</th>
<th>winner ~ loser</th>
<th>*\text{NON-INITIAL}[^{+}\text{low}]</th>
<th>*[^{-}\text{low}, \text{ATR}]</th>
<th>*[^{+}\text{low}]</th>
<th>\text{IDENT}[^{low}]</th>
<th>\text{IDENT}[^{ATR}]</th>
</tr>
</thead>
<tbody>
<tr>
<td>/??/</td>
<td>paso ~ ??</td>
<td>W or e</td>
<td>L</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

Because there are no surface forms in the language that contain [^{-}\text{low}, \text{ATR}] vowels, the constraint that favours [pasu] and causes it to win in (23) cannot be a faithfulness constraint. Input /pasu/ never maps faithfully onto itself in the target language, and so *[^{-}\text{low}, \text{ATR}] must dominate at least some faithfulness constraints that would prefer faithful mapping of /pasu/. The winner in (23) must therefore result from an unfaithful mapping of some other input, and this unfaithfulness must be driven by a markedness constraint that conflicts with *[^{-}\text{low}, \text{ATR}]$. Within the limited space that we are considering here, the only input that meets these criteria is /pasO/, and the conflicting markedness constraint which prefers the winner [pasu] must be *\text{NON-INITIAL}[^{+}\text{low}].\textsuperscript{13} We thus arrive at the fully specified ERC vector in (16).

In effect, learning from the Identity Map is only a modestly successful approach to learning an HS grammar. In cases where the observed output forms motivate the demotion of a markedness constraint to a position below faithfulness – as with the \text{IDENT}[^{low}] \gg *[^{+}\text{low}] ranking of Punu – the appropriate ranking can be achieved. In cases where a necessary ranking exists between two markedness constraints that are consistently satisfied at the surface, however, the Identity Map proves

\textsuperscript{13} In fact, both *[^{+}\text{low}] constraints prefer the winner in this ERC vector, but we saw in §3.2 that learning from both ERC vectors in (17) will allow learning of the correct ranking.
inadequate. Denied any informative errors, the learner fails to establish the crucial hidden ranking, and the end-state grammar proves insufficiently restrictive.

4 Possible solutions for the phonotactic learner

If the Identity Map is not sufficient to learn an HS grammar’s hidden rankings, what conclusions can be drawn about the nature of the learner and the grammar? We see three possibilities: (i) determining hidden rankings cannot be the goal of the HS learner during phonotactic learning, (ii) the learner must abandon HS as a model of grammar or (iii) the HS learner must abandon the Identity Map from the beginning, even when just learning phonotactics. This section briefly discusses these options, including the scope of the problem, and concludes in tentative support of the third possibility.

4.1 The scope of the hidden ranking problem

The origin of the hidden ranking problem lies in the limitations that HS places on the construction of the candidate set. As we have seen, in HS only candidates that differ from the current input by a single change are considered at each step in the derivation; this constrains the application of processes whose markedness benefits are only evident after further processes have applied. Coupled with the Identity Map, this means that only a very limited set of potential ‘losers’ is available for consideration in the construction of ERC vectors. Learning from the Identity Map presents a problem, then, for any derivational approach that requires harmonic improvement at each step. The question for this section is whether any refinements to the grammatical framework might facilitate the discovery of hidden rankings.

At least initially, Optimality Theory with Candidate Chains (OT-CC; McCarthy 2007a – see also Walker 2010) might appear to provide candidates that include enough information to avoid the learning problem discussed in this paper. In OT-CC, candidates are sequences of forms (‘chains’) comprising a mapping from input to output, with some chains involving multiple changes. These changes are still constrained, however, as each step in a candidate chain must be gradual and harmonically improving according to the constraint ranking of the language. In both HS and OT-CC, reliance on harmonic improvement makes (hidden) rankings of markedness constraints crucial – and reliance on the Identity Map means that these rankings cannot be consistently discovered.

