What do harmony-based grammars exclude?

Antón de la Fuente[(a)], Sarang Jeong[(a)], Arto Anttila[(a)], and Giorgio Magri[(b)]
(a) Department of Linguistics, Stanford University and (b) SFL, University of Paris 8, CNRS

1 Introduction

Maximum entropy (ME) grammars are widely adopted in current research on phonological variation (Hayes & Wilson 2008; Wilson 2025). What phonological assumptions does adopting ME entail? The question is difficult because ME consists of a bundle of assumptions. In this paper, we focus on one fundamental assumption: the probability of realizing an underlying form as a specific candidate surface form is proportional to that candidate's harmony score that depends *only* on the candidate's constraint violation profile. We call this the **harmony-basedness assumption**.

Section 2 makes the harmony-basedness assumption explicit. Section 3 asks whether this assumption is restrictive in and of itself, even when we make no assumptions about the specific way in which harmony scores are calculated. Section 4 considers the situation where the constraint set is distinctive, i.e., any two mappings are assigned different numbers of violations by at least one constraint. Here the answer is straightforwardly negative: the harmony-basedness assumption is not restrictive. Section 5 looks at situations where distinctiveness massively fails. This involves cases where we can construct a list of underlying forms such that each underlying form in the list has a candidate that shares a constraint violation profile with a candidate of the following underlying form and the final underlying form in the list has a candidate that shares a constraint violation profile with a candidate of the first underlying form. We call such a list of underlying forms a **cycle**.

Section 6 shows that the probabilities predicted by harmony-based grammars for the candidates in a cycle satisfy a special identity, no matter the mathematical shape of the harmony used to compute those harmony-based probabilities. This **cycle identity** thus shows that harmony-basedness is intrinsically restrictive. Furthermore, Section 7 shows that this cycle identity is all that there is to harmony-basedness: any empirical rates that satisfy this identity can be fitted through harmony-based probabilities, as long as we have complete freedom in the choice of the harmony scores. We conclude that the cycle identity provides a complete answer to our question of the intrinsic restrictiveness of harmony-basedness.

The rest of the paper starts to take a closer look at the cycle identity. Section 8 considers Noisy or Stochastic Harmonic Grammars (SHG; Boersma & Pater 2016)[1] and shows that they do not predict the cycle identity. In other words, the cycle identity distinguishes between ME and SHG. This raises the question of whether the cycle identity is empirically valid. As an initial contribution, Section 9 constructs cycles that contain an HG impossible mapping in the sense that no categorical HG grammar contains it and speculates that in such cases the cycle identity predicted by harmony-based grammars such as ME is paradoxical. Section 10 concludes that the cycle identity might provide a useful new tool for comparing ME and SHG phonology.

2 Harmony-based probabilistic grammars

Within constraint-based phonology, a **phonological system** is described through two ingredients. The first ingredient is a set Gen consisting of all the phonological mappings that are relevant for the description of

* Financial support from the France-Stanford Center for Interdisciplinary Studies is gratefully acknowledged.

[1] The literature often uses the acronym NHG instead of SHG. We prefer the latter to underscore the analogy with SOT.

Proceedings of *AMP 2023/2024*

the phonological processes of interest. We construe a phonological mapping as a pair (x, y) of an underlying form x and a surface realization y. The second ingredient is a set $\boldsymbol{C}$ consisting of a finite number n of constraints $C_1, \ldots, C_n$ that suffice to quantify all the phonologically relevant properties of the mappings listed by Gen. The number of violations assigned by a constraint C_k to a mapping (x, y) is denoted by $C_k(\mathsf{x}, \mathsf{y})$. The violation profile of a mapping (x, y) is the tuple $\boldsymbol{C}(\mathsf{x}, \mathsf{y}) = (C_1(\mathsf{x}, \mathsf{y}), \ldots, C_n(\mathsf{x}, \mathsf{y}))$ of its numbers of constraint violations, usually represented as a row in a familiar tableau.

Within this framework, we want to define a **probabilistic grammar** G that assigns a positive number $G(\mathsf{y} \,|\, \mathsf{x})$ to any mapping (x, y) listed by Gen. We interpret this number $G(\mathsf{y} \,|\, \mathsf{x})$ as the probability that the given underlying form x is realized as the designated surface form y from the candidate set $Gen(\mathsf{x})$, as suggested by the notation $G(\cdot \,|\, \cdot)$ for conditional probability. This probabilistic interpretation requires these numbers $G(\mathsf{y} \,|\, \mathsf{x})$ to be **normalized**: their sum over all surface forms y in the candidate set $Gen(\mathsf{x})$ must be equal to one, namely $\sum_{\mathsf{y} \in Gen(\mathsf{x})} G(\mathsf{y} \,|\, \mathsf{x}) = 1$ (we suppose that the candidate set $Gen(\mathsf{x})$ only lists finitely many surface forms, whereby this sum is always finite and normalization therefore always possible).

Because of normalization, it suffices to define the number $G(\mathsf{y} \,|\, \mathsf{x})$ only up to a multiplicative proportionality constant, as normalization then determines the value of that constant. In the most general case, this number $G(\mathsf{y} \,|\, \mathsf{x})$ can be proportional to a quantity that is **complex** because it depends not only on the designated candidate y but also on all other candidates of the underlying form x. To illustrate, we could order all the candidate surface forms in $Gen(\mathsf{x})$ in decreasing order of quality, observe that the designated candidate y is kth from the top of the list, and thus define its probability $G(\mathsf{y} \,|\, \mathsf{x})$ to be proportional to the power p^k (for some constant p between zero and one), as indeed assumed by Labov's (1969) "postulate of geometric ordering". Since the position of a candidate y in the list can only be determined by taking the entire candidate set $Gen(\mathsf{x})$ into consideration, Labov's postulate predicts the probability $G(\mathsf{y} \,|\, \mathsf{x})$ of a designated candidate y to be proportional to a quantity that depends on all other candidates in $Gen(\mathsf{x})$ as well. Against this background, we now make two restrictive assumptions.

Our **first assumption** is to move away from such fully general, complex probabilistic models and instead require the probability $G(\mathsf{y} \,|\, \mathsf{x})$ of a designated candidate y to be proportional to some quantity $H(\mathsf{x}, \mathsf{y})$ that is **simple** because it only depends on the designated candidate y considered but not on the other candidates in $Gen(\mathsf{x})$, as stated in (1). In other words, this score $H(\mathsf{x}, \mathsf{y})$ quantifies the phonological quality or **harmony** of the candidate y as a surface realization of the underlying form x in absolute terms, without comparing it to other candidate surface realizations listed by $Gen(\mathsf{x})$. Indeed, these other candidates enter into the final expression of $G(\mathsf{y} \,|\, \mathsf{x})$ only through the back door, namely only through normalization.

(1) $G(\mathsf{y} \,|\, \mathsf{x}) \propto H(\mathsf{x}, \mathsf{y})$

Our **second assumption** is that the harmony score $H(\mathsf{x}, \mathsf{y})$ of a phonological mapping (x, y) is actually a function solely of the constraint violations of the mapping (x, y), as stated in (2). In other words, the harmony score $H(\mathsf{x}, \mathsf{y})$ ignores any properties of underlying and surface forms x and y that are not encoded by the constraints. Also this second assumption is substantive because the constraint set $\boldsymbol{C}$ is allowed to list only a finite number n of constraints while Gen can list an infinite number of mappings, say, because it lists the infinitely many underlying strings that can be constructed out of a finite alphabet of segments. Because of this assumption (2), H has the mathematical shape of a function that takes n numbers (interpreted as numbers of constraint violations) and returns a positive score.

