The Implications of Hayes and Zuraw's Shifted Sigmoids Generalization for MaxEnt Phonology

Giorgio Magri

Section 1 reviews a generalization on the rates of application of variable phonological processes recently proposed by Bruce Hayes and Kie Zuraw (HZ; Zuraw & Hayes 2017, Hayes 2022). HZ observe that the popular Maximum Entropy (ME) grammars satisfy this generalization. Section 2 shows that, within harmony-based phonology, there are virtually no other options. In the sense that a harmony-based grammar satisfies HZ's generalization if and only if it is either a ME grammar or else it can be construed as a ME grammar through a transformation of the constraints. This constraint transformation is plausibly innocuous in the daily business of phonological analysis, where constraint violation profiles are usually sparse and only feature small numbers of violations. Section 3 concludes that HZ's generalization provides sort of an axiomatic justification of ME within harmony-based phonology.

1. HZ's Shifted Sigmoids Generalization

HZ's generalization has the form '*if*..., *then*...'. Subsection 1.1 formulates the consequent '*then*...' and subsection 1.2 turns to the antecedent '*if*...'. Subsection 1.3 finally puts the pieces together.

1.1. The consequent of HZ's generalization

We start by reasoning concretely on Tagalog consonant clusters derived through the concatenation of a nasal-final prefix with an obstruent-initial stem. These derived clusters undergo a process of **nasal substitution** that coalesces the nasal and the obstruent into a single consonant that retains the nasality of the former and the place of articulation of the latter: /maŋ+bigáj/ is realized as [mamigáj] 'to distribute' (Zuraw 2010). We focus on two factors that influence this process. One factor is the quality of the stem-initial obstruent. For simplicity, we consider only two options: the voiced labial stop /b/ and the voiceless velar stop /k/. Another factor is the identity of the nasal-final affix. Again, we consider only two options: /paŋ-/ and /maŋ-/. HZ show that the rates of nasal substitution for the four resulting underlying concatenations in the **square** (1) satisfy the following surprising property.

(1)

/maŋb/	/maŋk/
/paŋb/	/paŋk/

The **sigmoid** function $\mathbf{S}(x) = \frac{1}{1+e^{-x}}$ has an S-shaped graph that stretches between 0 and 1 (both excluded). Any rates can thus be plotted on this graph. For instance, figure 1a plots on the graph of the sigmoid the rates 0.916 and 0.993 of nasal substitution reported by Zuraw (2010) for underlying concatenations /maŋb/ and /maŋk/ that share the prefix /maŋ-/. Next, figure 1b plots as two points with the same abscissas the rates 0.434 and 0.909 of nasal substitution reported by Zuraw (2010) for underlying concatenations /paŋb/ and /paŋk/ that share the other prefix /paŋ-/. Figure 1c shows that the latter two points surprisingly sit on the graph of some shifted sigmoid, namely on the graph of the function $\mathbf{S}(x + \delta)$ for some shifting constant δ. In conclusion, HZ have discovered that these Tagalog rates of nasal substitution can be fitted on two shifted sigmoids at shared abscissas. Thus, these four rates can be described through only three parameters: the two shared abscissas x_1, x_2 and the shifting constant δ. Equivalently, whenever we know three of these rates, we can guess the fourth one.

[*] Author affiliations: CNRS, University of Paris 8, UMR 7023.

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

Proceedings of the 42nd West Coast Conference on Formal Linguistics, ed. Shweta Akolkar et al., 232-239. Somerville, MA: Cascadilla Proceedings Project.