Altering the mode of evaluation from ranked to weighted constraints, as in Harmonic Grammar (Legendre et al. 1990, Smolensky & Legendre

14 We are grateful to two anonymous reviewers for discussion of the issues in this section.
2006; see also Potts et al. 2010, Pater to appear), likewise cannot eliminate the learning problem, provided that the requirements of gradualness and harmonic improvement are maintained. Indeed, if a learning model that assigns the same high initial weight to all markedness constraints and low initial weight to all faithfulness constraints is adopted (e.g. Goldwater & Johnson 2003, Jesney & Tessier 2011), precisely the same problem as discussed in §3 will arise. Without the necessary evidence to force separation in the weights of markedness constraints that are consistently obeyed at the surface, no hidden ranking (or, in this case, hidden weighting) can be learned. Depending upon the specific assumptions about the initial state and the implementation of the bias, the Identity Map can also introduce further complications in a weighted-constraint version of HS. Because all constraint violations contribute to the selection of optima in weighted-constraint models, even very low-weighted faithfulness constraints can play a role. When learning from the Identity Map, faithfulness constraints consistently favour winners, and are therefore subject to increases in their weights. In order to establish a crucial hidden relationship between markedness constraints, then, the learner must achieve a large enough separation between their weights to overcome any faithfulness constraints that prefer a fully faithful but illegal mapping, like the one in (15).

The failure to learn hidden rankings also cannot be solved by encoding the constraint ranking in absolute numerical values and adding noise to the system, as in Stochastic OT (Boersma 1998, Boersma & Hayes 2001 and subsequent works). If the critical markedness constraints have the same basic initial ranking value, noise will succeed in placing the constraints in the necessary configuration in some cases – 50% of the time if there are

As illustration of this effect, we simulated learning of constraint weights in Punu, using the Wilson/George MaxEnt learner available at http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool (also see Wilson 2006). As data, the learner was given the identity mappings /posu/ → [posu] and /pOsu/ → [pOsu], with a candidate set that included just the faithful winner and outputs that applied a single change to one of the vowels. An essentially infinite \( \sigma^2 \) was specified for the markedness constraints, reducing the penalty for relatively high weights; a \( \sigma^2 \) of 50 was specified for the faithfulness constraints, biasing their values to remain close to zero. The resulting weights of \( \star[-\text{low}, -\text{ATR}] \) and \( \star\text{\textsc{non-initial}}[+\text{low}] \) were 11.43 and 11.33 respectively, while \( \star[+\text{low}] \) retained a weight of 0-00. Crucially, the faithfulness constraints had non-zero weights at the end of learning – \( \nu(\text{\textsc{ident}[\text{ATR}]) = 2.78 \) and \( \nu(\text{\textsc{ident}[\text{low}]) = 2.25 \). With these weights, input /pOsu/, which is unattested in the target language and was not part of the learning data, maps faithfully to itself with a probability of 0.79. These results cannot be attributed to overfitting of the observed data, due to an excessively high \( \sigma^2 \) value for faithfulness constraints. Even when \( \sigma^2 = 50 \), the probability of the observed input /pOsu/ mapping faithfully is only 0-86, while the probability for /pOsu/ is only 0-94. Further decreasing the value of \( \sigma^2 \) reduces the extent to which the unobserved form /pOsu/ maps faithfully, but also reduces the extent to which the observed forms are mapped appropriately. With \( \sigma^2 = 10 \), for instance, unattested /pOsu/ maps faithfully with a probability of 0-54, while /pOsu/ and /pOsu/ map faithfully with probabilities of 0-65 and 0-83 respectively. In sum, given the Identity Map and the assumption of a limited HS-style candidate set, the weighted-constraint learner fails to establish the necessary constraint weights to ensure restrictiveness.
only two constraints whose ranking is crucial—but this will only be a temporary effect. Because these constraints are not implicated in any errors when learning from the Identity Map, their basic ranking values will not be altered, and the hidden ranking will not be learned. Including this type of noise within an HS framework may well be an effective means of capturing variation (e.g. Kimper 2011), and of ensuring that a full ranking of the constraints is employed by the learner on each iteration of \textsc{Eval} (Boersma 2009), but it does not provide a solution to the restrictiveness problem inherent to learning an HS grammar using the Identity Map.

We consider finally the possibility of solving the problem by changing the contents and the structure of \textsc{Con}. First, could the hidden ranking problem be avoided by better regulating the nature of the markedness constraints and their definitions? While this is a possibility in individual cases, it is not clear that it is a viable general solution. The constraints involved in the Punu hidden ranking are simple feature co-occurrence and positional markedness constraints of the type used in countless analyses. The required ranking of these constraints—with *\textsc{Non-initial}[-\textsc{low}] dominating *[-\textsc{low}, –\textsc{ATR}]—might suggest a revision to the learning algorithm biasing positional constraints toward higher ranking than general markedness constraints. As discussed initially by Prince & Tesar (2004), however, this type of bias cannot be reliably implemented, because in the general case there is no reliable means of determining a priori which of two constraints targets a more specific context within a given language. Determining this relative specificity is a particular challenge when the features targeted are non-identical. In the current example, we note that, of the two markedness constraints, the higher-ranked one is if anything more general in its featural scope—that is, *\textsc{+low} rules out many more vowels than *[-\textsc{low}, –\textsc{ATR}] (see also the arguments in Tessier 2007).