(2) $H(\mathsf{x}, \mathsf{y}) = H\big(C_1(\mathsf{x}, \mathsf{y}), \ldots, C_n(\mathsf{x}, \mathsf{y})\big)$

By combining our two assumptions (1) and (2), we obtain the definition (3) of the probabilistic phonological grammar $G_H^{\boldsymbol{C}}$ **based** on a constraint set $\boldsymbol{C}$ through a harmony function H.[2] It predicts that a given underlying form x is realized as a designated candidate y with a probability $G_H^{\boldsymbol{C}}(\mathsf{y} \,|\, \mathsf{x})$ proportional to the harmony score of the constraint violation profile $(C_1(\mathsf{x}, \mathsf{y}), \ldots, C_n(\mathsf{x}, \mathsf{y}))$ of the phonological mapping (x, y). To illustrate, we assign to each constraint C_k some non-negative weight w_k whose relative size quantifies the relative importance of that constraint C_k and we define the harmony score $H(\mathsf{x}, \mathsf{y})$ of a mapping (x, y) as the exponential of the opposite of the sum of its weighted constraint violations, namely

[2] The notation $G_H^{\boldsymbol{C}}$ highlights the dependence of a harmony-based grammar on the constraint set $\boldsymbol{C}$ and the harmony function H but hides its dependence on Gen. We avoid the more cumbersome notation $G_H^{Gen, \boldsymbol{C}}$ because Gen can be reconstructed from $\boldsymbol{C}$ when the constraints are explicitly construed as functions from Gen to non-negative integers.

$H(\mathsf{x},\mathsf{y}) = \exp(-\sum_{k=1}^{n} w_k C_k(\mathsf{x},\mathsf{y}))$. These harmony scores satisfy assumption (2) because they are fully defined in terms of the constraint violations. The grammar $G_H^{\boldsymbol{C}}$ that predicts probabilities proportional to these harmony scores as in (3) is the familiar ME grammar.

(3) $G_H^{\boldsymbol{C}}(\mathsf{y}\,|\,\mathsf{x}) \propto H\big(C_1(\mathsf{x},\mathsf{y}),\ldots,C_n(\mathsf{x},\mathsf{y})\big)$

3 Is harmony-basedness intrinsically restrictive?

The preceding section has derived the harmony-basedness assumption (3) from the two assumptions (1) and (2) that probabilities are proportional to harmony scores that depend on a single candidate and only through its constraint violation profile. This paper addresses the following question: is this harmony-basedness assumption (3) **intrinsically restrictive** in and of itself? Namely, is this assumption restrictive even when we make no assumptions whatsoever about the shape of the harmony scores?

Let us make this question explicit. For each mapping (x,y) listed by Gen, we assume that we have a corpus or psycholinguistic or sociolinguistic tools to exactly measure the **empirical rate** $\widehat{R}(\mathsf{y}\,|\,\mathsf{x})$ with which the underlying form x is realized as the surface variant y by a certain speaker of a certain dialect in a certain register. The grammar $G_H^{\boldsymbol{C}}$ based on the relevant constraint set $\boldsymbol{C}$ through some harmony function H **fits** these empirical rates provided it ideally satisfies the identity (4) for any mapping (x,y) listed by Gen.

(4) $\widehat{R}(\mathsf{y}\,|\,\mathsf{x}) = G_H^{\boldsymbol{C}}(\mathsf{y}\,|\,\mathsf{x})$

The question of the intrinsic restrictiveness of harmony-basedness can now be formulated explicitly as follows: given a phonological system $Gen, \boldsymbol{C}$, are there patterns of empirical rates $\widehat{R}(\mathsf{y}\,|\,\mathsf{x})$ that cannot be fitted by the harmony-based grammar $G_H^{\boldsymbol{C}}$ in (3), no matter how we choose the harmony function H? If we can indeed construct such counterexample rates that cannot be fitted no matter the harmony H, we can conclude that the harmony-basedness assumption (3) is intrinsically restrictive – "intrinsically" in the sense that the assumption is restrictive in and of itself, without assumptions about the mathematical shape of the harmony functions H we are allowed to use. If it is instead always possible to cobble together a harmony function H that does the job no matter the rates, we conclude that harmony-basedness is not intrinsically restrictive.

This paper tackles the question of the intrinsic restrictiveness of harmony-basedness. Before getting started let us pause to discuss why this formal question is relevant for phonological theory. A number of authors in the probabilistic constraint-based literature have endorsed ME. But what phonological assumptions are we actually endorsing when we subscribe to ME? This is a difficult question because ME is a bundle of technical assumptions: the assumption that harmony scores depend on constraint violations only through their weighted sum; the assumption that the link between this sum and the harmony score is exponential; and so on. Underlying these multiple technical ME's assumptions there is the more basic assumption (3) of harmony-basedness. From this perspective, the most basic question raised by ME phonology is indeed the question of the intrinsic restrictiveness of harmony-basedness addressed here.

Let us make this point more concrete. Hayes and Zuraw (Zuraw & Hayes 2017; Hayes 2022) have recently argued that the rates of application of a number of variable phonological processes in a number of languages obey what they call the **wug curve generalization**. Furthermore, they observe that ME grammars succeed at predicting this empirical generalization. Magri (2025) adds that if we restrict ourselves to probabilistic grammars that are harmony-based in the sense of assumption (3) then ME grammars are the only grammars that predict the wug curve generalization.[3] In other words *harmony-basedness (3) + wug curve generalization = ME*. Can we conclude from this result that the wug curve generalization provides an empirical justification for ME phonology? Well, only if the result can be simplified down to *wug-curve generalization = ME* because the harmony-basedness assumption (3) can be made for free. And that in turn depends on the answer to the question of the intrinsic restrictiveness of harmony-basedness addressed here.

Another reason why this question is relevant concerns the comparison between ME and other models of probabilistic phonology such as SOT and SHG. Obviously, SOT and SHG grammars are constraint-based just like ME grammars: they are defined solely in terms of constraint violations. Yet, while ME grammars are explicitly defined in terms of harmony scores, SOT and SHG grammars are not. In order to better compare

[3] More precisely, a harmony-based grammar predicts HZ's generalization irrespectively of the constraints if and only if it is either a ME grammar or else it can be construed as a ME grammar through a constraint transformation.

ME versus SOT and SHG, we are thus interested in the following question: given a constraint set $\boldsymbol{C}$, is it always possible to construct some harmony function H such that an SOT or SHG grammar corresponding to that constraint set $\boldsymbol{C}$ coincides with the harmony-based grammar $G_H^{\boldsymbol{C}}$ in (3)? This question is a special case of the question of the intrinsic restrictiveness of harmony-basedness addressed here.

4 When the constraint set is distinctive

We start our discussion of the intrinsic restrictiveness of harmony-basedness by discussing a trivial case. We make the assumption that the relevant constraint set $\boldsymbol{C}$ is **distinctive** in the sense that any two different mappings listed by Gen are assigned different numbers of violations by at least one constraint in $\boldsymbol{C}$ and thus have different constraint violation profiles. When the constraint set $\boldsymbol{C}$ is this rich, the harmony-basedness assumption (3) turns out not to be intrinsically restrictive: any pattern of rates $\widehat{R}(\mathsf{y}\,|\,\mathsf{x})$ can be fitted by some harmony-based grammar $G_H^{\boldsymbol{C}}$, as long as we have complete freedom in the choice of the harmony function H from constraint violation profiles to positive harmony scores.