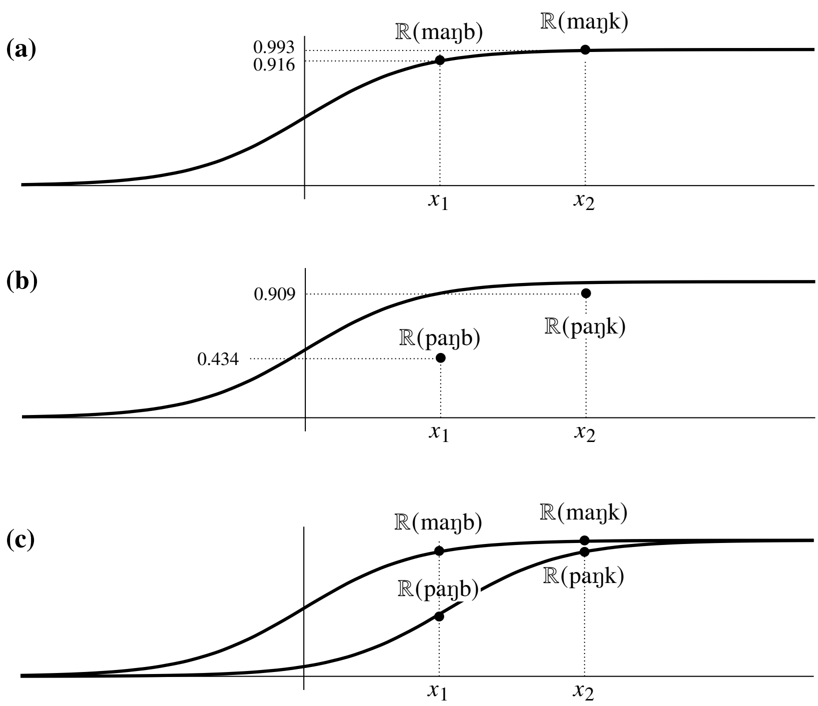

Figure 1: HZ plot the rates of nasal substitution for the four underlying forms (1) on two shifted sigmoids at shared abscissas

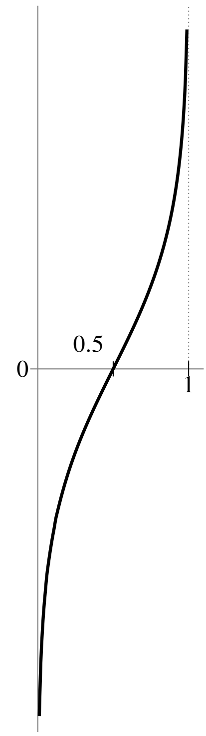

Figure 2: Logit function

Let us take a closer look at HZ's discovery. Figure 1c with the two shifted sigmoids effectively says that the rates $\mathbb{R}(/\text{maŋb}/)$, $\mathbb{R}(/\text{maŋk}/)$, $\mathbb{R}(/\text{paŋb}/)$, and $\mathbb{R}(/\text{paŋk}/)$ of nasal substitution on the four concatenations (1) satisfy the four identities in (2), in terms of the two shared abscissas x_1, x_2, the shifting constant δ, and the sigmoid function $\mathbf{S}(x) = \frac{1}{1+\exp(-x)}$.

(2)
$$\mathbb{R}(/\text{maŋb}/) = \mathbf{S}(x_1) \qquad \mathbb{R}(/\text{maŋk}/) = \mathbf{S}(x_2)$$
$$\mathbb{R}(/\text{paŋb}/) = \mathbf{S}(\delta + x_1) \qquad \mathbb{R}(/\text{paŋk}/) = \mathbf{S}(\delta + x_2)$$

The **logit** function $\mathbf{L}(x) = \log \frac{x}{1-x}$ (with the argument x between zero and one, both excluded) plotted in figure 2 inverts the sigmoid function $\mathbf{S}$: whenever we apply the sigmoid function $\mathbf{S}$ to some number x and then feed the result to the logit function $\mathbf{L}$, we get back the number x we started from, namely $\mathbf{L}(\mathbf{S}(x)) = x$. Thus, by applying the logit function $\mathbf{L}$ at both sides of the four identities in (2), we obtain the expressions in (3) for the logit rates of nasal substitution.

(3)
$$\mathbf{L}\mathbb{R}(/\text{maŋb}/) = x_1 \qquad \mathbf{L}\mathbb{R}(/\text{maŋk}/) = x_2$$
$$\mathbf{L}\mathbb{R}(/\text{paŋb}/) = \delta + x_1 \qquad \mathbf{L}\mathbb{R}(/\text{paŋk}/) = \delta + x_2$$

The difference $\mathbf{L}\mathbb{R}(/\text{maŋb}/) - \mathbf{L}\mathbb{R}(/\text{maŋk}/)$ between the logit rates on the first row of (3) is equal to $x_1 - x_2$. The difference $\mathbf{L}\mathbb{R}(/\text{paŋb}/) - \mathbf{L}\mathbb{R}(/\text{paŋk}/)$ between the logit rates on the second row is the same, because $(\delta + x_1) - (\delta + x_2) = x_1 - x_2$. In other words, the differences along the two rows are identical, as stated in (4a). The differences along the two columns are identical as well, as stated in (4b).[1] Thus, (2) entails (4). The reverse holds as well: (4) ensures that there exist three parameters x_1, x_2, and δ that satisfy (2). We conclude that (4) is an equivalent, more compact formulation of HZ's discovery in figure 1c.

(4) a. $\mathbf{L}\mathbb{R}(/\text{maŋb}/) - \mathbf{L}\mathbb{R}(/\text{maŋk}/) = \mathbf{L}\mathbb{R}(/\text{paŋb}/) - \mathbf{L}\mathbb{R}(/\text{paŋk}/)$

b. $\mathbf{L}\mathbb{R}(/\text{maŋb}/) - \mathbf{L}\mathbb{R}(/\text{paŋb}/) = \mathbf{L}\mathbb{R}(/\text{maŋk}/) - \mathbf{L}\mathbb{R}(/\text{paŋk}/)$

[1] Indeed, (4a) holds if and only if (4b) holds, whereby it suffices to consider only one of these two identities.