Second, could the problem be avoided by relying on positional faithfulness rather than positional markedness to restrict Punu’s non-low vowels to initial position? The answer here appears to be no. To suggest briefly why not, (24) provides a positional faithfulness translation of the Punu target ranking. The key here is the positional \textsc{ident}[^\textsc{low}]/\textsc{initial} constraint that is ranked above general *\textsc{+low} at the top of the hierarchy. This ensures that initial [+\textsc{low}] vowels, as in [pos], are preserved.

\begin{equation}
\textsc{ident}[^\textsc{low}]/\textsc{initial} \gg *\textsc{+low} \gg *[-\textsc{low}, –\textsc{ATR}] \gg \textsc{ident}[^\textsc{low}], \textsc{ident}[\textsc{ATR}]
\end{equation}

Crucially, this grammar also includes a hidden ranking: *\textsc{+low} \gg *[-\textsc{low}, –\textsc{ATR}]. Outside of initial position, the general *\textsc{+low} is now responsible for the first step in mapping unattested /pasO/ to [pasU] and then to [pasu]. This ranking is hidden, however, because outside of initial position there are no surface forms that violate either of these constraints, and therefore
there are no accessible errors that could motivate the constraints’ relative ranking. As long as the Identity Map assumption is maintained, the learning problem persists.

4.2 The need to learn hidden rankings without alternations

Given the results of the previous sections, it is worth emphasising that abandoning the goal of learning hidden rankings is not a viable option. Phonotactic learning must establish hidden rankings among undominated markedness constraints in order for various attested patterns to be captured. This is clear in Punu, where it appears that the morphology does not provide any alternations that could give evidence supporting the restriction on vowel heights outside of stem-initial position. The necessary hypothetical alternation would have to demonstrate that a surface form like [pasu] comes from the input /pasɔ/. While one could imagine scenarios that might provide this evidence – e.g. an opaque interaction whereby the underlying /ɔ/ triggers harmony before raising itself – Punu is not reported to have any such process. To deny the existence of hidden rankings learned solely from phonotactics, then, is to deny that Punu speakers should have any grammatical knowledge of the illegality of forms like /pasɔ/. In other words, it is to claim that their grammars can rule out non-initial mid vowels like [o], and also high [−ATR] vowels like [u], but cannot construct derivations that rule out both for a single input.

While we cannot provide word-likeness ratings from Punu speakers to demonstrate this knowledge, it is well established that native speakers of better-studied languages are highly sensitive to static phonotactic regularities in their lexicons (e.g. Pertz & Bever 1975, Moreton 2002, Hayes & Londe 2006, Berent et al. 2008). Experimental studies have demonstrated that these sensitivities are under grammatical control, with speakers showing particular responsiveness to patterns that are robustly attested cross-linguistically (e.g. Moreton 2008, Becker et al. 2011). Alternations, then, cannot be required in order for the grammar to instantiate phonotactic patterns of the type observed in Punu. The learner must be able to acquire the necessary rankings on the basis of static distributional evidence alone.

4.3 The trade-off: Harmonic Serialism vs. the Identity Map

Harmonic Serialism has several key advantages in typological scope and efficacy compared to classic OT. For example, it offers a means of avoiding too-many-solutions problems involving medial cluster reduction (McCarthy 2008a) and stress–syncope interactions (McCarthy 2008b), it better predicts the factorial typology of iterative footing and stress (Pruitt 2010, 2012) and it offers a solution to pathologies arising with positional faithfulness constraints in parallel OT (Jesney 2011). On the other hand, other recent work has suggested that the type of fell-swoop
derivations available in parallel OT are beneficial in capturing certain long-distance segmental processes (Walker 2010, Kaplan 2011; cf. Kimper 2012). Should we therefore consider the present paper’s learning result to be an argument for retaining classic OT’s fully parallel evaluation?