In fact, let us define the harmony score $H(\mathsf{x},\mathsf{y})$ of a phonological mapping (x,y) simply as its empirical rate $\widehat{R}(\mathsf{y}\,|\,\mathsf{x})$, as stated in (5). Because of the assumption that the constraint set $\boldsymbol{C}$ is distinctive, the resulting harmony H is indeed a function solely of the constraint violations, as required by (2). To appreciate this point, suppose instead that the constraint set $\boldsymbol{C}$ were not distinctive because some mappings (x,y) and $(\widehat{\mathsf{x}},\widehat{\mathsf{y}})$ share the same number of constraint violations $C_k(\mathsf{x},\mathsf{y}) = C_k(\widehat{\mathsf{x}},\widehat{\mathsf{y}})$ for every constraint C_k. Suppose furthermore that these two mappings (x,y) and $(\widehat{\mathsf{x}},\widehat{\mathsf{y}})$ have different rates $\widehat{R}(\mathsf{y}\,|\,\mathsf{x}) \neq \widehat{R}(\widehat{\mathsf{y}}\,|\,\widehat{\mathsf{x}})$. In this case, the position (5) would contradict the assumption (2) that the harmony scores only depend on the constraint violations.

(5) $H(\mathsf{x},\mathsf{y}) = \widehat{R}(\mathsf{y}\,|\,\mathsf{x})$

The trivial reasoning in (6) shows that the grammar $G_H^{\boldsymbol{C}}$ based as in (3) on the harmony function H defined in (5) succeeds at fitting the empirical rates because it validates the identity (4) for any mapping (x,y). Step (6a) holds because of the assumption (1) that harmony-based probabilities are proportional to harmony scores, with the proportionality constant notated called $1/Z$ as usual. Step (6b) holds because normalization requires the constant Z to be equal to the sum of the harmony scores $H(\mathsf{x},\mathsf{z})$ of all surface forms z in the candidate set $Gen(\mathsf{x})$. Step (6c) holds because of the definition (5) of the harmony scores. Finally, step (6d) holds because the rates are normalized, whereby the sum of the rates $\widehat{R}(\mathsf{z}\,|\,\mathsf{x})$ over all surface forms z in $Gen(\mathsf{x})$ is equal to one.

(6) $$G_H^{\boldsymbol{C}}(\mathsf{y}\,|\,\mathsf{x}) \overset{(a)}{=} \frac{H(\mathsf{x},\mathsf{y})}{Z} \overset{(b)}{=} \frac{H(\mathsf{x},\mathsf{y})}{\sum\limits_{\mathsf{z}\in Gen(\mathsf{x})} H(\mathsf{x},\mathsf{z})} \overset{(c)}{=} \frac{\widehat{R}(\mathsf{y}\,|\,\mathsf{x})}{\sum\limits_{\mathsf{z}\in Gen(\mathsf{x})} \widehat{R}(\mathsf{z}\,|\,\mathsf{x})} \overset{(d)}{=} \widehat{R}(\mathsf{y}\,|\,\mathsf{x})$$

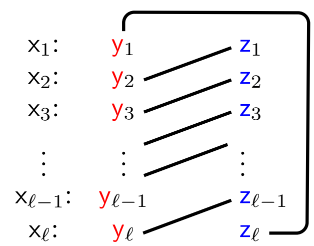
In conclusion, whenever Gen is described through a distinctive constraint set $\boldsymbol{C}$, the harmony-basedness assumption (3) is not intrinsically restrictive: any rates can be fitted by a harmony-based grammar, at least when we have complete freedom in the choice of the harmony function H. In order to investigate the intrinsic restrictiveness of harmony-basedness, we thus need to look at constraint sets $\boldsymbol{C}$ that are not distinctive because they **compress** two different mappings into the same constraint violation profile. Compression is abundant in the phonological literature (Anttila et al., 2019). Phonological processes apply to natural classes, namely to all forms that share some relevant property, ignoring the irrelevant ways in which they differ. Thus, in Latin, stress targets heavy syllables, but ignores vowel quality; in English, aspiration targets voiceless stops, but ignores place of articulation; in Finnish, vowel harmony targets [±back], but ignores the number of intervening consonants. Put differently, words with the same distribution of heavy and light syllables are stressed alike; voiceless stops are aspirated alike; and words with a different number of consonants harmonize alike. Such phonological equivalences are a key property of phonological systems and indeed the *raison d'être* of distinctive features.

5 When the constraint set is not distinctive: constraint cycles

Which patterns of constraint compression are relevant for testing the intrinsic restrictiveness of harmony-basedness? To address this question, we say that a list of ℓ (not necessarily distinct) underlying forms $\mathsf{x}_1, \ldots, \mathsf{x}_\ell$ is a **cycle** relative to a phonological system $(Gen, \boldsymbol{C})$ provided each underlying form x_i in the list has a candidate that shares the constraint violation profile with a candidate of the following underlying form x_{i+1} and the last underlying form x_ℓ has a candidate that shares the constraint violation profile with a candidate of the first underlying form x_1. Constraint distinctiveness thus massively fails because two candidates of any two consecutive underlying forms in the cycle are compressed into the same constraint violation profile.

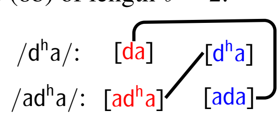
More explicitly, the underlying forms $x_1, \ldots, x_\ell$ form a cycle provided we can choose from each candidate set $Gen(x_i)$ two (different) candidates y_i and z_i that satisfy the identities in (7a) for every constraint C in the constraint set $\boldsymbol{C}$ considered. The candidate z_1 of the first underlying form x_1 has the same constraint violation profile as the candidate y_2 of the second underlying form x_2; the candidate z_2 of the second underlying form x_2 has the same constraint violation profile as the candidate y_3 of the third underlying form x_3; and so on until the candidate z_ℓ of the last underlying form x_ℓ, which has the same constraint violation profile as the candidate y_1 of the first underlying form x_1. We summarize this condition (7a) by drawing a line between any two candidates that share the same constraint violation profile, as in (7b).

(7) (a)
$$\begin{aligned} C(x_2, y_2) &= C(x_1, z_1) \\ C(x_3, y_3) &= C(x_2, z_2) \\ &\vdots \\ C(x_\ell, y_\ell) &= C(x_{\ell-1}, z_{\ell-1}) \\ C(x_1, y_1) &= C(x_\ell, z_\ell) \end{aligned}$$

(b)


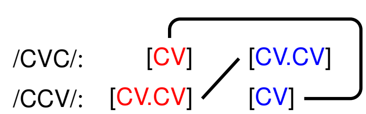
To illustrate, we consider the faithfulness constraints IDENT(voice) and IDENT(spread glottis) that protect underlying obstruent voicing and aspiration. We also consider the markedness constraints *[+voice], *[+spread], and *[+voice, +spread] against voiced, aspirated, and breathy voiced obstruents and *V[−voice]V, against intervocalic voiceless obstruents. The constraint identities in (8a) hold when C is one of the markedness constraints because no constraint that penalizes [d] or [dʰ] is sensitive to the distinction between intervocalic and non-intervocalic position. Furthermore, these identities in (8a) also hold when C is one of the faithfulness constraints because the two candidates compared by either identity stand in the same relationship (faithfulness or de-aspiration) with their underlying forms. When we draw a line between mappings that share the same constraint violation profile, we obtain the cycle (8b) of length $\ell = 2$.