The identity (4a) says that the difference $\mathbf{L}\mathbb{R}$(/maŋb/) – $\mathbf{L}\mathbb{R}$(/maŋk/) between the logit rates for concatenations that share the prefix /maŋ-/ is equal to the difference $\mathbf{L}\mathbb{R}$(/paŋb/) – $\mathbf{L}\mathbb{R}$(/paŋk/) between the logit rates for concatenations that share the other prefix /paŋ-/. In other words, the difference between the logit rates for two concatenations that share the same prefix does not depend on the shared prefix. Analogously, the identity (4b) says that the difference between the logit rates for two concatenations that share the same stem-initial obstruent does not depend on that shared obstruent. Magri & Flemming (2024) interpret HZ's observation (4) as an instance of a broader **Constant Difference Generalization** that was formulated by Labov (1969) for the rates of copula deletion in Black American English and that has preoccupied much of the early variationist literature (Kay & McDaniel 1979, Sankoff & Labov 1979).

1.2. The antecedent of HZ's generalization

Why should the differences (4a) between the logit rates of nasal substitution for two concatenations that only differ for the stem-initial obstruent be independent of the choice of the shared prefix? Well, intuitively because the quality of the stem-initial obstruent and the identity of the prefix are **independent** phonological factors. Indeed, the former is purely phonological while the latter pertains to morphology. More generally, the Constant Difference Generalization was interpreted within the variationist literature as the reflex at the level of rates of the independence of the grammatical factors that affect those rates. How can this intuition of independence be cashed out when we switch from a description of rates in terms of variationist factors to their modeling through the constraints of modern constraint-based phonology?

HZ offer the following answer. When modeling the frequencies of nasal substitution within constraint-based phonology, we obviously need some constraints that are sensitive to the quality of the stem-initial obstruent: whether it is voiced or voiceless, its place of articulation, and so on. Furthermore, we also need some constraints that are sensitive to the identity of the prefix, given that concatenations with different prefixes (but the same stem-initial obstruent) undergo nasal substitution with different rates. Yet, we expect no single constraint to be simultaneously sensitive to both the quality of the stem-initial obstruent **and** the identity of the prefix. Thus, no constraint can encode any interaction between those two factors.

To illustrate, we consider the constraint set **C** consisting of the following four constraints (a subset of those considered by Zuraw & Hayes 2017). C_1 = *NC assigns one violation for every sequence of a nasal followed by an obstruent. C_2 = *NC̥ assigns one violation for every sequence of a nasal followed by a voiceless obstruent. C_3 = Uniformity$_{\text{paŋ}}$ assigns one violation when the nasal at the end of the prefix /paŋ-/ coalesces with the obstruent at the beginning of the stem. Finally, C_4 = Uniformity$_{\text{maŋ}}$ is defined analogously, only with the prefix /maŋ-/ in place of /paŋ-/.

Constraints C_1 = *NC and C_2 = *NC̥ are insensitive to the identity of the prefix. They thus satisfy the vertical identities in (5a), namely are constant along the columns of the square (1). Vice versa, constraints C_3 = Unif$_{\text{paŋ}}$ and C_4 = Unif$_{\text{maŋ}}$ are insensitive to the quality of the stem-initial obstruent. They thus satisfy the horizontal identities in (5b), namely are constant along the rows of the square. In (5), we consider the nasal substitution candidates, but analogous identities hold for the candidates to which nasal substitution has not applied. We conclude that prefixes and stems are independent factors relative to this constraint set. In fact, every constraint is constant relative to either the prefixes or the stem-initial obstruents. Thus, no constraint is simultaneously sensitive to both factors and thus no constraint can encode any interaction between the two factors.

(5) a. Constraints insensitive to the prefix and thus constant along the columns:

*NC(/maŋb/, [mam])	*NC(/maŋk/, [maŋ])
‖	‖
*NC(/paŋb/, [pam])	*NC(/paŋk/, [paŋ])

*NC(/maŋb/, [mam])	*NC(/maŋk/, [maŋ])
‖	‖
*NC(/paŋb/, [pam])	*NC(/paŋk/, [paŋ])

b. Constraints insensitive to the stem-initial obstruent and thus constant along the rows:

Uni$_{\text{paŋ}}$(/maŋb/, [mam]) = Uni$_{\text{paŋ}}$(/maŋk/, [maŋ])	Uni$_{\text{maŋ}}$(/maŋb/, [mam]) = Uni$_{\text{maŋ}}$(/maŋk/, [maŋ])
Uni$_{\text{paŋ}}$(/paŋb/, [mam]) = Uni$_{\text{paŋ}}$(/paŋk/, [maŋ])	Uni$_{\text{maŋ}}$(/paŋb/, [mam]) = Uni$_{\text{maŋ}}$(/paŋk/, [maŋ])

