

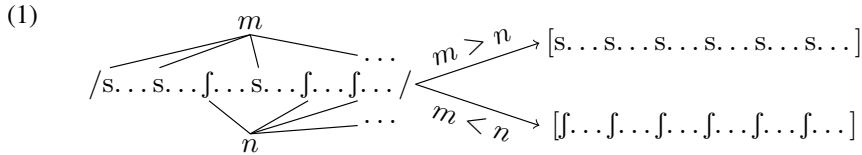
Majority Rule in Harmonic Serialism

Andrew Lamont

University of Massachusetts Amherst

1 Introduction

Majority Rule is an unattested process where agreement is controlled by the largest class in the input (Lombardi, 1999; Baković, 2000). (1) illustrates Majority Rule with a running example of sibilant harmony; here the classes are [+anterior] sibilants, such as /s/, and [−anterior] sibilants, such as /ʃ/. A disharmonic input cannot surface faithfully, and must either be mapped onto an output with only [+anterior] sibilants or one with only [−anterior] sibilants. The choice between these mappings depends only on how many members each class has in the input (represented in (1) by m and n): whichever class is larger will surface. If the input contains more [+anterior] sibilants ($m > n$), then the output contains only [+anterior] sibilants. Only [−anterior] sibilants surface if there are more underlyingly ($m < n$). In the event of a tie ($m = n$), a default value for [anterior] must be chosen to characterize Majority Rule as a total function.



As a function from inputs to outputs, Majority Rule requires more computational expressivity than do attested phonological transformations (Riggle, 2004; Gerdemann & Hulden, 2012; Heinz & Lai, 2013). Determining the correct output for a given input involves counting the members of each class to ascertain which class is larger. Because an input may be arbitrarily large, this requires a computational model with *unbounded* memory. Attested phonological transformations can be modeled using Finite-State Transducers (Johnson, 1972; Kaplan & Kay, 1994), which have *bounded* memory capacity. Any model that is only expressive enough to compute attested phonological transformations therefore cannot produce Majority Rule.

This paper examines how Majority Rule arises in parallel Optimality Theory (pOT) (Prince & Smolensky, 1993/2004) and Harmonic Serialism (HS) (Prince & Smolensky, 1993/2004; McCarthy, 2000). Section 2 reviews how pOT produces Majority Rule. Section 3 presents the conditions under which HS produces Majority Rule. Section 4 generalizes the problem of Majority Rule to a host of other pathologies. Section 5 proposes importing directional constraint evaluation (Eisner, 2000) into HS as a solution, and illustrates the proposal with an analysis of sibilant harmony in Tsuut’ina. Section 6 concludes.

2 Majority Rule in parallel Optimality Theory

Majority Rule is predicted to be the default agreement process in pOT. It was first described with AGREE constraints defined over adjacent segments (Lombardi, 1999; Baković, 2000), but generalizes to any output constraint that penalizes the co-occurrence of two classes. This paper focuses on agreement between non-adjacent segments, and does so for two reasons. First, as is discussed below, the relevant output constraints can be defined locally over tier-adjacent segments, or globally over every pair of adjacent or non-adjacent segments. Globally defined constraints are shown later to be the root of the problem (§4). Second, as §3 demonstrates, HS can only derive Majority Rule when agreement is between non-adjacent segments.

* This work has greatly benefited from discussions with Carolyn Anderson, Gaja Jarosz, John McCarthy, Joe Pater, Rachel Walker, Kristine Yu, participants in the UMass Sound Workshop, attendants of presentations at Swarthmore College, PhoNE 2018, 26mfm, and AMP 2018, and anonymous reviewers for AMP 2018. All remaining errors are of course my own.

The tableau in (2) illustrates Majority Rule with sibilant harmony. The constraint CORR(SIBILANT) requires that sibilants correspond with each other; correspondence is indicated on segments with an index i . CC-IDENT(ANTERIOR) penalizes corresponding segments that disagree in the feature [anterior]; the number of violations assigned by CC-IDENT(ANTERIOR) is intentionally excluded in the tableau and discussed below. Ranked high enough, these two constraints filter out faithful candidates whose sibilants do not correspond (2a) and those whose sibilants correspond, but do not agree in [anterior] (2b). This leaves unfaithful candidates whose sibilants correspond and agree in [anterior] (2c-d). Because both unfaithful candidates satisfy the output constraints, the choice between them falls to the lower-ranked faithfulness constraint IDENT(ANTERIOR). IDENT(ANTERIOR) prefers candidate (2c) with its two unfaithful sibilants to candidate (2d) with its three, and so chooses as optimal the candidate in which the minority class assimilates to the majority class.¹ It is not crucial that IDENT(ANTERIOR) *counts* violations, but it is crucial that it can *distinguish* between two violations and three. Through mark cancellation, EVAL itself does not count violations per se, but it is nonetheless expressive enough to model functions that rely on counting.

(2) Majority Rule in parallel OT: $/j...j...j...s...s/ \rightarrow [j...j...j...j...j]^2$

$/j...j...j...s...s/$	CORR(SIB)	CC-IDENT(ANT)	IDENT(ANT)
a. $j...j...j...s...s$	W 10		L
b. $j_i...j_i...j_i...s_i...s_i$		W	L
→ c. $j_i...j_i...j_i...j_i...j_i$			2
d. $s_i...s_i...s_i...s_i...s_i$			W 3

Majority Rule maximally satisfies the output constraints while minimally violating the faithfulness constraint; it is the globally optimal solution to agreement. The discussion above leaves the evaluation of CC-IDENT(ANTERIOR) vague, because the result does not crucially depend on its specific formulation. All that matters is that it penalizes candidates in which the two classes co-occur. Whether evaluated *globally*, assigning violations to every pair of corresponding segments that disagree in [anterior] (Bennett, 2013, 2015; Hansson, 2001, 2010; Rose & Walker, 2004; Walker, 2000), or evaluated *locally*, only assigning violations to tier-adjacent disagreeing segments (Hansson, 2007, 2014; Pulleyblank, 2002; Walker, 2015), CC-IDENT(ANTERIOR) motivates Majority Rule. In pOT, it does not matter exactly how many violations candidate (2b) incurs, provided it incurs more than candidates (2c-d). Section 3 elaborates on this, as this difference in evaluation is critical for Harmonic Serialism. It is worth noting that under either evaluation type, CC-IDENT is categorical (McCarthy, 2003); while a given segment may define multiple loci of violation, each locus of violation incurs exactly one violation.