(8) (a) C(/adʰa/, [adʰa]) = C(/dʰa/, [dʰa])
C(/dʰa/, [da]) = C(/adʰa/, [ada])

(b)


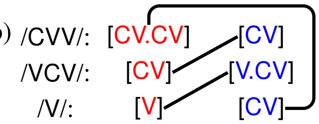
Next, we consider the markedness constraints NOCOMPONSET and NOCOMPCODA against complex syllable edges and ONSET and NOCODA against empty onsets and filled codas. We also consider the faithfulness constraints MAXC and MAXV against consonant and vowel deletion and DEPC and DEPV against consonant and vowel epenthesis. The constraint identities in (9a) hold when C is one of the markedness constraints because the two mappings compared share the same surface form. Furthermore, these identities also hold when C is one of the faithfulness constraints because the shared surface string is obtained through the same operations from the two underlying strings compared. For instance, the first identity in (9a) holds for the faithfulness constraints because the shared surface string [CV.CV] is obtained through vowel epenthesis from both underlying strings /CCV/ and /CVC/. When we draw a line between mappings that share the same constraint violation profile, we obtain the cycle (9b) of length $\ell = 2$.

(9) (a) C(/CCV/, [CV.CV]) = C(/CVC/, [CV.CV])
C(/CVC/, [CV]) = C(/CCV/, [CV])

(b)


Finally, we consider the same constraints for syllable structure just listed plus the markedness constraint NOHIATUS against two adjacent vowel nuclei. The identities in (10a) hold for each constraint C considered. The intuitive reason is that the string CV is "invisible" to markedness: two strings such as [V] and [V.CV] are identical relative to markedness just like x and $x+0$. When we draw a line between mappings that share the same constraint violation profile, we obtain the cycle (10b) of length $\ell = 3$.

(10) (a)
$$C(\text{/VCV/, [CV]}) = C(\text{/CVV/, [CV]})$$
$$C(\text{/V/, [V]}) = C(\text{/VCV/, [V.CV]})$$
$$C(\text{/CVV/, [CV.CV]}) = C(\text{/V/, [CV]})$$

(b) 

6 The behavior of harmony-based grammars on cycles

Given a cycle (7) relative to some phonological system $Gen, \boldsymbol{C}$, which probabilities are predicted for the candidates y_i and z_i in the cycle by the probabilistic grammar $G_H^{\boldsymbol{C}}$ that is based as in (3) on this constraint set $\boldsymbol{C}$ through some harmony function H? The reasoning in (11) answers this question: it shows that the product of the harmony-based probabilities of the candidates $y_1, \ldots, y_\ell$ is equal to the product of the probabilities of the candidates $z_1, \ldots, z_\ell$. Steps (11a) and (11d) hold because of the assumption (1) that the harmony-based probability of a candidate of the underlying form x_i is proportional to its harmony score. Equivalently, it is equal to that harmony score divided by a quantity $Z(x_i)$ that depends on the underlying form x_i but not on the specific candidate considered. Step (11b) holds because of the assumption (2) that the harmony scores depend only on the constraint violation profiles (not on the actual mappings) together with the assumption that the violation profiles satisfy the cycle identities in (7a). Step (11c) holds by simply reordering the numerators.

(11)
$$\prod_{i=1}^{\ell} G_H^{\boldsymbol{C}}(y_i \,|\, x_i) \overset{(a)}{=} \frac{H(x_1, y_1)}{Z(x_1)} \frac{H(x_2, y_2)}{Z(x_2)} \frac{H(x_3, y_3)}{Z(x_3)} \cdots \frac{H(x_\ell, y_\ell)}{Z(x_\ell)}$$
$$\overset{(b)}{=} \frac{H(x_2, z_\ell)}{Z(x_1)} \frac{H(x_3, z_1)}{Z(x_2)} \frac{H(x_4, z_2)}{Z(x_3)} \cdots \frac{H(x_{\ell-1}, z_{\ell-1})}{Z(x_\ell)}$$
$$\overset{(c)}{=} \frac{H(x_\ell, z_1)}{Z(x_1)} \frac{H(x_\ell, z_2)}{Z(x_2)} \frac{H(x_\ell, z_3)}{Z(x_3)} \cdots \frac{H(x_\ell, z_\ell)}{Z(x_\ell)} \overset{(d)}{=} \prod_{i=1}^{\ell} G_H^{\boldsymbol{C}}(z_i \,|\, x_i)$$

The identity between products of harmony-based probabilities obtained in (11) can be rewritten as the identity between ratios in (12a). These ratios make sense: the harmony-based probabilities in the denominators are different from zero because proportional to harmony scores that are strictly positive. For cycles of length $\ell = 2$, (12a) becomes (12b), which says in particular that, if the candidate y_1 is chosen with a larger probability than the candidate z_1 for the underlying form x_1 (whereby the ratio on the lefthand side is larger than one), then the candidate z_2 is chosen with a larger probability than y_2 for the other underlying form x_2 as well (the ratio on the righthand side must be larger than one as well). This makes sense because the two mappings in the two numerators share the same constraint violation profile as do the two mappings in the two denominators. From now on, we will refer to any of these equivalent formulations as the **cycle identity** predicted by harmony-based grammars, no matter the proper definition of the harmony scores.

(12) (a)
$$\frac{G_H^{\boldsymbol{C}}(y_1 \,|\, x_1)}{G_H^{\boldsymbol{C}}(z_1 \,|\, x_1)} = \prod_{i=2}^{\ell} \frac{G_H^{\boldsymbol{C}}(z_i \,|\, x_i)}{G_H^{\boldsymbol{C}}(y_i \,|\, x_i)}$$
(b)
$$\frac{G_H^{\boldsymbol{C}}(y_1 \,|\, x_1)}{G_H^{\boldsymbol{C}}(z_1 \,|\, x_1)} = \frac{G_H^{\boldsymbol{C}}(z_2 \,|\, x_2)}{G_H^{\boldsymbol{C}}(y_2 \,|\, x_2)}$$

To illustrate, we recall that (8) is a cycle of length $\ell = 2$ relative to the set $\boldsymbol{C}$ consisting of the six constraints for syllable structure listed above. Thus, no matter how we choose the harmony function H, the harmony-based grammar $G_H^{\boldsymbol{C}}$ based on this constraint set predicts the cycle identity (13). It says that the identities between constraint violation profiles that define the cycle in (8a) translate into an identity between the products of the harmony-based probabilities of the corresponding mappings.

(13)
$$G_H^{\boldsymbol{C}}(\text{[da]} \,|\, \text{/d}^{\text{h}}\text{a/}) \cdot G_H^{\boldsymbol{C}}(\text{[ad}^{\text{h}}\text{a]} \,|\, \text{/ad}^{\text{h}}\text{a/}) = G_H^{\boldsymbol{C}}(\text{[d}^{\text{h}}\text{a]} \,|\, \text{/d}^{\text{h}}\text{a/}) \cdot G_H^{\boldsymbol{C}}(\text{[ada]} \,|\, \text{/ad}^{\text{h}}\text{a/})$$

This cycle identity (13) between products can be rewritten as the identity (14) between ratios. This identity says that the ratio between the probability of de-aspiration (numerator) divided by the probability of faithful realization (denominator) for the underlying form /d$^{\text{h}}$a/ (lefthand side) is equal to the same ratio for the underlying form /ad$^{\text{h}}$a/ (righthand side). In other words, the ratio between the probabilities of removing or maintaining aspiration cannot depend on the difference between intervocalic versus non-intervocalic positions. This cycle identity thus extends to the probabilistic setting the fact that no categorical OT or HG grammar corresponding to the constraints considered can repair through de-aspiration one of the two underlying forms /dʰa/ or /adʰa/ while realizing the other underlying form faithfully. This prediction seems reasonable: de-aspiration is predicted to be insensitive to the difference between intervocalic versus non-intervocalic positions because the three constraints IDENT(spread glottis), *[+spread], and *[+voice, +spread] that are relevant for aspiration are insensitive to that positional difference.

