2017

Review of _Harmonic Grammar and Harmonic Serialism_

Eric Baković, *University of California, San Diego*

Review of *Harmonic Grammar and Harmonic Serialism*

Eric Baković, UC San Diego

Harmonic Grammar and Harmonic Serialism (henceforth *HGHS*) consists of thirteen chapters addressing matters of empirical, theoretical, and typological concern to Harmonic Grammar (HG) and Harmonic Serialism (HS).[1] As the editors note in their preface, HG and HS are "in some ways relatives of OT [Optimality Theory] in that they incorporate much of its structure (e.g., candidate comparison by markedness and faithfulness constraints)," except that "HS questions the choice of parallel over serial evaluation, while HG questions the assumption that constraints are ranked rather than weighted." Anyone interested in an introduction to and further developments within either HG or HS, or wondering why one would want to pursue one or the other of these theories rather than their better-established relative, will be more than profitably occupied by perusal of the chapters in *HGHS*.

In Part I ("Introductions to Harmonic Grammar and Harmonic Serialism"), the editors have each written a chapter: "Universal grammar with weighted constraints" (Ch. 1, by Pater) and "The theory and practice of Harmonic Serialism" (Ch. 2, by McCarthy). The rest of the chapters are divided into "Analysis and Typology" (Part II, Chs. 3-11) and "Learning" (Part III, Chs. 12 & 13). Part II is the focus of this review, with some discussion of Part I along the way and (with regret) little more than honorable mention of Part III.

More HS than HG

Ten of the thirteen chapters of *HGHS* are concerned almost exclusively with HS as opposed to HG. Two of these chapters (Elfner's "Stress-epenthesis Interactions in Harmonic Serialism," Ch. 9, and Staubs & Pater's "Learning serial constraint-based grammars," Ch. 12) contain some discussion of a hybrid theory, called Serial HG, which assumes both weighted constraints and serial evaluation. Pater's introduction also contains discussion of Serial HG, as well as of Maximum Entropy Grammar (MaxEnt, on which again also see Staubs & Pater's Ch. 12) and of Noisy HG (on which also see Boersma & Pater's "Convergence properties of a Gradual Learning Algorithm for Harmonic Grammar," Ch. 13.)

The bias toward HS is either welcome or unfortunate, depending on one's point of view. On the one hand, work in HS tends to address more 'traditional' phonological problems (locality vs. globality of process interaction; opaque process interacton), and it has the appealing veneer of a just-right compromise between constraint ranking and process ordering.

[1]Thanks to Andrew Nevins, Alan Prince, and Joe Pater for comments on earlier drafts of this review. Any remaining errors of fact, interpretation, and/or terminology are my responsibility.

On the other hand, HG is the more direct gateway from OT to theories such as MaxEnt and Noisy HG that are featured in more of the bleeding-edge work in our subfield on variation and learning — including Staubs & Pater's Ch. 12, where MaxEnt is also used in the context of learning (variation in) HS grammars. Being something of a traditionalist myself, I was generally unperturbed by the skew while reading *HGHS*; taking a step back afterward and considering what I could have learned from a book with this title, however, I felt a little more disquieted. But, as the preface notes, HG and HS are "the two theories that give this book its title", which hardly commits the editors to a balanced review of both.

Defining theories

The discussion to follow will benefit from a (very) brief overview of HG and HS and how each differs from OT. OT defines the mapping from underlying forms (more generally, 'inputs') to surface forms ('grammatical outputs') as follows. A given input is related to a set of output candidates by a generation function called Gen; each candidate in the set represents a set of modifications, possibly empty, of the input.[2] Output candidates are compared against a set of constraints referred to as Con: *markedness* constraints that penalize particular output structures and *faithfulness* constraints that penalize particular disparities between input and output. These constraints are *ranked* in a strict dominance hierarchy such that the grammatical output is the output candidate that, in every pairwise comparison with another output candidate, better satisfies the highest-ranked constraint on which the two output candidates differ. This is the evaluation function in OT, generally referred to as *Eval*.