As discussed in Section 1, the complexity of Majority Rule derives from comparing the sizes of two arbitrarily large classes. Majority Rule can be made even more complex by introducing overlapping classes. This happens when sibilants are required to agree in a second feature independent of [anterior]. The tableau in (3) below adds the constraints CC-IDENT(VOICE), which penalizes corresponding segments that disagree in the feature [voice], and IDENT(VOICE), which penalizes changing the [voice] feature of a given segment. With the output constraints ranked over the faithfulness constraints, sibilants must agree in both [anterior] and [voice], as is the case in various Berber languages (Hansson, 2010:169-175). As in the previous tableau (2), no active constraints prefer one value of [anterior] over the other or one value of [voice] over the other. The surface specifications of these features are determined independently by minimizing their respective faithfulness violations. The input to (3) contains more [−anterior] sibilants than [+anterior] sibilants as well as more [+voice] sibilants than [−voice] sibilants, so the output contains only [−anterior, +voice] sibilants.³

¹ Technically, this process is *Plurality* Rule. If agreement is enforced across three or more classes, the largest class always controls assimilation regardless of whether it is the *majority* class or not.

² Hansson (2007:403) presents a similar Majority Rule effect with the same constraints.

³ While Majority Rule over a single feature can be modeled using a Pushdown Transducer, an abstract machine more expressive than a Finite-State Transducer, multi-feature Majority Rule requires even more expressivity. For instance, the set of inputs that are mapped onto outputs which only contain the sibilant [ʒ] is not a Context-Free Language and therefore cannot be identified by a Pushdown Transducer. This can be proven by pumping the string $z^n ʒ^n j^m s^n j^n ʒ^n z^n$. I suspect multi-feature Majority Rule requires nothing less than a Linear Bounded Automaton, a type of Turing Machine.

- (3) Multi-feature Majority Rule in parallel OT:
- $/\mathfrak{z}\dots\mathfrak{f}\dots\mathfrak{z}\dots\mathfrak{s}\dots\mathfrak{z}/ \rightarrow [\mathfrak{z}\dots\mathfrak{z}\dots\mathfrak{z}\dots\mathfrak{z}\dots\mathfrak{z}]$

$/\mathfrak{z}\dots\mathfrak{f}\dots\mathfrak{z}\dots\mathfrak{s}\dots\mathfrak{z}/$	CORR(SIB)	CC-ID(ANT)	CC-ID(VC)	ID(ANT)	ID(VC)
a. $\mathfrak{z}\dots\mathfrak{f}\dots\mathfrak{z}\dots\mathfrak{s}\dots\mathfrak{z}$	W 10			L	L
b. $\mathfrak{z}_i\dots\mathfrak{f}_i\dots\mathfrak{z}_i\dots\mathfrak{s}_i\dots\mathfrak{z}_i$		W	W	L	L
c. $\mathfrak{f}_i\dots\mathfrak{f}_i\dots\mathfrak{f}_i\dots\mathfrak{f}_i\dots\mathfrak{f}_i$				2	W 3
→ d. $\mathfrak{z}_i\dots\mathfrak{z}_i\dots\mathfrak{z}_i\dots\mathfrak{z}_i\dots\mathfrak{z}_i$				2	2
e. $\mathfrak{s}_i\dots\mathfrak{s}_i\dots\mathfrak{s}_i\dots\mathfrak{s}_i\dots\mathfrak{s}_i$				W 3	W 3
f. $\mathfrak{z}_i\dots\mathfrak{z}_i\dots\mathfrak{z}_i\dots\mathfrak{z}_i\dots\mathfrak{z}_i$				W 3	2

3 Majority Rule in Harmonic Serialism

In pOT, GEN is unrestricted, and, as a consequence, produces candidates that differ from the input in arbitrarily many ways. For instance, the tableaux in Section 2 (2-3) contain candidates where assimilation has occurred enough times for the output constraints to be satisfied. In pOT, for any given input, the candidate set contains fully agreeing candidates including the Majority Rule candidate. In HS, GEN is restricted to producing only those candidates that differ from the input via the application of exactly one operation. For certain inputs, then, it may be the case that every member of the candidate set violates the output constraints, and the intermediate form chosen will be the candidate that best satisfies them. For example, the candidate set for the input $/\mathfrak{f}\dots\mathfrak{f}\dots\mathfrak{s}\dots\mathfrak{s}/$ will only contain candidates that violate CORR(SIBILANT), such as $\mathfrak{f}\dots\mathfrak{f}\dots\mathfrak{s}\dots\mathfrak{s}$ or CC-IDENT(ANTERIOR), such as $\mathfrak{f}_i\dots\mathfrak{f}_i\dots\mathfrak{f}_i\dots\mathfrak{s}_i$. The candidates that satisfy both constraints, such as $\mathfrak{f}_i\dots\mathfrak{f}_i\dots\mathfrak{f}_i\dots\mathfrak{f}_i$, are two operations removed from the input.

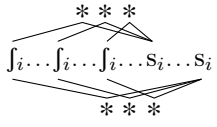
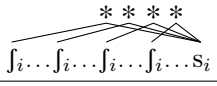
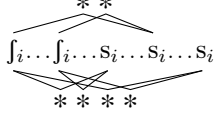
Majority Rule requires assimilating arbitrarily many segments, which takes as many steps in HS as there are segments to assimilate. In pOT, Majority Rule results from comparing the total number of assimilations candidates have undergone. Directly importing this strategy into HS amounts to comparing the lengths of all potential derivations for some input, which is impossible, because HS cannot lookahead to future derivational steps. In terms of their typological predictions, the only differences between pOT and HS are global optima (McCarthy, 2008). HS can only reach some of global optima pOT produces, as it often gets stuck in local optima. Given its nature, we might expect that Majority Rule is one of the global optima unique to pOT. However, as this section demonstrates, this is not the case. Global evaluation of CC-IDENT provides enough information for EVAL to choose the Majority Rule candidate at every step of an HS derivation (§3.1). This issue is not easily avoided, as local evaluation of CC-IDENT cannot motivate iterative harmony at all (§3.2).