(14) $$\frac{G_H^{\boldsymbol{C}}(\text{[da]}\,|\,\text{/d}^\text{h}\text{a/})}{G_H^{\boldsymbol{C}}(\text{[d}^\text{h}\text{a]}\,|\,\text{/d}^\text{h}\text{a/})} = \frac{G_H^{\boldsymbol{C}}(\text{[ada]}\,|\,\text{/ad}^\text{h}\text{a/})}{G_H^{\boldsymbol{C}}(\text{[ad}^\text{h}\text{a]}\,|\,\text{/ad}^\text{h}\text{a/})} \qquad \frac{\textit{de-aspiration}}{\textit{faithful}}$$

Next, the cycle identity for the cycle in (9) can be written directly in terms of ratios as in (15). This identity says that the ratio between the probability of repairing the coda of /CVC/ through consonant deletion (numerator) divided by the probability of repairing it through vowel epenthesis (denominator) is equal to the ratio between the probability of repairing the complex onset of /CCV/ through consonant deletion (numerator) divided by the probability of repairing it through vowel epenthesis (denominator). In other words, the ratio between the probabilities of the two repairs (deletion versus epenthesis) cannot depend on what is repaired (codas versus complex onsets). This cycle identity thus extends to the probabilistic setting the fact that no categorical OT or HG grammar corresponding to the constraints considered can choose one of the two repairs (consonant deletion versus vowel epenthesis) for codas but the other repair for complex onsets. In other words, no OT or HG grammar can model a language where the English word *proton* would be nativized as [ro.to.ni], where consonant deletion repairs the onset cluster while vowel epenthesis repairs the coda.

(15) $$\frac{G_H^{\boldsymbol{C}}(\text{[CV]}\,|\,\text{/CVC/})}{G_H^{\boldsymbol{C}}(\text{[CV.CV]}\,|\,\text{/CVC/})} = \frac{G_H^{\boldsymbol{C}}(\text{[CV]}\,|\,\text{/CCV/})}{G_H^{\boldsymbol{C}}(\text{[CV.CV]}\,|\,\text{/CCV/})} \qquad \frac{\textit{consonant deletion}}{\textit{vowel epenthesis}}$$

Finally, the cycle identity for the cycle in (10) can be written directly in terms of ratios as in (16). The lefthand side of this identity features the ratio (16a) between the probability of repairing /CVV/ through vowel deletion (numerator) divided by the probability of repairing it through consonant epenthesis (denominator). The righthand side features the somewhat analogous ratio (16b) between the probability of repairing /VCV/ through vowel deletion (numerator) divided by the probability of repairing the different underlying form /V/ through consonant epenthesis (denominator). Yet, the latter ratio is multiplied on the righthand side with the ratio (16c) between the probability of faithfully realizing /V/ (numerator) divided by the probability of faithfully realizing /VCV/ (denominator). This cycle identity is harder to interpret. We content ourselves with observing that no categorical OT or HG grammar corresponding to the constraints considered can realize the three underlying forms /CVV/, /VCV/, and /V/ in the cycle as [CV.CV], [CV], and [V] respectively, namely through consonant epenthesis, vowel deletion, and faithfully. Furthermore, no categorical OT or HG grammar corresponding to the constraints considered can realize the three underlying forms /CVV/, /VCV/, and /V/ as [CV], [V.CV], and [CV] respectively, namely through vowel deletion, faithfully, and through consonant epenthesis. In other words, the choices made by OT and HG categorical grammars for these three underlying forms must be mutually coherent.

(16) $$\underbrace{\frac{G_H^{\boldsymbol{C}}(\text{[CV]}\,|\,\text{/CVV/})}{G_H^{\boldsymbol{C}}(\text{[CV.CV]}\,|\,\text{/CVV/})}}_{(a)} = \underbrace{\frac{G_H^{\boldsymbol{C}}(\text{[CV]}\,|\,\text{/VCV/})}{G_H^{\boldsymbol{C}}(\text{[CV]}\,|\,\text{/V/})}}_{(b)} \underbrace{\frac{G_H^{\boldsymbol{C}}(\text{[V]}\,|\,\text{/V/})}{G_H^{\boldsymbol{C}}(\text{[V.CV]}\,|\,\text{/VCV/})}}_{(c)}$$

7 A complete characterization of harmony-based rates

The reasoning in (11) holds under no assumptions whatsoever on the shape of the harmony function H from n numbers to positive harmony scores. Thus, this reasoning provides an **initial** answer to our question of the intrinsic restrictiveness of harmony-basedness. Indeed, let us consider some empirical rates $\widehat{R}(\text{y}\,|\,\text{x})$ for the mappings listed by Gen and some constraint set $\boldsymbol{C}$ that is relevant for describing those mappings. We ask whether we can cook up some harmony function H such that the grammar $G_H^{\boldsymbol{C}}$ based on that constraint set $\boldsymbol{C}$ through that harmony function H as in (3) fits those rates. The reasoning in (11) says that a **necessary** condition for that to be the case is that the empirical rates considered satisfy condition (17) for any cycle (7) relative to the phonological system $Gen, \boldsymbol{C}$ considered.

(17) $\prod_{i=1}^{\ell} \widehat{R}(\mathsf{y}_i \,|\, \mathsf{x}_i) = \prod_{i=1}^{\ell} \widehat{R}(\mathsf{z}_i \,|\, \mathsf{x}_i)$

Is this initial answer to our question the **final** answer as well? In other words, is this necessary condition (17) on the empirical rates also **sufficient** for us to be able to construct a harmony function H that fits the rates? The answer to this question turns out to be positive as shown in a forthcoming longer version of this paper. This condition (17) on cycles is actually all it takes for a pattern of empirical rates to conform with the harmony-basedness assumption (3) as summarized in the following box.

> *Consider a collection Gen of phonological mappings; an assignment of rates $\widehat{R}(\mathsf{y} \,|\, \mathsf{x})$ to each mapping (x, y) in Gen; and a constraint set **C** for Gen. The following two conditions are equivalent:*
>
> ***(a)*** *There exists a harmony function H such that the grammar $G_H^{\boldsymbol{C}}$ based on this harmony fits the rates $\widehat{R}(\mathsf{y} \,|\, \mathsf{x})$ in the sense that the identity (4) holds for any mapping (x, y) listed by Gen.*
>
> ***(b)*** *For any cycle (7) of any length ℓ relative to the phonological system $Gen, \boldsymbol{C}$, the product of the empirical rates $\widehat{R}(\mathsf{y}_i \,|\, \mathsf{x}_i)$ is equal to the product of the empirical rates $\widehat{R}(\mathsf{z}_i \,|\, \mathsf{x}_i)$ as in (17).*

This boxed result entails in particular that if a constraint set ***C*** predicts no cycles the harmony-based assumption is not intrinsically restrictive: we can always construct a harmony function H that fits any rates. Obviously, a distinctive constraint set predicts no cycles. The observation in Section 4 that harmony-basedness is not intrinsically restrictive when the constraint set ***C*** is distinctive thus follows as corollary.