HG differs from OT primarily in terms of how the evaluation function works. In HG, constraints are *weighted* rather than ranked as they are in OT. Violations of a given constraint by an output candidate (represented as a negative integer) are multiplied by that constraint's specified (and positive) weight; the sum of the weighted constraint violations incurred by a given output candidate is that output candidate's *Harmony*, and the grammatical output is the output candidate with the highest Harmony value.[3]

HS differs from OT primarily in terms of how the candidate generation function works. In HS, Gen is limited to making up to one modification to the input rather than the potentially larger set of modifications allowed by OT's Gen. All such minimal modifications are compared against the set of ranked constraints, exactly as in OT's *Eval*, but the winner of this comparison is then resubmitted as a new input to Gen, which generates another round of minimal modifications that are again compared — and so on until the winner is the output candidate with no modification at all. This final 'step' is the point of *convergence*, and the winning output candidate at this step is the grammatical output.

A final, terminological note. Both HG and HS have the word 'Harmonic' in their name, perhaps giving the impression that these are more related to each other than either is to OT. Such an impression would be a false one, however. The term *Harmony* refers specifically

[2]The cardinality of this candidate set is in principle infinite, but the real requirement imposed by the theory is that it be well-defined. See Prince (2016) for relevant discussion.

[3]On the origin of the word 'Harmony' in this context, see the exchange between Paul Smolensky and Mark Johnson summarized by Joe Pater at `https://blogs.umass.edu/comphon/2016/01/14/whats-harmony/`.

to the sum of weighted violations of all constraints by a given output candidate; that is, to the evaluation function as it is defined in HG. Harmony *per se* is thus nowhere to be found in either OT or HS, where *no number of violations* of a lower-ranked constraint is sufficient to outweigh *a single violation* of a higher-ranked constraint due to the strictness of strict domination property of OT/HS constraint ranking. So while the term is found in the name 'Harmonic Serialism' and in phrases like 'harmonic bounding' and 'harmonic completeness' in OT/HS work, it stands at some distance from its original meaning in these contexts.

Constraints

Alongside the usual questions of interest to phonologists (*what are the phonological representations of inputs and outputs? what are the right analyses of particular phonological patterns?* etc.), a key question to be addressed in OT, HG, and HS is: *what are the constraints?* Answers to this question are typically informed by evidence from various sources: analytical concerns, formal commitments, functional convictions, and typological considerations.

In this context, a *typology* is a set of grammars predicted by a given hypothesis about the content of Con.[4] As Pater discusses in Ch. 1, the difference between the evaluation functions in OT and HG means that the typology predicted in OT will in general be a subset of the typology predicted in HG, given the same set of constraints.[5] Assuming that other sources of evidence (analytical, formal, and functional) are not relevant, then if the 'right' (= sufficiently expressive and appropriately restrictive) typology is the smaller one predicted in OT, then a different set of constraints will be necessary to generate that same typology in HG; conversely, if the 'right' typology is the larger one predicted in HG, then a different set of constraints will be necessary to generate that same typology in OT.

Jesney's "Positional constraints in Optimality Theory and Harmonic Grammar," Ch. 6, offers a particularly convincing demonstration of this point. As Jesney shows, accounting for complex positional licensing patterns requires two types of constraints in OT: positional markedness for conjunctive licensing (given two licensing contexts A and B, a feature F is licensed only if it appears in *both* A *and* B) and positional faithfulness for disjunctive licensing (F is allowed only if it appears in *either* A *or* B). In HG, both conjunctive and disjunctive licensing patterns can be accounted for with just one or the other type of positional constraints, given the right weighting conditions. Adopting HG thus accounts for both types of positional licensing pattern while reducing the set of constraint types, and thus (more significantly) avoids pathological predictions particular to positional faithfulness.

The difference between parallel and serial evaluation in OT and HS, respectively, also

[4] A hypothesis about the specifics of Gen is also necessary, as recent work in modern rigorous OT emphasizes (see e.g. Prince 2016, Alber et al. 2016, Merchant & Prince to appear, and references cited therein). This point applies in a specific way to HS, as discussed in the following section.