1.3. Putting the pieces together: HZ's generalization

We are now ready to put the pieces together into an explicit formulation of HZ's generalization. We focus on a specific variable phonological process, described as the SPE rule A→B/X_Y. We consider four underlying forms x_{TL}, x_{TR}, x_{BL}, and x_{BR} that can be targeted by this process, because they contain one (and only one) instance of the structural description XAY. We arrange these forms into a **square** (6), whereby the subscripts "T" and "B" (for top and bottom) specify the row an underlying form sits in while the subscripts "L" and "R" (for left and right) specify the column. Next, we denote by y_{TL}, y_{TR}, y_{BL}, and y_{BR} the forms obtained by applying the phonological process considered to these four underlying forms (mnemonically, "*y*" is the initial of 'yes', the process has applied); we denote by z_{TL}, z_{TR}, z_{BL}, and z_{BR} the forms obtained when the process does not apply.

(6)
$$\boxed{\begin{array}{cc} x_{\mathrm{TL}} & x_{\mathrm{TR}} \\ x_{\mathrm{BL}} & x_{\mathrm{BR}} \end{array}}$$

We suppose that the "natural" constraint set **C** that describes this phonological process consists of constraints *C* that all satisfy the following disjunction. Either *C* satisfies the vertical identities in the disjunct (7a) because it is is constant along the columns: *C* assigns the same number of violations to the two mappings in the left column as well as to the two mappings in the right column. Or else *C* satisfies the horizontal identities in the disjunct (7b) because it is constant along the rows: *C* assigns the same number of violations to the two mappings in the top row as well as to the two mappings in the bottom row. Thus, no constraint in the set **C** is simultaneously sensitive to the rows **and** the columns, capturing the intuition that rows and columns correspond to independent phonological dimensions.

(7) a. Constant along the columns:

$$\boxed{\begin{array}{cc} C(x_{\mathrm{TL}}, y_{\mathrm{TL}}) & C(x_{\mathrm{TR}}, y_{\mathrm{TR}}) \\ \parallel & \parallel \\ C(x_{\mathrm{BL}}, y_{\mathrm{BL}}) & C(x_{\mathrm{BR}}, y_{\mathrm{BR}}) \end{array}}$$

$$\boxed{\begin{array}{cc} C(x_{\mathrm{TL}}, z_{\mathrm{TL}}) & C(x_{\mathrm{TR}}, z_{\mathrm{TR}}) \\ \parallel & \parallel \\ C(x_{\mathrm{BL}}, z_{\mathrm{BL}}) & C(x_{\mathrm{BR}}, z_{\mathrm{BR}}) \end{array}}$$

b. Constant along the rows:

$$\boxed{\begin{array}{c} C(x_{\mathrm{TL}}, y_{\mathrm{TL}}) = C(x_{\mathrm{TR}}, y_{\mathrm{TR}}) \\ C(x_{\mathrm{BL}}, y_{\mathrm{BL}}) = C(x_{\mathrm{BR}}, y_{\mathrm{BR}}) \end{array}}$$ (mappings where the process has applied)

$$\boxed{\begin{array}{c} C(x_{\mathrm{TL}}, z_{\mathrm{TL}}) = C(x_{\mathrm{TR}}, z_{\mathrm{TR}}) \\ C(x_{\mathrm{BL}}, z_{\mathrm{BL}}) = C(x_{\mathrm{BR}}, z_{\mathrm{BR}}) \end{array}}$$ (mappings where the process has not applied)

Under this assumption that the rows and columns of the square (6) are independent, HZ's generalization predicts that the rates $\mathbb{R}(x_{\mathrm{TL}})$, $\mathbb{R}(x_{\mathrm{TR}})$, $\mathbb{R}(x_{\mathrm{BL}})$, and $\mathbb{R}(x_{\mathrm{BR}})$ of application of the variable process considered to the four underlying forms fit onto two shifted sigmoids at shared abscissas as illustrated in figure 1. Equivalently, the difference between the logit rates corresponding to two underlying forms that sit on the same row is independent of the choice of the row (top versus bottom), as stated in (8a). Equivalently, the difference between the logit rates corresponding to two underlying forms that sit on the same column is independent of the choice of the column (left versus right), as stated in (8b).

(8) a. $\mathbf{L}\mathbb{R}(x_{\mathrm{TL}}) - \mathbf{L}\mathbb{R}(x_{\mathrm{TR}}) = \mathbf{L}\mathbb{R}(x_{\mathrm{BL}}) - \mathbf{L}\mathbb{R}(x_{\mathrm{BR}})$
b. $\mathbf{L}\mathbb{R}(x_{\mathrm{TL}}) - \mathbf{L}\mathbb{R}(x_{\mathrm{BL}}) = \mathbf{L}\mathbb{R}(x_{\mathrm{TR}}) - \mathbf{L}\mathbb{R}(x_{\mathrm{BR}})$

Hayes (2022) reviews a number of variable phonological processes in a number of languages: vowel harmony in Hungarian; consonant liaison in French; inversion of final devoicing in Dutch; genitive plurals in Finnish; schwa/zero alternations in French; stress placement in Hupa. In all these cases, he considers four underlying forms that can be organized into a two-by-two square along two phonological dimensions. He argues that these two dimensions are independent in the sense that no relevant phonological constraint is simultaneously sensitive to both dimensions. And he then shows that the rates with which the variable phonological process considered applies to those four underlying forms can indeed be fitted on two shifted sigmoids at shared abscissas with good approximation. Equivalently, the logit differences relative to one dimension are close to constant relative to the other dimension.