3.1 Globally evaluated CC-IDENT produces Majority Rule in HS The tableaux in (4-5) below illustrate Majority Rule in HS with the same input and constraint set as in the tableau in (2) above. In these tableaux, CC-IDENT(ANTERIOR) is evaluated globally, assigning violations to every pair of corresponding segments that disagree in [anterior]; loci of violation of CC-IDENT(ANTERIOR) are explicitly marked with asterisks *. Candidates whose sibilants do not correspond are ruled out by CORR(SIBILANT). In the first step (4), this leaves only candidates that violate CC-IDENT(ANTERIOR) to various degrees. Crucially, global evaluation of CC-IDENT assigns different numbers of violations depending on how many members of each class a candidate contains. Both the faithful candidate (4b) and the candidate where one $/\mathfrak{f}/$ assimilated (4d) violate CC-IDENT(ANTERIOR) six times. The Majority Rule candidate (4c), where one $/\mathfrak{s}/$ assimilated, violates CC-IDENT(ANTERIOR) only four times, and is chosen as optimal.⁴ For space, a second Majority Rule candidate where the second $/\mathfrak{s}/$ assimilated $\mathfrak{f}_i\dots\mathfrak{f}_i\dots\mathfrak{f}_i\dots\mathfrak{s}_i\dots\mathfrak{f}_i$ is excluded from the tableau, despite its tying with (4c). At any given step, Majority Rule is optimal, and all Majority Rule candidates tie (see the Appendix for the general case). These ties are *convergent* (Pruitt, 2009); choosing one candidate over the other does not affect the ultimate output. The entire minority class assimilates gradually over the course of the derivation and the order in which specific segments assimilate is inconsequential.

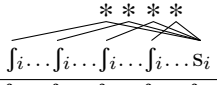
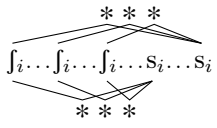
The optimal candidate serves as input to the second step (5) of the derivation. At this point, the minority class has exactly one member, and assimilating that remaining segment satisfies CC-IDENT(ANTERIOR) (5c). That candidate serves as input to a third step (not shown), where the derivation converges.

⁴ Correspondence between multiple segments similarly produces Majority Rule effects in paradigm uniformity (McCarthy, 2004).

(4) Majority Rule in HS, Step 1: /f...f...f...s...s/ → f...f...f...f...s

/f...f...f...s...s/	CORR(SIB)	CC-IDENT(ANT)[GLOBAL]	IDENT(ANT)
a. f...f...f...s...s	W 10		L
b. 		W 6	L
→ c. 		4	1
d. 		W 6	1

(5) Majority Rule in HS, Step 2: f...f...f...f...s → [f...f...f...f...f]

f...f...f...f...s	CORR(SIB)	CC-IDENT(ANT)[GLOBAL]	IDENT(ANT)
a. f...f...f...f...s	W 10		L
b. 		W 4	L
→ c. f...f...f...f...f			1
d. 		W 6	1

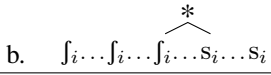
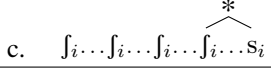
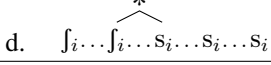
Multi-feature Majority Rule follows a similar derivational path. Assuming changing one [anterior] feature is an unfaithful operation and changing one [voice] feature is another, the derivation will first model Majority Rule over one feature and then the other, with the order depending on the rankings of CC-IDENT(ANTERIOR) and CC-IDENT(VOICE) relative to one another. For instance, if CC-IDENT(ANTERIOR) is ranked higher, the input from the tableau in (3) /ʒ...ʒ...ʒ...s...z/ will first be mapped gradually onto ʒ...ʒ...ʒ...ʒ...ʒ, where the two [+anterior] sibilants assimilate to the three [−anterior] sibilants. This intermediate form will then be mapped gradually onto the output [ʒ...ʒ...ʒ...ʒ...ʒ], where the two [−voice] sibilants assimilate to the three [+voice] sibilants.⁵

In pOT, Majority Rule results from comparing how many faithfulness violations equally well-formed candidates incur. Because GEN is unrestricted in pOT, this involves comparing arbitrarily many violations. In HS, faithfulness constraints are defined directly in terms of operations (McCarthy, 2016:§2.5), and, because GEN is restricted, cannot assign more than one violation to any given candidate. Majority Rule occurs in HS because it maximally satisfies CC-IDENT, but only when it is evaluated globally.

3.2 Locally evaluated CC-IDENT cannot produce Majority Rule in HS Whereas global evaluation of CC-IDENT always produces Majority Rule, local evaluation yields very different outcomes. Locally defined constraints cannot always motivate assimilation (Wilson, 2003; Pater et al., 2007), and, as the tableau in (6) below illustrates, this is true for CC-IDENT. As in previous tableaux, the high-ranked CORR(SIBILANT) rules out candidates whose sibilants do not correspond to each other (6a), leaving only candidates that violate CC-IDENT(ANTERIOR) (6b-d). Evaluated locally, CC-IDENT only assigns violations to *adjacent* sibilants, where adjacency is defined over the tier consisting of corresponding segments, not the timing tier. Assimilation in either direction fails to improve on CC-IDENT, and the unfaithful candidates are harmonically bounded (6c-d). The derivation converges on this step without harmonizing the input.

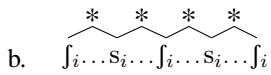
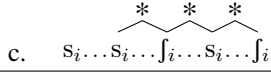
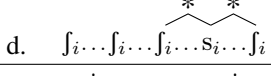
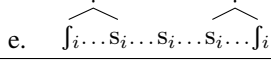
⁵ Breaking multi-feature Majority Rule down into multiple steps does not affect its formal complexity. Majority Rule over [anterior] can be modeled by one Pushdown Transducer, and Majority Rule over [voice] can be modeled by a second, with the entire function defined as their composition. Pushdown Transducers are not closed under composition, so the resulting function is not necessarily one that can itself be modeled with a Pushdown Transducer.