[5] The exception to this general statement, as Pater notes in fn. 4 on p. 21, concerns the strictness of strict domination property of OT noted above and hypothetically unbounded constraint violations. This exception is generally irrelevant to practical concerns (the "directly observable typology" in Pater's terms), where violations are naturally bounded. See also Prince (2002, 2006, 2007a), Legendre et al. (2006), Pater et al. (2007), Potts et al. (2009), Tesar (2007), and Magri (2013), among others, for relevant discussion.

means that the typology predicted in OT will differ from the typology predicted in HS. McCarthy discusses this point in Ch. 2, focusing mainly on Jesney's (2011) HS solution to the same problems with positional faithfulness that were alluded to just above. This solution relies on the hypothesis that faithfulness constraints in HS assess violations locally, to the single disparities between inputs and outputs incurred at each step of an HS derivation, rather than globally, to the (potentially) larger sets of disparities between underlying and surface forms (as they do in OT, where there is only one relevant 'step').[6]

Kimper's "Positive constraints and finite goodness in Harmonic Serialism," Ch. 7, also illustrates the point. As Kimper shows, HS can include constraints that *reward* output candidates — in addition to, not instead of, constraints that penalize them — without necessarily running afoul of the "infinite goodness" problem that inevitably arises when such positive constraints are allowed to roam free in OT (Prince 2007b). Briefly, the infinite goodness problem is this: given a markedness constraint $\mathbb{M}$ that rewards the presence of some structure X, then under certain ranking conditions X will be inserted *ad infinitum* in order to better satisfy $\mathbb{M}$. This problem is avoided in HS to the extent that insertion of the relevant X requires more than one modification: if any modification on the way to X-insertion does not satisfy $\mathbb{M}$, then it will not win the local comparative evaluation and the path to ultimate X-insertion will be blocked. In this way, Kimper argues that a positive feature-spreading constraint avoids the infinite goodness problem to the extent that (a) only spreading of a feature satisfies the constraint, and (b) insertion of a segment and spreading of a feature to it are distinct modifications.

Modifications

As McCarthy notes in Ch. 2, a specific question arises in HS about the nature of Gen: *what counts as one modification?* McCarthy helpfully illustrates a method for answering this question within the context of a particular analysis, and links the result of that illustration to a theoretically-grounded hypothesis. The illustration involves syncope and resyllabification in Cairene Arabic, and the particular version of the question is whether these occur separately (because each counts as a modification) or in tandem (because only one counts as a modification). McCarthy provides evidence internal to Cairene Arabic that syncope and resyllabification must count as one modification in tandem, and points to the consistency of this result with McCarthy's (2007) proposal that unfaithful modifications (such as syncope, which violates Max-V) count as modifications while faithful modifications (such as resyllabification, which violates no faithfulness constraint) do not.

Several other chapters in *HGHS* confront the issue of whether — and if so, in what way or to what extent — (re)syllabification counts as a modification. These chapters generally acknowledge and cite each other's sometimes competing claims (and those made in other HS works not in *HGHS*), but the only explicit acknowledgement that some common ground must be reached — and a proposal for how to get there — is made in a footnote by Pater

[6]McCarthy claims that this hypothesis "also limits the usefulness of local conjunction of faithfulness constraints" and refers the reader to "Section 2.8 for further discussion", but unfortunately there is no further discussion of this claim to be found there or elsewhere in the chapter.

in Ch. 1 (fn. 7, p. 35): "serial initial syllabification, and parallel resyllabification."[7]

The more general what-counts-as-one-modification theme emerges with respect to processes other than (re)syllabification in other chapters. Moore-Cantwell's "Contexts for epenthesis in Harmonic Serialism," Ch. 8, outlines a specific theoretical hypothesis, that "epenthesis respects EXHAUSTIVITY" — more transparently, that an inserted constituent must be dominated by a constituent at the next level up the prosodic hierarchy. EXHAUSTIVITY was proposed as a violable constraint in Selkirk (1995) to allow for e.g. weak layering (Ito & Mester 2003); Moore-Cantwell recruits it as part of a hard restriction on GEN, and demonstrates how it constrains HS derivations in such a way as to allow epenthesis to resolve segmental and syllabic markedness but not to resolve metrical markedness.[8]

This key question aside, the assumed gradualness of HS derivations vs. the relative gestalt of OT mappings means that the typology predicted in HS is both more restrictive (because it curtails excessively global interactions) and more expressive (because it allows certain opaque interactions) than the typology predicted in OT. Arguments for the benefits of both of these differences are highlighted in several chapters of *HGHS*.