2. Implications of HZ's generalization for ME phonology

Subsection 2.1 recalls that ME grammars satisfy HZ's generalization. Subsection 2.2 presents the main result of the paper, namely a complete characterization of the harmony-based grammars that satisfy HZ's generalization. Subsection 2.3 uses this characterization to argue that ME grammars are "virtually" the only harmony-based grammars that satisfy HZ's generalization.

2.1. ME grammars satisfy HZ's generalization

To construct a ME grammar, we start by listing into *Gen* all the phonological mappings that are relevant for the description of the phonological system of interest. We then devise a finite set **C** of constraints $C_1, \ldots, C_n$ that quantify all the relevant properties of the mapping listed by *Gen*. We denote by $C_k(x, y)$ the number of violations assigned by a constraint C_k to the mapping of an underlying form x to a candidate surface realization y. These numbers are usually collected together into the constraint violation profile specified by a row of a familiar OT tableau. We represent this profile as a constraint violation tuple or vector $\mathbf{C}(x, y) = (C_1(x, y), \ldots, C_n(x, y))$.

The **ME grammar** $G^{\mathrm{ME}}_{\mathbf{w}}$ corresponding to constraint weights $\mathbf{w} = (w_1, \ldots, w_n)$ predicts that an underlying form x is realized as a surface candidate y with a probability $G^{\mathrm{ME}}_{\mathbf{w}}(y \,|\, x)$ proportional to the ME harmony score $H^{\mathrm{HG}}_{\mathbf{w}}(\mathbf{C}(x, y))$ of the constraint violation vector $\mathbf{C}(x, y)$ of the mapping of that underlying form x to that surface realization y, as stated in (9a). This harmony score is computed by applying the exponential to the opposite of the weighted sum of constraint violations, as stated in (9b).

(9) $$G^{\mathrm{ME}}_{\mathbf{w}}(y \,|\, x) \overset{(a)}{\propto} H^{\mathrm{ME}}_{\mathbf{w}}\big(\mathbf{C}(x, y)\big) \overset{(b)}{=} \exp\left\{-\sum_{k=1}^{n} w_k C_k(x, y)\right\}$$

HZ observe that ME grammars satisfy their Shifted Sigmoids Generalization in the following sense. We consider four underlying forms x_{TL}, x_{TR}, x_{BL}, and x_{BR} organized into a two-by-two square (6). We assume that *Gen* assigns them only two candidates: the candidates y_{TL}, y_{TR}, y_{BL}, and y_{BR} obtained by applying the variable phonological process of interest and the candidates obtained when the process has not applied. We consider next a constraint set **C** that ensures that the rows and the columns of the square are independent phonological dimensions because each constraint C in the set **C** is constant along the columns as in (7a) or along the rows as in (7b). ME satisfies HZ's generalization in the sense that, no matter the choice of the constraint weights **w**, the ME grammar $G^{\mathrm{ME}}_{\mathbf{w}}$ corresponding to *Gen* and **C** predicts that the four underlying forms x_{TL}, x_{TR}, x_{BL}, and x_{BR} undergo the phonological process of interest with ME probabilities that satisfy the identities (10) and therefore predict HZ's rate identities (8).

(10) a. $\mathbf{L}G^{\mathrm{ME}}_{\mathbf{w}}(y_{\mathrm{TL}} \,|\, x_{\mathrm{TL}}) - \mathbf{L}G^{\mathrm{ME}}_{\mathbf{w}}(y_{\mathrm{TR}} \,|\, x_{\mathrm{TR}}) = \mathbf{L}G^{\mathrm{ME}}_{\mathbf{w}}(y_{\mathrm{BL}} \,|\, x_{\mathrm{BL}}) - \mathbf{L}G^{\mathrm{ME}}_{\mathbf{w}}(y_{\mathrm{BR}} \,|\, x_{\mathrm{BR}})$

b. $\mathbf{L}G^{\mathrm{ME}}_{\mathbf{w}}(y_{\mathrm{TL}} \,|\, x_{\mathrm{TL}}) - \mathbf{L}G^{\mathrm{ME}}_{\mathbf{w}}(y_{\mathrm{BL}} \,|\, x_{\mathrm{BL}}) = \mathbf{L}G^{\mathrm{ME}}_{\mathbf{w}}(y_{\mathrm{TR}} \,|\, x_{\mathrm{TR}}) - \mathbf{L}G^{\mathrm{ME}}_{\mathbf{w}}(y_{\mathrm{BR}} \,|\, x_{\mathrm{BR}})$

This observation is not surprising. In fact, the variationist literature had already observed that logistic regression (actually, any generalized linear model) predicts the Constant Difference Generalization that differences relative to one factor are constant relative to the other factors (Kay & McDaniel 1979). ME satisfies HZ's generalization because ME is the constraint-based implementation of logistic regression and because HZ's generalization is a special case of the Constant Difference Generalization.