- (6) No iterative harmony in HS with local evaluation:
- $/f...f...f...s...s/ \rightarrow [f...f...f...s...s]$

$/f...f...f...s...s/$	CORR(SIB)	CC-IDENT(ANT)[LOCAL]	IDENT(ANT)
a. $f...f...f...s...s$	W 10		
\rightarrow b. $f...f...f...s_i...s_i$ 		1	
c. $f...f...f...f...s_i...s_i$ 		1	W 1
d. $f...f...s_i...s_i...s_i...s_i$ 		1	W 1

When local CC-IDENT does motivate harmony, it is not powerful enough to guarantee Majority Rule. To improve on CC-IDENT, assimilation must *remove* loci of violation without *replacing* them. Unfaithful segments must agree with their neighbors on the tier. This happens when the unfaithful segment is at the edge of a tier, e.g., $/s...f...f/ \rightarrow [f...f...f]$ or $/f...f...s/ \rightarrow [f...f...f]$, or sandwiched between agreeing segments, e.g., $/f...s...f/ \rightarrow [f...f...f]$. As the tableau in (7) illustrates, local contexts cannot identify the larger class in the input, or even prefer one class over another. The input consists of alternating sibilants which, when brought into correspondence, define four loci of violation of CC-IDENT(ANTERIOR) (7b). Targeting either edge sibilant only removes one locus of violation (7c); targeting an interior sibilant removes two (7d-e). Because all interior sibilants offer equal improvement on CC-IDENT(ANTERIOR), this choice produces a *disjunctive* tie (Pruitt, 2009): whichever class the structural change belongs to in this step will control harmony throughout the derivation. If candidate (7d) is chosen, the derivation will eventually converge on the Majority Rule output $[f...f...f...f...f]$. Choosing candidate (7e) instead leads to the output $[s...s...s...s...s]$, where the derivation has assimilated a superfluous sibilant. While this input can only be mapped onto fully harmonized outputs, longer strings of alternating sibilants can get stuck in local optima with partial harmony.

- (7) Disjunctive tie in HS with local evaluation:
- $/f...s...f...s...f/ \rightarrow [f...f...f...f...f] \sim [s...s...s...s...s]$

$/f...s...f...s...f/$	CORR(SIB)	CC-IDENT(ANT)[LOCAL]	IDENT(ANT)
a. $f...s...f...s...f$	W 10		L
b. $f...s_i...f...s_i...f...s_i...f$ 		W 4	L
c. $s_i...s_i...f...s_i...f...s_i...f$ 		W 3	1
\rightarrow d. $f...f...f...s_i...f...s_i...f$ 		2	1
\rightarrow e. $f...s_i...s_i...s_i...s_i...f$ 		2	1

Local evaluation of CC-IDENT reflects the fundamental difference between pOT and HS: while pOT always selects the global optimum, HS can get stuck and converge on a local optimum. When CC-IDENT is evaluated locally, pOT produces Majority Rule, and HS produces a range of possible outputs.

4 The scope of the problem

Global and local evaluation of CC-IDENT pose complementary problems for HS. Global evaluation motivates harmony, but it overgenerates and produces Majority Rule. Local evaluation undergenerates, and is often unable to motivate harmony. One approach would be to abandon CC-IDENT as the constraint motivating harmony in favor of an alternative such as targeted constraints (Wilson, 2003), the SHARE family (McCarthy, 2010), or positive constraints (Kemper, 2016). While these alternatives are all viable solutions for harmony, they do not necessarily address the much more general problem of global evaluation.

The crucial difference between local and global evaluation is the number of violations candidates can incur. Under local evaluation, loci of violation are tier-adjacent segments, and the maximum number of loci grows *linearly* in the length of a candidate: a candidate with n segments has at most $n - 1$ loci. Under global

evaluation, *any* pair of segments, adjacent and non-adjacent, is a possible locus, and the maximum number grows *quadratically* in the length of a candidate: a candidate with n segments has at most $\binom{n}{2} = \frac{n^2-n}{2}$ loci.



It is unsurprising that global evaluation produces functions that are not regular, because constraints that assign quadratically many violations are themselves not regular functions (Eisner, 1997a; Bíró, 2003). Such quadratic constraints produce many non-regular functions in pOT and HS. For example, the Midpoint Pathology, where a floating element such as a high tone docks in the center of an input, e.g., $/\sigma\sigma\sigma\sigma\sigma-H/ \rightarrow [\sigma\sigma\acute{\sigma}\sigma\sigma]$, relies on a quadratic constraint (Eisner, 1997b), which assigns one violation for every pair of syllables to the left of the high tone and one for every pair of syllables to the right of the high tone (see the Appendix for details). Similarly, modeling alphabetical sorting in HS requires a constraint that assigns violations to every pair of out-of-order segments (Lamont, 2018).

This property is so reliable that it can easily be extended to novel pathologies. For example, consider the output constraint defined in (8), a variant of the Generalized OCP family without adjacency (Suzuki, 1998). This constraint assigns violations to every pair of segments in a candidate that are identical.



(8) *X...X: Assign one violation for every pair of segments $\alpha\dots\beta$ where $\alpha = \beta$.

The tableaux in (9-10) illustrate the effect of *X...X, a pattern I'm dubbing *congruent dissimilation*. For any input, the derivation eventually converges on an output with equally sized classes (see the Appendix for details). In the derivation below, an input with 4 [+anterior] sibilants is gradually mapped onto an output with 2 [+anterior] sibilants and 2 [-anterior] sibilants; the order they occur in is irrelevant.⁶ Congruent dissimilation is the complement of Majority Rule: it derives outputs with multiple, equally sized classes.

(9) Congruent dissimilation in HS, Step 1

/ssss/	*X...X	ID(ANT)
a. 	W 6	L
→ b. 	3	1

(10) Congruent dissimilation in HS, Step 2

/ʃsss	*X...X	ID(ANT)
a. 	W 3	L
→ b. 	2	1

Quadratic constraints are problematic, but banning them from CON hamstrings HS.⁷ By hypothesis, the only structures output constraints can refer to are strings and subsequences (Heinz, 2018). Strings are objects composed of contiguous elements defined on some tier. Local evaluation of CC-IDENT penalizes strings with the correspondence class defining the tier. Subsequences are objects composed of ordered elements which need not be contiguous (Heinz, 2007, 2010; Rogers et al., 2010). Global evaluation of CC-IDENT penalizes subsequences. Banning quadratic constraints entails banning subsequences as possible loci of violation. This leaves only constraints that refer to strings, which, as Section 3.2 demonstrated, significantly impairs the empirical adequacy of HS. Adopting an alternative analysis for harmony may problematically enrich the set of structures CON has access to, and ultimately requires unrelated patches to avoid the Midpoint Pathology, alphabetical sorting, congruent dissimilation, and other as yet undescribed pathologies. We need a general solution that addresses these problems without compromising phonological sensibility or empirical adequacy.

5 Directional constraint evaluation in Harmonic Serialism

As discussed in Section 3.2, locally-defined constraints cannot motivate harmony in HS. This is unfortunate especially because they effectively represent known properties of harmony: spreading, for the most part, only depends on adjacent segments (Wilson, 2003, 2006; Walker, 2010), exceptions to harmony are local (Finley, 2010), and blocking effects in harmony require a notion of tier-adjacency (McMullin & Hansson, 2015; McMullin, 2016). These properties all point to locally-defined constraints driving harmony.