Right out of the introductory gate, McCarthy, Pater, & Pruitt's "Cross-level interactions in Harmonic Serialism," Ch. 3, defends HS's curtailment of excessively global interactions by demonstrating that both top-down and bottom-up interactions between syllabification and foot structure, which formed the basis of early arguments for OT's parallel architecture and "against serialism," are still accounted for with HS's serial architecture because those early arguments were "arguments against specific serial theories that have very little to do with [HS]." In other words, these authors argue that interactions in HS are 'global enough' to account for a range of observed effects that were heretofore thought to require OT's parallelism. Elsman's "Parallelism vs. serialism, or constraints vs. rules? Tongan stress and syllabification revisted," Ch. 4, makes a similar point with a focus on a single case. And while these two chapters focus on the *sufficiency* of HS's globality, the focus of Staubs's "Serial restrictions on feature/stress interactions," Ch. 5, is on its *limits*. More specifically, Staubs shows that a set of pathological interactions between features and stress predicted by OT's more expressive globality, identified by de Lacy (2002), are avoided in HS.[9]

Near the other end of Part II lie two cases for HS's allowance of certain opaque interactions. Elfner's "Stress-epenthesis interactions in Harmonic Serialism," Ch. 9, makes the case for stress assignment and epenthesis being separate, serially-ordered operations in HS to account for the behavior and extent of well-known examples in which epenthetic vowels are 'invisible' to stress. This account is helpfully contrasted with Alderete's (1999) HEAD-DEP account, which Elfner argues to be both too expressive and too restrictive. Torres-Tamarit's "Compensatory and opaque vowel lengthening in Harmonic Serialism," Ch. 10, makes the

[7]As Pater also notes, this is the approach taken in Torres-Tamarit's "Compensatory and opaque vowel lengthening in Harmonic Serialism," Ch. 10, though it is not explicitly offered there as a path to reconciliation.

[8]Moore-Cantwell argues that augmentation for the sake of word minimality is a special case, requiring abandonment of the theory that minimal word conditions derive from those on minimal feet (Garrett 1999).

[9]For some recent push-back on the question of curtailing OT's globality, see now Adler (2016), Zymet (submitted), and Adler & Zymet (in preparation), the latter citing an earlier observation in class notes by Albright & Flemming (2013). These works highlight apparent cases of "irreducible parallelism" (McCarthy 2013), in which an excessively gradual derivation leads to (at best) missed generalizations.

case for vowel lengthening and consonant deletion/devoicing being separate, serially-ordered operations in HS to account for compensatory and other opaque vowel lengthening patterns.

Rounding out Part II is Wolf's "Cyclicity and non-cyclicity in Maltese: Local ordering of phonology and morphology in OT-CC," Ch. 11. This chapter is technically concerned with McCarthy's (2007) OT-CC ("OT with candidate chains"), a distinct theory with HS under the hood and further elaborated to accommodate opaque process interactions that are beyond the reach of OT, HG, and HS. In OT-CC, an HS-like system constructs from a given input a finite set of *candidate chains*: ordered sets of representations that consist of the input, an output, and harmonically-improving, one-modification-at-a-time representations in between. These candidate chains are compared against a constraint hierarchy that includes so-called PREC constraints, which essentially require particular precedence relations between modifications within the candidate chains. In McCarthy (2007) these PREC constraints are largely limited to typical phonological operations (e.g. insertion, deletion, and feature changes); Wolf extends these to include morphological spell-out operations to address an apparent paradox in the cyclic interaction of stress and syncope in Maltese.