8 Non-harmony-based grammars can behave differently on cycles

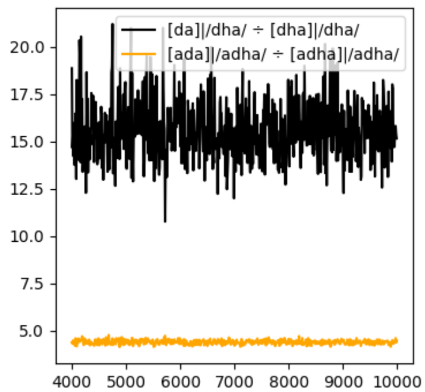
Because of the assumption (2) that harmony scores are a function solely of the constraint violation profiles, all harmony-based grammars are constraint-based in the traditional, more basic sense of being defined solely in terms of constraint violations. As recalled in Section 3, SHG grammars are also constraint-based in this traditional, basic sense. This section uses the three examples of cycles from Section 5 to show that SHG nonetheless does not predict the cycle identity that characterizes harmony-basedness. A couple of technical remarks concerning SHG are in order before we start.

The SHG probability of a mapping (x, y) given a weight vector ***w*** cannot be computed exactly. Thus, we estimate it in the usual way: for a certain number N of times, we add noise to each component of the weight vector ***w***; we use the resulting noisy weight vector to compute the categorical HG winner $\widehat{\mathsf{y}}$ for the underlying form x; and we estimate the SHG probability of the mapping (x, y) through the ratio η/N, where η is the total number of times the predicted winner $\widehat{\mathsf{y}}$ is equal to the intended winner y. In the simulations described below, we choose $N = 100,000$, which is shown to give sufficiently accurate estimates of the SHG probabilities.

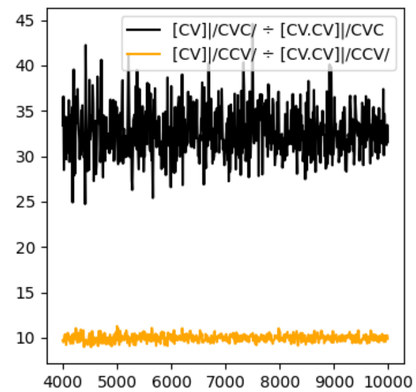
The SHG literature traditionally samples the noise with a gaussian distribution $p(x) = \frac{1}{\sqrt{2\pi}}\exp(-x^2/2)$ with zero mean and small variance. Yet, the choice of the gaussian distribution allows the noise to be negative and thus runs into the problem of negative noisy weights (Hayes & Kaplan 2023). In our simulations, we sample the noise with the exponential distribution $p(x) = \exp(-x)\mathbb{I}_{[0,+\infty)}(x)$ instead.[4] Thus, the noise is deterministically non-negative and the problem of negative noisy weights is avoided without compromising SHG's typological predictions (Magri & Anttila 2025). In any case, the choice of the distribution is irrelevant for the simulations reported below (we obtained comparable results with gaussian noise).

With these preliminaries in place, we now investigate the behavior of SHG grammars on the cycle in (8). We consider the constraint weights in (18a). Furthermore, we assume that Gen lists both underlying forms in the cycle with the four candidates obtained by changing the underlying values of the features voicing and spread glottis in all possible ways. Thus, the underlying form /dʰa/ is listed with the candidates [da] and [dʰa] in the cycle plus the two candidates [tʰa] and [ta]. Analogously, the underlying form /adʰa/ is listed with the candidates [adʰa] and [ada] in the cycle plus the two candidates [atʰa] and [ata]. The SHG grammar $G_{\mathrm{SHG}}^{\boldsymbol{C}}$ corresponding to these candidates, these constraints, and these weights predicts the probabilities in (18b).

[4] The indicator function $\mathbb{I}_S$ of a set S is equal to one when the argument belongs to S and is equal to zero otherwise.



(a)



(b)

Figure 1: Our estimation of SHG probabilities is sufficiently accurate to detect the failure of the cycle identity in the SHG grammars considered.

(18) (a)
IDENT(voice): 8.6796
IDENT(spread): 5.7517
*[+voice]: 8.1379
*[+spread]: 4.1604
*[+voice, +spread]: 1.8933
*V[−voice]V: 13.6517

(b)
G^{C}_{SHG}([ta] | /d^{h}a/) = .0297 $\quad$ G^{C}_{SHG}([ata] | /ad^{h}a/) = .0
G^{C}_{SHG}([da] | /d^{h}a/) = .2444 $\quad$ G^{C}_{SHG}([ada] | /ad^{h}a/) = .8162
G^{C}_{SHG}([t^{h}a] | /d^{h}a/) = .7128 $\quad$ G^{C}_{SHG}([at^{h}a] | /ad^{h}a/) = .0
G^{C}_{SHG}([d^{h}a] | /d^{h}a/) = .0160 $\quad$ G^{C}_{SHG}([ad^{h}a] | /ad^{h}a/) = .1833

The ratio between the SHG probability of de-aspirating the underlying form /d^{h}a/ divided by the SHG probability of realizing it faithfully is equal to $0.2444/0.0160 = 15.2750$. Analogously, the ratio between the SHG probability of de-aspirating the underlying form /ad^{h}a/ divided by the SHG probability of realizing it faithfully is equal to $0.8162/0.1833 = 4.4528$. The two ratios are therefore different because the latter is almost a fourth of the former, as stated in (19). Figure 1a plots the left- and righthand side ratios (on the vertical axis) as a function of the number N of noisy weight vectors used to estimate the SHG probabilities. This figure shows that the difference between the two rates is real, not an artifact of inaccurate estimation. We conclude that the cycle identity (14) that characterizes harmony-based grammars does not extend to SHG.

(19) $$\underbrace{\frac{G^{C}_{\text{SHG}}([\text{da}] \mid /\text{d}^{\text{h}}\text{a}/)}{G^{C}_{\text{SHG}}([\text{d}^{\text{h}}\text{a}] \mid /\text{d}^{\text{h}}\text{a}/)}}_{15.2750} \neq \underbrace{\frac{G^{C}_{\text{SHG}}([\text{ada}] \mid /\text{ad}^{\text{h}}\text{a}/)}{G^{C}_{\text{SHG}}([\text{ad}^{\text{h}}\text{a}] \mid /\text{ad}^{\text{h}}\text{a}/)}}_{4.4528}$$

A detailed explanation of the SHG simulation results is deferred to a forthcoming longer version of this paper. Here is the intuition briefly. Since the constraint *V[−voice]V has a large weight, the intervocalic devoiced candidates [ata] and [at^{h}a] have no probability mass in (18b), while the non-intervocalic devoiced candidates [ta] and [t^{h}a] have some mass. As a result, when we switch from the non-intervocalic position to the intervocalic position, we need to reallocate probability mass from the devoiced to the non-devoiced candidates. Crucially, SHG reallocates this probability mass differently to the two non-devoiced candidates. Indeed, the probability of the non-devoiced faithful candidate grows more than 11 times from the non-intervocalic position ([d^{h}a]) to the intervocalic position ([ad^{h}a]), because $0.1833/0.0160 = 11.4562$. The probability of the non-devoiced de-aspirated candidate instead grows less than four times from the non-intervocalic position ([da]) to the intervocalic position ([ada]), because $0.8162/0.2444 = 3.3396$.