In a rule-based framework, harmony can be represented as a self-feeding iterative rule with a locally-defined structural description such as $/s/ \rightarrow [ʃ] / _ X ʃ$, where X does not contain any sibilants. The structural change at one step defines the environment for the following step, producing iterative harmony, e.g.,

⁶ With two classes, this defines a Context-Free stringset; three or more classes derive a Context-Sensitive stringset.

⁷ Such a move also has consequences for the definition of CORR constraints, which are evaluated globally.

By virtue of only modeling regular functions, directional constraint evaluation not only eliminates Majority Rule, but all the other complex processes described above. Intuitively, Majority Rule in HS involves identifying the minority class on every step and assimilating one of its members to the majority class. As discussed in Section 3.1, targeting any member of the minority class offers equal improvement on the output constraints, and the grammar is free to choose among them. Directional evaluation removes this freedom by specifying exactly which locus of violation offers the most improvement. As illustrated in the tableau above (11), targeting the rightmost or leftmost locus is strictly better than targeting any other locus, and this positional aspect is independent of the size of the classes.

There is empirical support for directional evaluation as well, such as sibilant harmony in Tsuut'ina (formerly referred to as Sarcee) (Cook, 1979, 1984). In Tsuut'ina, [+anterior] sibilants assimilate to following [−anterior] sibilants. This is an instance of dominant-directional harmony: only [−anterior] spreads, and it only spreads regressively. The data below illustrate that [−anterior] spreading is regressive (13) and iterative (14), and that [−anterior] does not spread progressively (15-16). [+anterior] sibilants are free to surface after [−anterior] sibilants, but not before them.

- (13) a. /sí-tsàɣà/ → [sìsáyà] ‘my hair’ (Cook, 1979:26)
 b. /sí-tʃógò/ → [ʃitʃógò] ‘my flank’ (Cook, 1979:26)
- (14) a. /tó-gu-si-s-k'az/ → [tógùsk'áz] ‘he painted it (e.g., a floor) red’ (Cook, 1984:204)
 b. /tó-si-s-tʃídʒ/ → [tófítʃídʒ] ‘I painted it (e.g., a stick)’ (Cook, 1984:204)
 c. /ts'aasi tʃak'a-i/ → [tʃ'ááʃí tʃák'ā-ā] ‘Crow-Ribs (personal name)’ (Cook, 1984:102)
- (15) a. /tʃiz/ → [tʃíz] ‘duck’ (Cook, 1979:26)
 b. /sí-tʃiz-aʔ/ → [ʃitʃidzàʔ] ‘my duck’ (Cook, 1979:26)
 c. /tʃiz k'ùs/ → [tʃiz k'ùs] ‘duck neck’ (Cook, 1984:68)
 d. /tʃiz mi-si-t'ùd-í/ → [tʃiz mìst'ùdí] ‘duck pipe’ (Cook, 1984:209)
- (16) a. [áftʃ'á] ‘split’ (Cook, 1984:49)
 b. [áftʃ'á-sùz] ‘I will split it’ (Cook, 1984:48)
 c. [áftʃ'á-ázùz] ‘it will split’ (Cook, 1984:48)

The analysis sketched below is almost identical to the toy harmony example given above (11). $\text{AGREE}(\text{SIBILANT}, \text{ANTERIOR})_{R \rightarrow L}$ evaluates right-to-left and dominates $\text{IDENT}(\text{ANTERIOR})$. Two additional faithfulness constraints complete the analysis. To prevent [−anterior] segments from assimilating, $\text{IDENT}(-\text{ANTERIOR})$ dominates the AGREE constraint. As the tableau in (17) makes clear, a positional faithfulness constraint that protects the rightmost sibilant is also necessary (Jurgec, 2011; Bennett, 2015): $\text{IDENTTAIL}(\text{SIBILANT}, \text{ANTERIOR})$. Like the AGREE constraint, this faithfulness constraint projects a sibilant tier in order to define the rightmost sibilant locally. The derivation in (17) converges on the fully faithful candidate (17a) because assimilation either fatally targets a [−anterior] sibilant (17b) or the rightmost sibilant (17c). Because both unfaithful candidates satisfy the output constraint, they must be ruled out by other constraints.

(17) Sibilant harmony in Tsuut'ina (1)

/tʃiz/	ID(−ANT)	IDTAIL(SIB, ANT)	AGREE(SIB, ANT) _{R→L}	ID(ANT)
→ a. tʃiz			tʃ ₁ z ₂	
b. tsiz	W 1		L	W 1
c. tʃiʒ		W 1	L	W 1

As the tableau above illustrates, IDENTTAIL prevents progressive assimilation from occurring when only one target is present. With longer spans of potential targets, the directionality of AGREE prevents progressive assimilation. This is illustrated in the tableau in (18) below. This derivation also converges on the fully faithful candidate (18a), even though it violates AGREE . Regressive assimilation fatally targets a [−anterior] sibilant (18b), and progressive assimilation pushes the locus further to the right, incurring a worse violation of AGREE (18c).

(18) Sibilant harmony in Tsuut'ina (2)

/tʃiz mi-si-t'ùd-í/	ID(-ANT)	IDTAIL(SIB, ANT)	AGREE(SIB, ANT) _{R→L}	ID(ANT)
→ a. tʃiz mist'ùdí			tʃ ₁ z ₂	
b. tsiz mist'ùdí	W 1		L	W 1
c. tʃiz mist'ùdí			W	3 ₂ S ₃ W 1

6 Conclusion and future directions

This paper aimed to demonstrate a very general problem for Harmonic Serialism: global output constraints defined over subsequences are too powerful under optimization, while local output constraints defined over strings are not powerful enough given the restrictions on GEN. The focus was on harmony, but the problem is much more general. It is certainly not a new result that optimization can generate non-regular functions (Frank & Satta, 1998; Smolensky & Legendre, 2006), and while alternative analyses for some phenomena have successfully addressed specific pathologies, such as limiting alignment to avoid the Midpoint Pathology (Hyde, 2012, 2016), the problem demands a very general solution.

Importing directional constraint evaluation into Harmonic Serialism was proposed as a general solution, and it appears to address a number of unrelated problems. Directional evaluation maintains local generalizations without impairing the grammar's ability to model iterative processes. Whether it turns out to be empirically adequate is a question demanding further investigation. It may offer a solution to the problem of divergent ties in Harmonic Serialism (Pruitt, 2009), and may indeed limit the grammar to only modeling regular functions (Lamont, 2019). It is important moving forward to carefully characterize its properties and formalize its typological predictions in order to compare it against standard analyses and frameworks.