2.2. Characterization of the grammars that satisfy HZ's generalization

ME has played the role of the king in the literature. Yet, besides the ME harmony H^{ME}, we can construct a plethora of other harmony functions H that assign harmony scores $H(\mathbf{C}(x, y))$ to constraint violation vectors. Thus, besides ME grammars, we can construct a plethora of other **harmony-based**

grammars through the position (11), that generalizes (9a) in predicting that the probability that an underlying form x is realized as a surface candidate y is proportional to the harmony score (whatever that is) of the corresponding constraint violation vector $\mathbf{C}(x, y)$. Which of these harmony-based grammars satisfy HZ's generalization as well? In other words, what is crucial in the definition (9b) of ME grammars for them to satisfy the generalization? and what is instead tradable?

(11) $G^H(y \mid x) \;\propto\; H\big(\mathbf{C}(x, y)\big)$

To answer this question, we say that a harmony function H is **separable** provided all harmony scores have the shape (12), where w_k is an arbitrary constraint weight and h_k is a unary function (that is, a function from numbers to numbers). This condition says that the harmony score separates into the product of n functions h_k each of which attends only to the violations assigned by the corresponding constraint C_k. Two constraints C_{k_1} and C_{k_2} thus only interact in a limited and simple way, namely through the product of the values returned by the corresponding functions h_{k_1} and h_{k_2}.

(12) $H\big(\mathbf{C}(x, y)\big) = \prod_{k=1}^{n} h_k^{w_k}\big(C_k(x, y)\big)$

To illustrate, the ME harmony $H_{\mathbf{w}}^{\mathrm{ME}}$ recalled in (13a) can be rewritten as a product as in (13b), because the exponential of a sum is the product of the exponentials. Thus, the ME harmony separates into the product of unary functions all identical to the exponential function $h(x) = \exp(-x)$. In other words, when we switch from the ME harmony $H_{\mathbf{w}}^{\mathrm{ME}}$ to an arbitrary separable harmony we are simply replacing the exponentials in the product that defines the ME harmony with arbitrary unary functions.

(13) $H_{\mathbf{w}}^{\mathrm{ME}}(\mathbf{C}(x, y)) \overset{(a)}{=} \exp\left\{-\sum_{k=1}^{n} w_k C_k(x, y)\right\} \overset{(b)}{=} \prod_{k=1}^{n} \Big(\exp(-C_k(x, y))\Big)^{w_k}$

The main result of this paper boxed below says that a harmony function H encodes a mode of constraint interaction that complies with HZ's generalization if and only if (this is a complete characterization) H is a separable harmony (for a proof, see Magri 2024). Intuitively, this result can be made sense of as follows. The antecedent of HZ's generalization says that the constraints considered are simple. In the sense that they are constant either along the columns as in (7a) or along the rows as in (7b) and thus fail to capture any interaction between these two phonological dimensions. The consequent of HZ's generalization says that the pattern of rates that needs to be predicted is simple as well. In the sense that these rates satisfy the logit identities in (8), whereby they depend on only three parameters and are therefore so tightly connected that we can guess one of the rates from the others. The boxed result captures the intuition that, in order for a simple constraint set to predict a simple pattern of rates, we need a simple mode of constraint interaction. In the sense that the constraints can interact only is a limited way, namely only through the products of a separable harmony function.

> A harmony function H separates into the product (12) of some n unary functions h_k attuned to a single constraint each if and only if the harmony-based grammar G^H corresponding to this harmony H through (11) satisfies HZ's generalization.

2.3. Any separable grammar is a ME grammar up to a constraint transformation

To understand the phonological implications of this boxed result, we make three assumptions on the unary functions h_k that go into the definition (12) of a separable harmony. The first assumption is that each unary function h_k is **strictly positive**. This assumption is needed to ensure that the resulting separable harmony H is strictly positive. The second assumption is that each unary function h_k is **decreasing**: $h_k(x)$ achieves its largest value at $x = 0$ and then gets smaller as its argument x increases away from zero. This assumption (together with the assumption that the weights w_k are non-negative) is needed to ensure that the separable harmony scores sensibly decrease when constraint violations increase. The third assumption is that each unary function h_k is "**normalized**": its largest value taken on at $x = 0$ is equal to one, namely $h_k(0) = 1$. This assumption can be made without loss of generality, because the unary functions h_k and thus the resulting harmony H can be rescaled without affecting the corresponding harmony-based grammar G^H in (11) and thus the probabilistic phonology predicted. Unary functions that are positive, decreasing, and normalized include the exponential function $h(x) = \exp(-x)$; the hyperbolic secant $h(x) = \mathrm{sech}(x) = \frac{2}{\exp(x)+\exp(-x)}$; and the inverse function $h(x) = \frac{1}{1+x}$, plotted in figure 3.