Software check

One of the extraordinary benefits of working in formal linguistics these days, and in OT, HG, and HS in particular, is the availability of software tools to aid in the analysis of individual language patterns, in the analysis of typology, and in the verification of both. Several authors mention their use of OT-Help (Staubs et al. 2010), a tool developed at UMass Amherst (where many if not all of the authors in *HGHS* have been affiliated) specifically to do (comparative) analytical work in parallel OT, HG, HS, and Serial HG.[10]

At many points in my reading of *HGHS*, I wished for easy access to the files created by these authors so that I could verify specific analytical claims myself, or so that I could quickly follow up on an idea generated by consideration of those claims. Of course, one could correspond with the individual authors to see if they would be willing to share their files, but a centralized, accessible repository of online supplementary materials to accompany the work would be ideal. I'm pleased to add that the editors of *HGHS* agree with this suggestion, and now plan to create such a repository; a link to it will eventually be made available at `http://blogs.umass.edu/pater/software/`.

Concluding remarks

The editors point out in their preface that both HG and HS technically antedate OT: "[HG] is the name of the weighted constraint predecessor of OT first characterized in Legendre et al. (1990), while [HS] was used in Prince and Smolensky's early analysis of Berber syllabification that became the first substantive chapter of [Prince & Smolensky (1993)]." The use of OT

[10] I take this opportunity to also recommend OTWorkplace (Prince et al. 2016), a tool developed at Rutgers University specifically to do analytical work in "modern rigorous Optimality Theory, including Harmonic Serialism." For an example of OTWorkplace's usefulness in HS work, see Merchant (2015).

as a yardstick is thus not due to its prior chronological claim to our attention, but rather due to its relative establishment (or entrenchment, depending on one's point of view). For better or for worse, the benchmark status of OT means that much of the content in *HGHS* is taken up by (sometimes defensive-seeming) comparison: OT predicts this good thing (e.g. limited global interactions) and HG/HS can, too; OT predicts this bad thing (e.g. excessive global interactions) and HG/HS can't; OT doesn't predict this good thing (e.g. some opaque interactions) and HG/HS does. While it's hard to see how this could have been avoided, it leaves less room for the reader of *HGHS* to assess HG and HS simply on their own terms.

An example that particularly struck me can be found in McCarthy, Pater, & Pruitt's Ch. 3. The authors compare OT and HS analyses of the same set of facts (epenthesis, shortening, and apocope in Yawelmani), requiring two overlapping constraints in the HS analysis (*Complex-Coda and *Superheavy) but only the latter in the OT analysis. The distinction is argued to be independently necessary in both OT and HS by appeal to another set of facts in a different language (syncope in Sudanese Arabic). This argument left me wondering when the postulation of a constraint in an HS analysis will be defensible solely on HS's own terms, without asking whether it is also necessary in OT.

And yet, the overall impression I am left with is distinctly positive. The chapters of *HGHS* are required reading for anyone interested in HG, HS, OT, and related theories, from the minds of many who have been working most intently on them.

References

Adler, Jeff. 2016. *Parallelism and conspiracy*. UC Santa Cruz Master's thesis.

Adler, Jeff & Jesse Zymet. in preparation. Irreducible parallelism in process interactions. Ms., UC Santa Cruz and UCLA.

Alber, Birgit, Natalie DelBusso & Alan Prince. 2016. From intensional properties to universal support. *Language* 92(2). e88–e116. ROA-1235, Rutgers Optimality Archive, http://roa.rutgers.edu.

Albright, Adam & Edward Flemming. 2013. A note on parallel and serial derivations. Class notes, MIT.

Alderete, John. 1999. Head dependence in stress-epenthesis interactions. In Ben Hermans & Marc van Oostendorp (eds.), *The Derivational Residue in Optimality Theory*, 29–50. Amsterdam: Benjamins.

Garrett, Edward. 1999. Minimal words aren't minimal feet. In Matthew Gordon (ed.), *UCLA Working Papers in Linguistics* vol. 1, 68–105.

Ito, Junko & Armin Mester. 2003. Weak layering and word binarity. In Takeru Honma, Masao Okazaki, Toshiyuki Tabata & Shin-ichi Tanaka (eds.), *A New Century of Phonology and Phonological Theory: A Festschrift for Professor Shosuke Haraguchi on the Occasion of His Sixtieth Birthday*, 26–65. Tokyo: Kaitakusha. Originally circulated in 1992 as working paper LRC-92-09, Linguistic Research Center, UC Santa Cruz.