Appendix

This appendix derives the general cases of the pathologies discussed in the paper. Output constraints are evaluated traditionally, dispreferring candidates that incur more violations relative to other candidates.

$$(19) \quad (\# \text{ violations by Cand}_i < \# \text{ violations by Cand}_j) \leftrightarrow (\text{Cand}_i > \text{Cand}_j)$$

Majority Rule in HS with two classes For simplicity, I only consider strings made of two symbols: a and b . Consider the output constraint AGREE-GLOBAL, which assigns one violation to every disagreeing pair of segments in a candidate: $\{a \dots b, b \dots a\}$. For each a in a candidate, AGREE-GLOBAL assigns one violation for each b in the candidate. The total number of violations therefore equals the product of the number of a 's and the number of b 's. It is always optimal to assimilate a member of the smaller class to the larger class.

Consider an input made of a 's and b 's with $n > 2$ total segments. Define A as the total number of a 's in the input and B as the total number of b 's in the input: $A + B = n$. Let there be more a 's than b 's: $A > B$. Let GEN have only one operation which changes an a to a b or a b to an a . There are three candidate types at each step: the fully faithful candidate $\text{Cand}_{\text{faithful}}$, the candidate(s) where an a became a b $\text{Cand}_{a \rightarrow b}$, and the Majority Rule candidate(s) where a b became an a $\text{Cand}_{b \rightarrow a}$.

$\text{Cand}_{\text{faithful}}$ has the same number of a 's and b 's as the input, and incurs $A \times B = AB$ violations of AGREE-GLOBAL. $\text{Cand}_{a \rightarrow b}$ has one more a , and one fewer b , and incurs $(A - 1) \times (B + 1) = AB + A - B - 1$ violations. $\text{Cand}_{b \rightarrow a}$ has one more b , and one fewer a , and incurs $(A + 1) \times (B - 1) = AB - A + B - 1$ violations. AGREE-GLOBAL prefers $\text{Cand}_{b \rightarrow a}$ because it incurs the fewest violations:

$$\text{Cand}_{b \rightarrow a} > \text{Cand}_{\text{faithful}} \succeq \text{Cand}_{a \rightarrow b}$$

$$\# \text{ violations by Cand}_{b \rightarrow a} < \# \text{ violations by Cand}_{\text{faithful}} \leq \# \text{ violations by Cand}_{a \rightarrow b}$$

$$AB - A + B - 1 < AB \leq AB + A - B - 1$$

$$B - A < 1 \leq A - B$$

This confirms that targeting the minority class always improves the most on AGREE-GLOBAL. Any input with more a 's than b 's will be mapped onto a form with one fewer b and one more a . After B steps, all the b 's will have assimilated to all the a 's, and the derivation will converge. Symmetrically, if there are more b 's than a 's in the input the derivation will converge on a form containing only b 's after A steps.

Plurality Rule in HS with more than two classes Plurality Rule requires the same output constraint AGREE-GLOBAL, but a larger set of symbols: a, b, c, \dots, m . For each a in a candidate, AGREE-GLOBAL assigns one violation for each b , one violation for each c , and so on, yielding one violation for every segment that is not an a . The total number of violations is the sum over the alphabet of these products. It is always optimal to assimilate a member of the smallest class to the largest class.

Consider an input made of a 's, b 's, c 's, and so on with $n > 2$ total segments. Define A as the total number of a 's in the input, B as the total number of b 's in the input, and so on. Let the class of a 's be the largest, and let the class of b 's be the smallest: $A > C, D, \dots, M > B$. Let GEN have only one operation which changes one segment into another. There are numerous candidate types on any given step, but only three matter: the fully faithful candidate $\text{Cand}_{\text{faithful}}$, the Plurality Rule candidate(s) where a b became an a $\text{Cand}_{b \rightarrow a}$, and the other unfaithful candidate(s) where some segment changed to some other segment $\text{Cand}_{i \rightarrow j}$.

$\text{Cand}_{\text{faithful}}$ incurs $A(n - A) + B(n - B) + \dots + M(n - M)$ violations of AGREE-GLOBAL. $\text{Cand}_{b \rightarrow a}$ incurs $(A + 1)(n - A - 1) + (B - 1)(n - B + 1) + \dots + M(n - M)$ violations. $\text{Cand}_{i \rightarrow j}$ incurs $A(n - A) + B(n - B) + \dots + (I - 1)(n - I + 1) + (J + 1)(n - J - 1) + \dots + M(n - M)$ violations. AGREE-GLOBAL prefers $\text{Cand}_{b \rightarrow a}$ because it incurs the fewest violations:

$$\text{Cand}_{b \rightarrow a} \succ \text{Cand}_{\text{faithful}}$$

$$\# \text{ violations by } \text{Cand}_{b \rightarrow a} < \# \text{ violations by } \text{Cand}_{\text{faithful}}$$

$$(A + 1)(n - A - 1) + (B - 1)(n - B + 1) + \dots + M(n - M) < A(n - A) + B(n - B) + \dots + M(n - M)$$

After some algebra, we have $A > B + 1$, which we know to be true. AGREE-GLOBAL prefers the Plurality Rule candidate(s) to the fully faithful candidate.

$$\text{Cand}_{b \rightarrow a} \succ \text{Cand}_{i \rightarrow j}$$

$$\# \text{ violations by } \text{Cand}_{b \rightarrow a} < \# \text{ violations by } \text{Cand}_{i \rightarrow j}$$

$$(A + 1)(n - A - 1) + (B - 1)(n - B + 1) + \dots + I(n - I) + J(n - J) + \dots + M(n - M) < A(n - A) + B(n - B) + \dots + (I - 1)(n - I + 1) + (J + 1)(n - J - 1) + \dots + M(n - M)$$

After some algebra, we have $A - B > J - I$, which we know to be true: the difference between the largest class and the smallest class is greater than the difference between any two other classes. AGREE-GLOBAL prefers the Plurality Rule candidate(s) to other unfaithful candidates.

This confirms that assimilating the smallest class to the largest improves the most on AGREE-GLOBAL.

Congruent dissimilation in HS For space, I only derive congruent dissimilation for two symbols, a and b . It is straightforward to generalize it to arbitrarily large alphabets. Consider the output constraint $*X \dots X$, which assigns one violation to every identical pair of segments in a candidate: $\{a \dots a, b \dots b\}$. For each a in a candidate, $*X \dots X$ assigns one violation for every other a , and for each b in a candidate $*X \dots X$ assigns one violation for every other b .

