What do harmony-based grammars exclude?

Antón de la Fuente, Sarang Jeong, Arto Anttila (Stanford), and Giorgio Magri (CNRS)

- Maximum entropy grammars are popular models of the empirical frequency $\mathbb{P}(\mathsf{y}\,|\,\mathsf{x})$ with which an underlying form x (say *cost me*) is realized as a surface candidate y (say *cos'me*, with t-deletion). Maximum entropy phonology makes two assumptions. The **first assumption** is that the frequency $\mathbb{P}(\mathsf{y}\,|\,\mathsf{x})$ is proportional to the harmony score $H(\mathbf{C}(\mathsf{x},\mathsf{y}))$ assigned by some harmony function H to the vector $\mathbf{C}(\mathsf{x},\mathsf{y}) = (C_1(\mathsf{x},\mathsf{y}),\dots,C_n(\mathsf{x},\mathsf{y}))$ of constraint violations $C_k(\mathsf{x},\mathsf{y})$ of the mapping (x,y), as in (1). The **second assumption** is that this harmony function H has a very specific shape, namely it returns the exponential of the opposite of the weighted sum of constraint violations, as in (2). Is the general harmony-based assumption (1) intrinsically restrictive in and of itself without specific assumptions about the harmony H? In other words, given a constraint set $\mathbf{C}$, are there empirical frequency patterns that contradict assumption (1) no matter the choice of the harmony function H?

(1) $\mathbb{P}(\mathsf{y}\,|\,\mathsf{x}) \propto H(\mathbf{C}(\mathsf{x},\mathsf{y}))$

(2) $H = \exp\left\{-\sum_{k=1}^{n} w_k C_k\right\}$

- To start with the simplest case, let us suppose that the constraint set $\mathbf{C}$ does not **compress** mappings: no two candidates of two different underlying forms share the same constraint violation vector. In other words, mappings and their constraint violation vectors are in a one-to-one correspondence. In this case, no matter the frequencies $\mathbb{P}$, we can trivially construct a harmony function H that satisfies (1). Yet, as illustrated below, a reasonable constraint set $\mathbf{C}$ might compress into the same constraint violation vector two mappings that only differ for properties that are not encoded into $\mathbf{C}$ because they are irrelevant relative to the phonological problem at hand. Does such compression pose problems for the harmony-basedness assumption (1)?

- To answer this question, we say that a sequence $\langle \mathsf{x}_1,\mathsf{x}_2,\dots,\mathsf{x}_\ell\rangle$ of underlying forms (not necessarily all distinct) is a **cycle** relative to a constraint set $\mathbf{C}$ provided each underlying form x_i in the sequence shares a constraint violation vector with the following underlying form x_{i+1} and the last underlying form x_ℓ shares a constraint violation vector with the first underlying form x_1. In other words, each underlying form x_i has two different candidates $\mathsf{y}_i, \mathsf{z}_i \in Gen(\mathsf{x}_i)$ that satisfy the identities (3). To illustrate, we consider the constraint set $\mathbf{C}$ consisting of the eight constraints Onset, Coda, MaxC, MaxV, DepC, DepV, CompCoda, CompOnset. A cycle of length $\ell = 4$ is $\langle$/CCV/, /CVC/, /VCC/, /CVC/$\rangle$, as verified in (4). For instance, the two mappings (/CCV/, [CV]) and (/CVC/, [CV]) share the same constraint violation vector because they share the same surface form and violate only one faithfulness constraint DepV.

(3)
$$\begin{aligned}
\mathbf{C}(\mathsf{x}_1,\mathsf{z}_1) &= \mathbf{C}(\mathsf{x}_2,\mathsf{y}_2)\\
\mathbf{C}(\mathsf{x}_2,\mathsf{z}_2) &= \mathbf{C}(\mathsf{x}_3,\mathsf{y}_3)\\
&\vdots\\
\mathbf{C}(\mathsf{x}_{\ell-1},\mathsf{z}_{\ell-1}) &= \mathbf{C}(\mathsf{x}_\ell,\mathsf{y}_\ell)\\
\mathbf{C}(\mathsf{x}_\ell,\mathsf{z}_\ell) &= \mathbf{C}(\mathsf{x}_1,\mathsf{y}_1)
\end{aligned}$$

(4)
$\mathbf{C}$(/CCV/, [CV]) = $\mathbf{C}$(/CVC/, [CV])
$\mathbf{C}$(/CVC/, [VC]) = $\mathbf{C}$(/VCC/, [VC])
$\mathbf{C}$(/VCC/, [CVC]) = $\mathbf{C}$(/CVC/, [CVCC])
$\mathbf{C}$(/CVC/, [CV.CV]) = $\mathbf{C}$(/CCV/, [CV.CV])

- We can now state the result boxed below. The result says that the harmony-basedness assumption (1) really boils down to a prediction about the empirical frequencies on cycles.