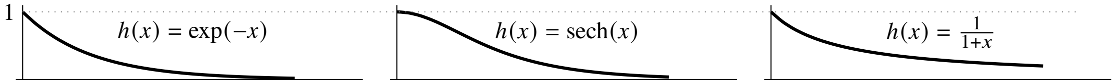

Figure 3: Examples of unary functions that are positive, decreasing, and equal to one at zero.

As observed in (13), ME harmonies separate into the product of unary functions all equal to the exponential function $h(x) = \exp(-x)$. But is there any reason to prefer the ME harmony against alternative separable harmonies? Equivalently, is there any reason to prefer ME's exponential unary function to other unary functions? To reason on this question concretely, let us focus on the harmony that separates into the product of unary functions all equal to the inverse function $h(x) = \frac{1}{1+x}$. I will denote this harmony as $H^{\mathrm{IN}}_{\mathbf{w}}$ and the corresponding harmony-based grammar as $G^{\mathrm{IN}}_{\mathbf{w}}$. So, is there any reason to prefer the ME grammar $G^{\mathrm{ME}}_{\mathbf{w}}$ against $G^{\mathrm{IN}}_{\mathbf{w}}$? The rest of this section argues that there is no reason.

Given some constraints $C_1, \dots, C_n$, we transform them into the new constraints $\widehat{C}_1, \dots, \widehat{C}_n$ through the transformation $\widehat{C}_k = \log(1 + C_k)$. Each transformed constraint $\widehat{C}_k$ takes values that are always non-negative but not always integral. I ignore this issue of non-integrality and thus say that a mapping (x, y) comes with both the vector $\mathbf{C}(x, y)$ of violations assigned by the **original** constraints $C_1, \dots, C_n$ and the vector $\widehat{\mathbf{C}}(x, y)$ of violations assigned by the **transformed** constraints $\widehat{C}_1, \dots, \widehat{C}_n$. Crucially, the harmony $H^{\mathrm{IN}}_{\mathbf{w}}$ based on the inverse unary functions assigns to the original constraint violation vector $\mathbf{C}(x, y)$ the same harmony score that the ME harmony $H^{\mathrm{ME}}_{\mathbf{w}}$ assigns to the transformed constraint violation vector $\widehat{\mathbf{C}}(x, y)$, as stated in (14). In other words, the difference between the two harmony functions is neutralized through our constraint transformation.

(14) $H^{\mathrm{IN}}_{\mathbf{w}}\big(\mathbf{C}(x, y)\big) = H^{\mathrm{ME}}_{\mathbf{w}}\big(\widehat{\mathbf{C}}(x, y)\big)$

The original constraint C_k is satisfied (it assigns zero violations) if and only if the transformed constraint $\widehat{C}_k$ is satisfied as well (it assigns zero violations as well). Furthermore, when the original constraint C_k is not satisfied but assigns only a small number of violations, the transformed constraint $\widehat{C}_k$ assigns a comparable number of violations. To illustrate, when the original constraint C_k assigns 1 or 2 violations, the transformed constraint $\widehat{C}$ assigns 0.7 and 1.1 violations, whereby the transformation effectively amounts to halving the numbers of violations. In the daily practice of phonological analysis, we often restrict ourselves to mappings that have sparse and small constraint violations, namely violate only a few constraints and only by a small amount. In this case, the original constraints $C_1, \dots, C_n$ and the transformed constraints $\widehat{C}_1, \dots, \widehat{C}_n$ are similar, as stated in (15).

(15) $C_1, \dots, C_n \simeq \widehat{C}_1, \dots, \widehat{C}_n$

We are now ready put the pieces together. Because of the identity (14) between the harmony scores assigned by the harmony functions $H^{\mathrm{IN}}_{\mathbf{w}}$ and $H^{\mathrm{ME}}_{\mathbf{w}}$ to the original and the transformed constraints $\mathbf{C}$ and $\widehat{\mathbf{C}}$, the grammar $G^{\mathrm{IN}}_{\mathbf{C},\mathbf{w}}$ based on the original constraint set $\mathbf{C}$ through the harmony function $H^{\mathrm{IN}}_{\mathbf{w}}$ coincides with the ME grammar $G^{\mathrm{ME}}_{\widehat{\mathbf{C}},\mathbf{w}}$ based on the set $\widehat{\mathbf{C}}$ of transformed constraints (and the same weight vector $\mathbf{w}$), as stated in (16a). Furthermore, because of (15), this constraint transformation yields only a small deformation of constraint violation profiles, as long as the original constraint violations are sparse and small. We thus expect the ME grammar $G^{\mathrm{ME}}_{\widehat{\mathbf{C}},\mathbf{w}}$ based on the transformed constraints $\widehat{\mathbf{C}}$ to be very similar to the ME grammar $G^{\mathrm{ME}}_{\mathbf{C},\mathbf{w}}$ based on the original constraints $\mathbf{C}$, as stated in (16b).