For simplicity, consider an even-parity input made of a 's and b 's with $2n$ total segments; odd-parity inputs follow along similar lines. Consider two candidate types: the congruent candidate $\text{Cand}_{A=B}$ and a non-congruent candidate $\text{Cand}_{A \neq B}$. Define k as the difference between A and B in $\text{Cand}_{A \neq B}$, where $1 \leq k \leq n$. Let there be more a 's than b 's in $\text{Cand}_{A \neq B}$: $k = A - B$. $*X \dots X$ prefers $\text{Cand}_{A=B}$, because it incurs the fewest violations:

$$\text{Cand}_{A=B} \succ \text{Cand}_{A \neq B}$$

$$\# \text{ violations by } \text{Cand}_{A=B} < \# \text{ violations by } \text{Cand}_{A \neq B}$$

$$\frac{\overbrace{\# a \dots a}^{n(n-1)}}{2} + \frac{\overbrace{\# b \dots b}^{n(n-1)}}{2} < \frac{\overbrace{\# a \dots a}^{(n+k)(n+k-1)}}{2} + \frac{\overbrace{\# b \dots b}^{(n-k)(n-k-1)}}{2}$$

After some algebra, we have $0 < k^2$. Incongruent candidates incur an additional k^2 violations of $*X \dots X$. Reducing k at each step reduces the total number of violations, yielding gradual congruent dissimilation in HS. In even-parity inputs, k reduces to 0, and in odd-parity inputs, k reduces to 1.

Midpoint Pathology in HS The mechanisms underlying the Midpoint Pathology and congruent dissimilation are identical. Candidates are divided into two classes and the difference in their cardinalities is minimized. In congruent dissimilation, the classes are defined by segment type, e.g., *a*'s and *b*'s. In the Midpoint Pathology, classes are defined by location relative to the center. Following Eisner's (1997b) example of centering a high tone, I define one class *l* as the syllables to the left of the high tone, and the other class *r* as the syllables to the right of the high tone:

$$\underbrace{\sigma \dots \sigma \sigma \sigma}_l \acute{\sigma} \underbrace{\sigma \sigma \sigma \dots \sigma}_r$$

n

Consider an output constraint ALIGN, which assigns one violation to every pair of syllables to the left of the high tone, and every pair of syllables to the right of the high tone. Note that this assigns fewer total violations than Eisner's (1997b) constraint. Consider candidates of length n with one high-toned syllable. Define L as the number of syllables to the left of the high tone and R as the number of syllables to its right: $L + R = n - 1$. ALIGN prefers the high tone to dock as close to the center of the string as possible; in even-parity inputs, it will surface adjacent to the absolute center, e.g., $\sigma \acute{\sigma} \sigma \sigma \sim \sigma \sigma \acute{\sigma} \sigma$.

For simplicity, consider an odd-parity input with $2n + 1$ syllables and two candidate types: the centered candidate $\text{Cand}_{L=R}$, and an off-center candidate $\text{Cand}_{L \neq R}$. Let k be the difference between L and R in $\text{Cand}_{L \neq R}$, where $1 \leq k \leq n$. Let there be more syllables to the left of the high tone in $\text{Cand}_{L \neq R}$: $k = L - R$. ALIGN prefers the $\text{Cand}_{L=R}$, because it incurs the fewest violations:

$$\text{Cand}_{L=R} \succ \text{Cand}_{L \neq R}$$

$$\# \text{ violations by } \text{Cand}_{L=R} < \# \text{ violations by } \text{Cand}_{L \neq R}$$

$$\frac{\# l \dots l}{2} + \frac{\# r \dots r}{2} < \frac{\# l \dots l}{2} + \frac{\# r \dots r}{2}$$

$$\frac{n(n-1)}{2} + \frac{n(n-1)}{2} < \frac{(n+k)(n+k-1)}{2} + \frac{(n-k)(n-k-1)}{2}$$

This is exactly the same setup as with congruent dissimilation. The difference between L and R in the off-center candidate incurs additional violations of ALIGN, and the centered candidate is preferred.

References

- Baković, Eric (2000). *Harmony, dominance and control*. Ph.D. thesis, Rutgers, The State University of New Jersey.
- Bennett, Wm. G. (2013). *Dissimilation, Consonant Harmony, and Surface Correspondence*. Ph.D. thesis, Rutgers, The State University of New Jersey.
- Bennett, Wm. G. (2015). *The Phonology of Consonants*. Cambridge University Press, Cambridge.
- Bíró, Tamás (2003). Quadratic alignment constraints and finite state Optimality Theory. *Proceedings of the Workshop on Finite-State Methods in Natural Language Processing (FSM/NLP)*, 119–126.
- Cook, Eung-Do (1979). Palatalization and related rules in Sarcee. Cook, Eung-Do & Jonathan Kaye (eds.), *Linguistic Studies of Native Canada*, University of British Columbia Press, Vancouver, 19–35.
- Cook, Eung-Do (1984). *A Sarcee Grammar*. University of British Columbia Press, Vancouver.
- Eisner, Jason (1997a). Efficient generation in Primitive Optimality Theory. *Proceedings of the 35th Annual ACL and 8th European ACL*, 313–320.
- Eisner, Jason (1997b). What constraints should OT allow? Paper presented at LSA 71.
- Eisner, Jason (2000). Directional constraint evaluation in Optimality Theory. *Proceedings of the 18th International Conference on Computational Linguistics*, 257–263.
- Finley, Sara (2008). *Formal and Cognitive Restrictions on Vowel Harmony*. Ph.D. thesis, Johns Hopkins University.
- Finley, Sara (2009). Directionality in vowel harmony. Lehnert-LeHouillier, Heike & Alex B. Fine (eds.), *University of Rochester Working Papers in the Language Sciences*, vol. 5, 69–88.
- Finley, Sara (2010). Exceptions in vowel harmony are local. *Lingua* 120, 1549–1566.
- Frank, Robert & Giorgia Satta (1998). Optimality Theory and the generative complexity of constraint violability. *Computational Linguistics* 24:2, 307–315.
- Gerdemann, Dale & Mans Hulden (2012). Practical finite state Optimality Theory. *Proceedings of the 10th International Workshop on Finite State Methods and Natural Language Processing*, 10–19.
- Hansson, Gunnar Ólafur (2001). *Theoretical and Typological Issues in Consonant Harmony*. Ph.D. thesis, University of California, Berkeley.