Jesney, Karen. 2011. Positional faithfulness, non-locality, and the Harmonic Serialism solution. In Suzi Lima, Kevin Mullin & Brian Smith (eds.), *Proceedings of the 39th North*

East Linguistics Society, 239–440. Amherst, MA: GLSA Publications. ROA-1018, Rutgers Optimality Archive, http://roa.rutgers.edu.

de Lacy, Paul. 2002. *The Formal Expression of Markedness*. Amherst, MA, University of Massachusetts Doctoral dissertation.

Legendre, Géraldine, Yoshiro Miyata & Paul Smolensky. 1990. Harmonic Grammar – A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, 388–395. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Legendre, Géraldine, Antonella Sorace & Paul Smolensky. 2006. The Optimality Theory–Harmonic Grammar connection. In Paul Smolensky & Géraldine Legendre (eds.), *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar* vol. 2, chap. 20, 339–402. Cambridge, MA: MIT Press.

Magri, Giorgio. 2013. HG has no computational advantages over OT: towards a new toolkit for computational OT. *Linguistic Inquiry* 44(4). 569–609.

McCarthy, John. 2013. Irreducible parallelism and desirable serialism. Talk presented at the 2013 Annual Meeting on Phonology.

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

Merchant, Nazarré. 2015. Recursive Join Learning using Candidate Maps. ROA-1232, Rutgers Optimality Archive, http://roa.rutgers.edu.

Merchant, Nazarré & Alan Prince. to appear. *The Mother of All Tableaux*. London: Equinox. ROA-1285, Rutgers Optimality Archive, http://roa.rutgers.edu.

Pater, Joe, Rajesh Bhatt & Christopher Potts. 2007. Linguistic optimization. ROA-924, Rutgers Optimality Archive, http://roa.rutgers.edu.

Potts, Christopher, Joe Pater, Karen Jesney, Rajesh Bhatt & Michael Becker. 2009. Harmonic Grammar with Linear Programming: From Linear Systems to Linguistic Typology. *Phonology* 27. 77–117.

Prince, Alan. 2002. Anything goes. In Takeru Honma, Masao Okazaki, Toshiyuki Tabata & Shin-ichi Tanaka (eds.), *New Century of Phonology and Phonological Theory*, 66–90. Tokyo: Kaitakusha. ROA-536, Rutgers Optimality Archive, http://roa.rutgers.edu.

Prince, Alan. 2006. Harmony at Base Omega: Utility Functions for OT. ROA-798, Rutgers Optimality Archive, http://roa.rutgers.edu.

Prince, Alan. 2007a. Let the decimal system do it for you: a very simple utility function for OT. ROA-943, Rutgers Optimality Archive, http://roa.rutgers.edu.

Prince, Alan. 2007b. The pursuit of theory. In Paul de Lacy (ed.), *The Cambridge Handbook of Phonology*, 33–60. Cambridge: Cambridge University Press.

Prince, Alan. 2016. What is OT? ROA-1271, Rutgers Optimality Archive, http://roa.rutgers.edu.

Prince, Alan & Paul Smolensky. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. Rutgers University Center for Cognitive Science: Technical Report RuCCS-TR-2. ROA-537, Rutgers Optimality Archive, http://roa.rutgers.edu.

Prince, Alan, Bruce Tesar & Nazarré Merchant. 2016. OTWorkplace. Software package, Available at `https://sites.google.com/site/otworkplace/`.

Selkirk, Elisabeth O. 1995. The prosodic structure of function words. In Jill Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.), *Papers in Optimality Theory*, GLSA

Publications.
Staubs, Robert, Michael Becker, Christopher Potts, Patrick Pratt, John J. McCarthy & Joe Pater. 2010. OT-Help 2.0. Software package, `http://people.umass.edu/othelp/`.
Tesar, Bruce. 2007. A comparison of lexicographic and linear numeric optimization using violation difference ratios. ROA-939, Rutgers Optimality Archive, http://roa.rutgers.edu.
Zymet, Jesse. submitted. A case for parallelism: reduplication-repair interaction in Maragoli. Available at `http://www.linguistics.ucla.edu/people/grads/jzymet/maragoli_parallelism.pdf`.