(16) $G^{\mathrm{IN}}_{\mathbf{C},\mathbf{w}} \overset{(a)}{=} G^{\mathrm{ME}}_{\widehat{\mathbf{C}},\mathbf{w}} \overset{(b)}{\simeq} G^{\mathrm{ME}}_{\mathbf{C},\mathbf{w}}$

In conclusion, the harmony-based grammar $G^{\mathrm{IN}}_{\mathbf{C},\mathbf{w}}$ based on a constraint set $\mathbf{C}$ through the harmony $H^{\mathrm{IN}}_{\mathbf{w}}$ that separates into the product of inverse unary functions is very close to the ME grammar corresponding to the same constraint set $\mathbf{C}$ (and the same weight vector $\mathbf{w}$). This conclusion easily extends from the harmony function $H^{\mathrm{IN}}_{\mathbf{w}}$ considered so far to any harmony function that separates into the product of any positive, decreasing, and normalized unary functions. In other words, since these unary functions (like those plotted in figure 3) all have a very similar shape, the corresponding separable harmonies can be transformed one into the other through constraint transformations that in practice do not affect much of the phonology encoded into the constraints. It follows that the choice of the ME harmony relative to other separable harmonies is phonologically inconsequential because any grammar G^H based as in (11) on a separable harmony H can be interpreted as a ME grammar through a suitable constraint transformation.

3. Conclusions

HZ's generalization says that the rates of application of a variable phonological process on four underlying forms that differ along two independent phonological dimensions can be fitted on two shifted sigmoids at shared abscissas. Equivalently, the differences between logit rates along one dimension are constant relative to the other dimension. Through this reinterpretation, HZ's generalization can be construed as special case of the Constant Difference Generalization from the early variationist literature.

This paper has investigated the implications of HZ's generalization for the harmony-based implementation of probabilistic constraint-based phonology. The main result has been a complete characterization of the harmony-based grammars that satisfy HZ's generalization as those grammars based on a harmony that separates into the product of unary functions, each attuned to a single constraint. These separable harmonies intuitively capture a particularly simple mode of constraint interaction, as two constraints can only interact through the product of the corresponding unary functions.

ME harmonies are indeed separable. Furthermore, any grammar based on a separable harmony function can be construed as a ME grammar through a suitable constraint transformation. Since the constraint violation profiles in the daily practice of phonological analysis are usually sparse and small, it is reasonable to assume that this transformation does not affect the phonology encoded into the constraints. We conclude that a harmony-based grammar satisfies HZ's generalization if and only if it is either an ME grammar or else it can be construed as an ME grammar through an innocuous constraint transformation. This conclusion construes HZ's generalization as sort of an axiomatic justification of ME within harmony-based phonology.

References

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

Kay, Paul & Chad McDaniel. 1979. On the logic of variable rules. *Language in Society* 8. 151–187.

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

Magri, Giorgio. 2024. Principles of constraint interaction in probabilistic phonology: Hayes and Zuraw's generalization and the principled derivation of MaxEnt grammars. To appear in *Linguistic Inquiry*.

Magri, Giorgio & Edward Flemming. 2024. Variation. In Adam Jardine & Paul de Lacy (eds.), *The cambridge handbook of phonology (2nd edition)*. Cambridge University Press.

Sankoff, David & William Labov. 1979. On the uses of variable rules. *Language in Society* 8. 189–222.

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

Zuraw, Kiw. 2010. A model of lexical variation and the grammar with application to tagalog nasal substitution. *Natural Language and Linguistic Theory* 28. 417–472.

Proceedings of the 42nd West Coast Conference on Formal Linguistics

edited by Shweta Akolkar,
Amber Galvano, Akil Ismael,
Kang Franco Liu, and Line Mikkelsen

Cascadilla Proceedings Project   Somerville, MA   2025

Copyright information

ISBN 978-1-57473-484-3 hardback

Ordering information

Orders for the printed edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Magri, Giorgio. 2025. The Implications of Hayes and Zuraw's Shifted Sigmoids Generalization for MaxEnt Phonology. In *Proceedings of the 42nd West Coast Conference on Formal Linguistics*, ed. Shweta Akolkar et al., 232-239. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #3826.