- Hansson, Gunnar Ólafur (2007). Blocking effects in Agreement by Correspondence. *Linguistic Inquiry* 38:2, 395–409.
- Hansson, Gunnar Ólafur (2010). *Consonant Harmony: Long-Distance Interactions in Phonology*. University of California Press, Berkeley, CA.
- Hansson, Gunnar Ólafur (2014). (Dis)agreement by (non)correspondence: Inspecting the foundations. *Agreement by Correspondence: ABC↔Conference Archive*, 3–62.
- Heinz, Jeffrey (2007). *Inductive Learning of Phonotactic Patterns*. Ph.D. thesis, University of California Los Angeles.
- Heinz, Jeffrey (2010). Learning long-distance phonotactics. *Linguistic Inquiry* 41, 623–661.
- Heinz, Jeffrey (2018). The computational nature of phonological generalizations. Hyman, Larry M. & Frans Plank (eds.), *Phonological Typology*, Mouton, 126–195.
- Heinz, Jeffrey & Regine Lai (2013). Vowel harmony and subsequenceality. *Proceedings of the 13th Meeting on the Mathematics of Language*, 52–63.
- Hyde, Brett (2012). Alignment constraints. *Natural Language and Linguistic Theory* 30, 789–836.
- Hyde, Brett (2016). *Layering and Directionality*. Equinox, Sheffield.
- Johnson, C. Douglas (1972). *Formal Aspects of Phonological Description*. Mouton, The Hague.
- Jurcec, Peter (2011). *Feature Spreading 2.0: A Unified Theory of Assimilation*. Ph.D. thesis, University of Tromsø.
- Kaplan, Ronald & Martin Kay (1994). Regular models of phonological rule systems. *Computational Linguistics* 20, 331–378.
- Kimper, Wendell (2016). Positive constraints and finite goodness in Harmonic Serialism. McCarthy, John J. & Joe Pater (eds.), *Harmonic Grammar and Harmonic Serialism*, Equinox Publishing, London, 221–235.
- Lamont, Andrew (2018). Precedence is pathological: The problem of alphabetical sorting. Submitted to *Proceedings of WCCFL 36*.
- Lamont, Andrew (2019). Directional Harmonic Serialism. Qualifying paper, University of Massachusetts Amherst.
- Lombardi, Linda (1999). Positional faithfulness and voicing assimilation in Optimality Theory. *Natural Language & Linguistic Theory* 17, 267–302.
- McCarthy, John J. (2000). Harmonic serialism and parallelism. Hirotani, Masako, Andries Coetzee, Nancy Hall & Ji-Yung Kim (eds.), *Proceedings of the North East Linguistics Society 30*, GLSA, Amherst, MA, 501–524.
- McCarthy, John J. (2003). OT constraints are categorical. *Phonology* 20, 75–138.
- McCarthy, John J. (2004). Optimal paradigms. Downing, Laura J., T. Alan Hall & Renate Raffelsiefen (eds.), *Paradigms in Phonological Theory*, Oxford University Press, Oxford, 170–210.
- McCarthy, John J. (2008). Restraint of analysis. Blaho, Sylvia & Patrik Bye (eds.), *Freedom of Analysis?*, De Gruyter, 203–231.
- McCarthy, John J. (2010). Autosegmental spreading in Optimality Theory. Goldsmith, John A., Elizabeth Hume & W. Leo Wetzels (eds.), *Tones and Features*, Walter de Gruyter, Berlin, 195–222.
- McCarthy, John J. (2016). The theory and practice of Harmonic Serialism. McCarthy, John J. & Joe Pater (eds.), *Harmonic Grammar and Harmonic Serialism*, Equinox Publishing, London, 47–87.
- McMullin, Kevin (2016). *Tier-Based Locality in Long-Distance Phonotactics: Learnability and Typology*. Ph.D. thesis, The University of British Columbia.
- McMullin, Kevin & Gunnar Ólafur Hansson (2015). Long-distance phonotactics as tier-based strictly 2-local languages. Albright, Adam & Michelle A. Fullwood (eds.), *Proceedings of the 2014 Annual Meeting on Phonology*.
- Pater, Joe, Rajesh Bhatt & Christopher Potts (2007). Linguistic optimization. Unpublished manuscript.
- Prince, Alan & Paul Smolensky (1993/2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Blackwell Publishing, Malden, MA.
- Pruitt, Kathryn (2009). Two types of ties (some notes). Unpublished manuscript.
- Pulleyblank, Douglas (2002). Harmony drivers: No disagreement allowed. Larson, Julie & Mary Paster (eds.), *Proceedings of the Twenty-Eighth Annual Meeting of the Berkeley Linguistics Society*, 249–267.
- Riggle, Jason (2004). *Generation, Recognition, and Learning in Finite State Optimality Theory*. Ph.D. thesis, University of California Los Angeles.
- Rogers, James, Jeffrey Heinz, Gil Bailey, Matt Edlefsen, Molly Visscher, David Wellcome & Sean Wibel (2010). On languages piecewise testable in the strict sense. *The Mathematics of Language* 10/11, 255–265.
- Rose, Sharon & Rachel Walker (2004). A typology of consonant agreement as correspondence. *Language* 80:3, 475–531.
- Smolensky, Paul & Géraldine Legendre (2006). Formalizing the principles II: Optimization and grammar. Smolensky, Paul & Géraldine Legendre (eds.), *The harmonic mind: From neural completeness to Optimality-Theoretic grammar*, MIT Press, Cambridge, MA, vol. I, 207–234.
- Suzuki, Keiichiro (1998). *A Typological Investigation of Dissimilation*. Ph.D. thesis, University of Arizona.
- Walker, Rachel (2000). Long-distance consonantal identity effects. Billerey, Roger & Brook Lillehaugen (eds.), *Proceedings of WCCFL 19*, Cascadilla Press, Somerville, MA, 532–545.
- Walker, Rachel (2010). Nonmyopic harmony and the nature of derivations. *Linguistic Inquiry* 41:1, 169–179.
- Walker, Rachel (2015). Surface correspondence and discrete harmony triggers. Albright, Adam & Michelle A. Fullwood (eds.), *Proceedings of the 2014 Annual Meeting on Phonology*, Linguistic Society of America, Washington, DC.
- Wilson, Colin (2003). Analyzing unbounded spreading with constraints: marks, targets and derivations. Unpublished manuscript.
- Wilson, Colin (2006). Unbounded spreading is myopic. Paper presented at Current Perspectives on Phonology. Indiana University.