Rather than concluding in favour of one framework over the other, our current goal is instead to identify the fundamental properties that distinguish parallel OT and HS, and then to ask what type of learner can best accommodate those properties. With regard to the learning problem raised by hidden rankings, we see the most crucial difference between the two frameworks to be the scope of the output candidate set available in any one evaluation. The limitation of the candidate set to forms that differ from the current input by a single change means that derivational paths which might ultimately lead to greater overall well-formedness are blocked if each step in the derivation is not itself harmonically improving. This has two effects: the candidate set in HS is considerably smaller than in parallel OT, and the ranking of constraints that are always satisfied at the surface becomes crucial.

One consequence of the difference in the candidate sets’ scope might be that the child learning a parallel OT grammar is in greater need of a simplifying assumption like the Identity Map in order to identify informative errors and begin learning. The Identity Map assumption may be drastic in ignoring the possibility of unfaithful mappings in the target language, but it is sufficiently revealing to allow the parallel OT learner to learn key facts about the grammar’s phonotactics. The HS learner, on the other hand, encounters a more manageable candidate set for any given input, but must also explore a wider range of input–output mappings to ensure that hidden rankings are discovered and a sufficiently restrictive grammar is built. The upshot might be the following trade-off: if the phonological grammar entertains a smaller set of output options for any given input form, its learner needs a larger set of input options in order to appropriately rank the relevant constraints. The Identity Map cannot be an effective simplifying assumption when it fails to provide the inputs necessary for adequate phonotactic learning.

5 Future directions

This paper has demonstrated that learning a restrictive grammar within Harmonic Serialism requires that the Identity Map be abandoned, so that unfaithful underlying forms are considered even during phonotactic learning. One means of addressing this issue is to give the learner complete access to the entire rich base of possible inputs, and then construct a grammar that will ensure that each of these inputs map to some observed

16 These papers are primarily concerned with the gradualness requirements of OT with Candidate Chains (McCarthy 2007a).
surface form. Starting with Tesar & Smolensky’s (1998, 2000) work, assuming a rich-base approach has allowed for considerable progress on a number of OT learnability fronts; in particular, Jarosz’s (2006) Maximum Likelihood learner explicitly entertains the entire rich base when acquiring a lexicon and grammar. While effective in many ways, it remains to us unclear that a rich-base approach can be effectively adapted to the real child’s task of processing the impoverished data to which they are exposed. From our perspective, learning from the Identity Map was a highly effective starting point in parallel OT, and for the reasons discussed in §1, it remains worthwhile to consider some version of it in the HS context. The primary issue for future research on this question, then, is to establish a principled means of finding those input candidates that will allow the learner to make informative errors based on unfaithful mappings.

One avenue of investigation begins with the fact that the candidate set in HS is not only smaller than in parallel OT, but also more manageably structured. With this in mind, Tessier (2013) sketches an HS learning technique which examines candidates that lose out in competition with observed surface forms, and then reasons backwards from such loser forms to the markedness constraints responsible, iteratively looking among these constraints for crucial hidden rankings. This technique exploits the gradualness property of HS, by initially expanding the set of inputs considered to include a carefully selected subset of forms that differ from the learner’s current optimum by a single change. Staubs & Pater (to appear) tackle the issue of selecting underlying forms when learning an HS grammar along somewhat similar lines. In looking for an efficient solution to the problem of learning hidden structure and interacting processes, they strategically restrict their learner’s input hypothesis space to include the Identity Map and a few crucial underlying forms that differ only in their violations of the key constraints.

Further research will determine whether any algorithmic approach to finding the necessary ERC vectors can be effectively applied to a range of cases without sacrificing too much of the simplicity afforded by the Identity Map. One potential complication lies in the possibility of multiple dependent hidden rankings—i.e. cases where more than two markedness constraints are crucially ranked with respect to one another (e.g. *A ≫ *B ≫ *C), with no surface evidence demonstrating their necessary configuration. In this circumstance, the learner will observe surface forms where losing candidates include violations of *C, but the search for unfaithful mappings will need to walk back through losers which violated *B, and then still other losers which violated *A, to find the entire hierarchy. Whether this process would prove tractable remains unknown. Regardless, our larger goal has been to demonstrate the difficulties in importing the learning assumptions of parallel Optimality Theory into the derivational world of Harmonic Serialism. It is the properties of each individual grammatical framework that will best determine how its languages can and should be learned.


Anne-Michelle Tessier and Karen Jesney