> There exists a harmony function H that satisfies the harmony-basedness condition (1) if and only if the empirical frequencies $\mathbb{P}(\mathsf{y}\,|\,\mathsf{x})$ satisfy the identity (5) for every sequence $\langle \mathsf{x}_1,\dots,\mathsf{x}_\ell\rangle$ of underlying forms that is a cycle (3) with respect to the constraint set $\mathbf{C}$.
>
> (5) $\prod_{i=1}^{\ell} \mathbb{P}(\mathsf{y}_i\,|\,\mathsf{x}_i) = \prod_{i=1}^{\ell} \mathbb{P}(\mathsf{z}_i\,|\,\mathsf{x}_i)$

- We can now finally show that the harmony-basedness assumption (1) is restrictive in and of itself without any specific assumptions on the harmony function H. We do this by exhibiting grammars that are (constraint-based but) not harmony-based and that flout this identity (5) that characterizes harmony-basedness. For illustrative purposes, we use Noisy Harmonic Grammar (NHG), which is not harmony-based. We provide two examples summarized in the table on the following page. We discuss the first example in detail; the second example works analogously.

- In our first example, the constraint set $\mathbf{C}$ consists of NoCoda and NoCompOnset together with MaxC and DepV, as in (6). A cycle of length $\ell = 2$ is $\langle$/CCV/, /CVC/$\rangle$, as verified in (7). For this cycle, the product identity (5) predicted by harmony-based phonology (1) boils down to the identity in (8). The left-hand side of this identity is the product $\mathbb{P}$([CV] | /CCV/) · $\mathbb{P}$([CV.CV] | /CVC/) between the probability that the complex onset of /CCV/ is repaired through consonant deletion times the probability that the coda of /CVC/ is repaired through vowel epenthesis. The right-hand side $\mathbb{P}$([CV.CV] | /CCV/) · $\mathbb{P}$([CV] | /CVC/) of the identity is the product of the probabilities of the opposite repairs: the complex onset is repaired through vowel epenthesis and the coda is repaired through consonant deletion. This identity between products would fail in a language that repairs with high probability the onset cluster of /CCV/ through deletion but the coda of /CVC/ through epenthesis because the product on the left-hand side would be larger than the product on the right-hand side. Imagine a language where the English word *proton* would be nativized [ro.to.ni], where C-deletion removes the onset cluster and V-epenthesis removes the coda. Such examples could not be captured in a harmony-based grammar under our constraints.

□ We now want to test whether NHG complies with this identity between products as well. Since we cannot compute exactly the NHG probability of a mapping (x, y) given a weight vector $\boldsymbol{w}$, we estimate it in the usual way: for a certain number N of repetitions, we add Gaussian noise (with standard deviation $\sigma^2 = 1$) to each component of the weight vector $\boldsymbol{w}$; we use the noisy weight vector to compute the categorical HG winner $\widehat{\mathsf{y}}$ for the underlying form x; and we estimate the NHG probability of the mapping (x, y) through the ratio n/N, where n is the total number of times the predicted winner $\widehat{\mathsf{y}}$ is equal to the intended winner y. Obviously, as the number N of repetitions increases, we get more accurate estimates of the NHG probability of the mapping (x, y).

□ We consider the constraint weights in (6) obtained through a random search. The blue curve in (9) plots (on the vertical axis) the product between the NHG probability of the mapping (/CCV/, [CV]) times the NHG probability of the mapping (/CVC/, [CV.CV]) as a function of N (on the horizontal axis). Analogously, the orange curve plots the product of the NHG probabilities of the mappings (/CCV/, [CV.CV]) and (/CVC/, [CV]). The figure thus shows that NHG flouts the identity (5) between the two products because the estimates of the two products converge to different values.

□ In conclusion, we have reported cases where harmony-based models such as ME are more restrictive than NHG, even without making any specific assumptions about the harmony function H.

	first example	**second example**
(6)	NoCoda 2.1412628240990808 MaxC 2.5761596520119734 DepV 3.1380702119334614 NoCompOnset 6.676092855818892	Ident[voice] 8.754864400168293 Ident[spread] 2.7445353035043576 Ident[voice]/_V 8.544706252916026 Ident[spread]/_V 6.460122017813542 *[+voice] 9.506542790584021 *VTV 6.194812329710796 *[+spread] 2.4311096042189595 *[+voice, +spread] 8.287659091134277
(7)	**C**(/CCV/, [CV.CV]) = **C**(/CVC/, [CV.CV]) **C**(/CVC/, [CV]) = **C**(/CCV/, [CV])	**C**(/d^h^a/, [da]) = **C**(/ad^h^a/, [ada]) **C**(/ad^h^a/, [ad^h^a]) = **C**(/d^h^a/, [d^h^a])
(8)	$\mathbb{P}$([CV] | /CCV/) · $\mathbb{P}$([CV.CV] | /CVC/) = = $\mathbb{P}$([CV.CV] | /CCV/) · $\mathbb{P}$([CV] | /CVC/)	$\mathbb{P}$([$\mathsf{d^ha}$] | /$\mathsf{d^ha}$/) · $\mathbb{P}$([ada] | /$\mathsf{ad^ha}$/) = = $\mathbb{P}$([da] | /$\mathsf{d^ha}$/) · $\mathbb{P}$([$\mathsf{ad^ha}$] | /$\mathsf{ad^ha}$/)
(9)