Next, we investigate the behavior of SHG grammars on the cycle in (9). We restrict ourselves to only the relevant constraints, listed in (20a) with the corresponding weights. Furthermore, we suppose that Gen lists both underlying forms in the cycle with the faithful candidate plus the two candidates obtained through consonant deletion and vowel epenthesis. Thus, the underlying form /CVC/ is listed with the candidates [CV] and [CV.CV] in the cycle plus the faithful candidate [CVC]. Analogously, the underlying form /CCV/ is listed with the two candidates [CV.CV] and [CV] in the cycle plus the faithful candidate [CCV]. The SHG grammar G^{C}_{SHG} corresponding to these candidates, constraints, and weights predicts the probabilities in (20b).

(20) (a) NoCoda: 13.7876
NoCompOn: 4.1995
MaxC: 3.4745
DepV: 5.1756

(b) $G^{C}_{\text{SHG}}(\text{[CV.CV]}\,|\,\text{/CVC/}) = .0888$ $\quad G^{C}_{\text{SHG}}(\text{[CV.CV]}\,|\,\text{/CCV/}) = .0230$
$G^{C}_{\text{SHG}}(\text{[CV]}\,|\,\text{/CVC/}) = .9093$ $\quad G^{C}_{\text{SHG}}(\text{[CV]}\,|\,\text{/CCV/}) = .7495$
$G^{C}_{\text{SHG}}(\text{[CVC]}\,|\,\text{/CVC/}) = .0$ $\quad G^{C}_{\text{SHG}}(\text{[CCV]}\,|\,\text{/CCV/}) = .2301$

The ratio between the SHG probability of repairing the coda of /CVC/ through deletion divided by the SHG probability of repairing it through vowel epenthesis is equal to $0.9093/0.0888 = 10.2398$. Analogously, the ratio between the SHG probability of repairing the complex onset of /CCV/ through deletion divided by the SHG probability of repairing it through vowel epenthesis is equal to $0.7495/0.0230 = 32.5869$. The two ratios are therefore different because the latter is more than three times the former, as stated in (21). Figure 1b shows that the difference between the left- and righthand side ratios is not an artifact of inaccurate estimation of the SHG probabilities. We conclude that the prediction (15) that characterizes harmony-based grammars does not extend to SHG.

(21) $$\underbrace{\frac{G^{C}_{\text{SHG}}(\text{[CV]}\,|\,\text{/CVC/})}{G^{C}_{\text{SHG}}(\text{[CV.CV]}\,|\,\text{/CVC/})}}_{10.2398} \neq \underbrace{\frac{G^{C}_{\text{SHG}}(\text{[CV]}\,|\,\text{/CCV/})}{G^{C}_{\text{SHG}}(\text{[CV.CV]}\,|\,\text{/CCV/})}}_{32.5869}$$

Finally, we investigate the behavior of SHG grammars on the cycle in (10). We restrict ourselves to only the relevant constraints listed in (22a) with the corresponding weights. Furthermore, we assume that Gen lists the three underlying forms in the cycle with the faithful candidate plus the two candidates obtained through consonant epenthesis and vowel deletion. Thus, the underlying form /CVV/ is listed with the epenthetic and deletion candidates [CV.CV] and [CV] in the cycle plus the faithful candidate [CVV]. The underlying form /VCV/ is listed with the deletion and faithful candidates [CV] and [V.CV] in the cycle plus the epenthetic candidate [CV.CV]. Finally, the underlying form /V/ is listed with the faithful and epenthetic candidates [V] and [CV] in the cycle plus the deletion candidate [∅] (the empty string). The SHG grammar G^{C}_{SHG} corresponding to these candidates, constraints, and weights predicts the probabilities in (22b).

(22) (a) Onset: 8.2286 NoHia: 12.6905 MaxV: 8.5195 DepC: 10.1334

(b) $G^{C}_{\text{SHG}}(\text{[CV.CV]}\,|\,\text{/CVV/}) = .0991$ $\quad G^{C}_{\text{SHG}}(\text{[CV]}\,|\,\text{/VCV/}) = .3663$ $\quad G^{C}_{\text{SHG}}(\text{[V]}\,|\,\text{/V/}) = .6213$
$G^{C}_{\text{SHG}}(\text{[CV]}\,|\,\text{/CVV/}) = .9002$ $\quad G^{C}_{\text{SHG}}(\text{[V.CV]}\,|\,\text{/VCV/}) = .6218$ $\quad G^{C}_{\text{SHG}}(\text{[CV]}\,|\,\text{/V/}) = .0097$
$G^{C}_{\text{SHG}}(\text{[CVV]}\,|\,\text{/CVV/}) = .0004$ $\quad G^{C}_{\text{SHG}}(\text{[CV.CV]}\,|\,\text{/VCV/}) = .0093$ $\quad G^{C}_{\text{SHG}}([\emptyset]\,|\,\text{/V/}) = .3691$

As shown in (23), the prediction (16) that characterizes harmony-based grammars does not extend to SHG.

(23) $$\underbrace{\frac{G^{C}_{\text{SHG}}(\text{[CV]}\,|\,\text{/CVV/})}{G^{C}_{\text{SHG}}(\text{[CV.CV]}\,|\,\text{/CVV/})}}_{9.0837} \neq \underbrace{\frac{G^{C}_{\text{SHG}}(\text{[CV]}\,|\,\text{/VCV/})}{G^{C}_{\text{SHG}}(\text{[V.CV]}\,|\,\text{/VCV/})}\frac{G^{C}_{\text{SHG}}(\text{[V]}\,|\,\text{/V/})}{G^{C}_{\text{SHG}}(\text{[CV]}\,|\,\text{/V/})}}_{0.5890\cdot 64.0515=37.7263}$$

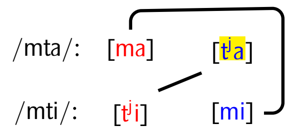
9 Is the behavior of harmony-based grammars phonologically sensible?

Having observed that the cycle identity predicted by harmony-based grammars such as ME does not extend to non-harmony-based grammars such as SHG, the following question now naturally arises: is the cycle identity empirically supported? Or is it possible to construct cycles where the prediction is pathological and can thus be construed as evidence against harmony-basedness? The examples considered so far are of little help in addressing this question. In this final section, we turn to cycles that feature one mapping that is **impossible**, so crazy that no categorical OT or HG grammar chooses it. We speculate that the cycle identity predicted by harmony-based grammars is paradoxical for such cycles.

We consider the markedness constraints NoPalatal that penalizes palatals, *[ti] that penalizes a non-palatalized coronal in front of a high front vowel, and NoNasal that penalizes nasals. Next, we consider the faithfulness constraints IdentPlace that penalizes in particular the palatalization of an underlying coronal, MaxInitial that penalizes the deletion of a word-initial segment (Beckman 1999), and MaxPreV that penalizes the deletion of a consonant followed by a vowel. The constraint identities in (24a) hold for each constraint C considered. The reason is that there is only one constraint *[ti] that is sensitive to the different vowel qualities in the two underlying forms and that constraint only distinguishes between the candidates [ti] and [ta] that do not partake in the cycle. When we draw a line between mappings that share the same constraint violation profile, we obtain the cycle (24b).

(24) (a) C(/mti/, [tʲi]) $=$ C(/mta/, [tʲa])
C(/mta/, [ma]) $=$ C(/mti/, [mi])

(b)



Section 6 tells us that any probabilistic grammar $G_H^{\boldsymbol{C}}$ based on the constraint set $\boldsymbol{C}$ considered through any harmony function H satisfies the cycle identity (25). This identity says that the ratio between the probability of deleting the first consonant and palatalizing the second consonant (numerator) divided by the probability of deleting the second consonant (denominator) for the underlying form /mta/ (lefthand side) is equal to the same probability ratio for the underlying form /mti/ (righthand side).

(25) $$\frac{G_H^{\boldsymbol{C}}(\text{[tʲa]} \mid \text{/mta/})}{G_H^{\boldsymbol{C}}(\text{[ma]} \mid \text{/mta/})} = \frac{G_H^{\boldsymbol{C}}(\text{[tʲi]} \mid \text{/mti/})}{G_H^{\boldsymbol{C}}(\text{[mi]} \mid \text{/mti/})}$$

To understand the implications of this cycle identity (25), we observe that it entails in particular the material implication (26) between probability inequalities. In fact, suppose that the inequality in the antecedent of (26) holds. Equivalently, the lefthand side ratio of (25) is smaller than one. The righthand side ratio must therefore be smaller than one as well. Equivalently, the consequent of (26) must hold as well.

(26) If $G_H^{\boldsymbol{C}}(\text{[tʲa]} \mid \text{/mta/}) < G_H^{\boldsymbol{C}}(\text{[ma]} \mid \text{/mta/})$
then $G_H^{\boldsymbol{C}}(\text{[tʲi]} \mid \text{/mti/}) < G_H^{\boldsymbol{C}}(\text{[mi]} \mid \text{/mti/})$

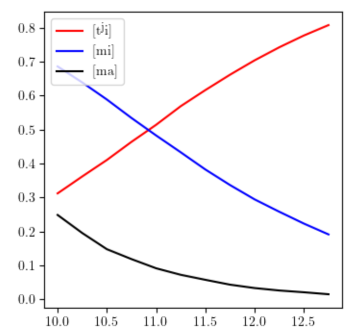


Figure 2: SHG probabilities that contradict the implication (26)

The violation of the faithfulness constraint IDENTPLACE in the mapping (/mta/, [tʲa]) does not yield any improvement in markedness: palatalization is gratuitous in front of a low central vowel. This mapping should therefore be categorically impossible (Moreton 2008). And it is indeed impossible in both categorical OT and HG: given our constraints, no OT or HG grammar is able to palatalize the underlying coronal stop of /mta/. To highlight this fact, this mapping is highlighted in yellow throughout. The other three mappings that appear in (26) are instead sensible and indeed all OT and HG possible.

Since the mapping (/mta/, [tʲa]) is impossible while the mapping (/mta/, [ma]) is decent and indeed OT and HG possible, we expect that the probability of the former cannot be larger than the probability of the latter in any language (or at least in any language where our constraints are the relevant ones). In other words, the antecedent of the implication (26) ought to be always true. It follows that the consequent should be always true as well. Yet, we see no reason why some languages shouldn't flout the consequent of the implication (26) by repairing the underlying cluster of /mti/ as [tʲi] with a larger probability than as [mi].

The problem is that the antecedent of the implication (26) features a candidate [tʲa] that is impossible; hence it is conceivable that the antecedent inequality holds universally. The consequent of (26) instead compares two mappings that are both possible; hence it is conceivable that the consequent inequality holds only for some but not all languages. The material implication (26) predicted by harmony-based grammars is therefore a paradox because its antecedent is tautological while the consequent is contingent.

Indeed, we can construct SHG grammars based on our constraints that defy this paradoxical implication (26). To illustrate, we assume that the two underlying forms /mta/ and /mti/ are listed by Gen with the two candidates in the cycle plus the candidates [ti] and [ta]. Furthermore, we set the weights of the constraints NONASAL, NOPALATAL, and IDENTPLACE equal to zero. We choose a large weight for the constraint *[ti], say equal to 200. Finally, we choose an arbitrary positive weight w for the faithfulness constraint MAXINITIAL (the actual value of this weight is irrelevant for the computation of the SHG probabilities). Figure 2 plots the SHG probabilities (on the vertical axis) of the three mappings (/mta/, [ma]), (/mti/, [tʲi]), and (/mti/, [mi]) compared in (26), as a function of the weight of MaxPrev (plotted on the horizontal axis) that varies between w and $w + 3$ (to draw the plot, we have chosen $w = 10$). The SHG probability of the fourth mapping (/mta/, [tʲa]) is not plotted because it is equal to zero no matter the choice of the weights (because HG impossible mappings always have zero SHG probabilities). When the weight of MaxPrev is larger than $w + 1 = 11$, the SHG probability of the mapping (/mta/, [ma]) is larger than zero whereby the antecedent of the material implication (26) holds. Yet, the SHG probability of the mapping (/mti/, [tʲi]) is larger than that of the mapping (/mti/, [mi]) whereby the consequent of (26) fails.

10 Conclusion

We explored the typological restrictiveness of two types of probabilistic grammars. Our main result is that harmony-based grammars (e.g., Maximum Entropy, ME) are inherently restrictive in a specific sense that non-harmony-based grammars (e.g., Stochastic Harmonic Grammar, SHG) are not. We demonstrated this using the notion of a **cycle** that involves chains of candidates with identical violation profiles across inputs and illustrated the restrictiveness of harmony-basedness with two types of examples. We further showed by counter-example that non-harmony-based grammars such as SHG are not subject to the same restriction and do not make the same predictions. We leave it to further work to adjudicate whether these facts favor ME over SHG. More exploration is needed to decide whether harmony-basedness is a genuine property of natural language or whether non-harmony-based grammars such as SHG are better suited to model some phonological phenomena.

References

Anttila, Arto, Scott Borgeson & Giorgio Magri (2019). Equiprobable mappings in weighted constraint grammars. Nicolai, Garrett & Ryan Cotterell (eds.), *Proceedings of the 16th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, Association for Computational Linguistics, 125–134.

Beckman, Jill N. (1999). *Positional Faithfulness: An Optimality Theretic Treatment of Phonological Asymmetries*. Garland Publishing (Outstanding dissertations in Linguistics), New York.

Boersma, Paul & Joe Pater (2016). Convergence properties of a gradual learning algorithm for Harmonic Grammar. McCarthy, John & Joe Pater (eds.), *Harmonic Grammar and Harmonic Serialism*, Equinox Press, London.

Hayes, Bruce (2022). Deriving the wug-shaped curve: A criterion for assessing formal theories of linguistic variation. *Annual Review of Linguistics* 8, 473–494.

Hayes, Bruce & Aaron Kaplan (2023). Zero-weighted constraints in Noisy Harmonic Grammar. *Linguistic Inquiry* 1–14.

Hayes, Bruce & Colin Wilson (2008). A Maximum Entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39, 379–440.

Labov, William (1969). Contraction, deletion, and inherent variability of the English copula. *Language* 45, 715–762.

Magri, Giorgio (2025). Principles of constraint interaction in probabilistic phonology: Hayes and Zuraw's Shifted Sigmoids Generalization and the axiomatic derivation of OT and HG phonology.

Magri, Giorgio & Arto Anttila (2025). The stochastic generalization and its implications for probabilistic constraint-based phonology.

Moreton, Elliott (2008). Non-computable functions in Optimality Theory. McCarthy, John J. (ed.), *Optimality Theory in Phonology: A reader*, Wiley-Blackwell, Malden: MA, 141–163.

Wilson, Colin (2025). Maximum entropy grammars. Jardine, Adam & Paul de Lacy (eds.), *The Cambridge Handbook of Phonology (2nd edition)*, Cambridge University Press.

Zuraw, Kie & Bruce Hayes (2017). Intersecting constraint families: an argument for Harmonic Grammar. *Language* 93.3, 497–546.