Analyzing Maximum Entropy typologies through implicational universals with negated consequents

Heidi Durresi (MIT) and Giorgio Magri (CNRS)

1 Introduction

Generative phonology has taken a probabilistic turn. At the empirical level, categorical data collected through introspection and field work are now routinely complemented with quantitative data extracted from corpora and experiments. At the theoretical level, the focus is shifting from **categorical** optimality theoretic (OT; Prince & Smolensky 2004) and harmonic grammars (HG; Potts et al. 2010) to **probabilistic** grammars that make quantitative predictions, such as maximum entropy grammars (ME; Wilson 2025) and noisy or stochastic harmonic grammars (SHG; Boersma & Pater 2016). New tools are needed to analyze the linguistic structure encoded by typologies of probabilistic ME or SHG grammars—that is, to do probabilistic phonology with the same theoretical ambition that has characterized categorical generative phonology.

Anttila and Magri (AM; Anttila & Magri 2018; Anttila et al. 2019; Magri & Anttila 2023; Anttila et al. 2024; Magri & Anttila 2025) develop one such tool. They extend the Greenbergian **implicational universals** from the categorical to the probabilistic setting through probability inequalities that must be satisfied by every grammar in the typology considered, as reviewed in section 2. They then use this new tool to show that, while probabilistic SHG typologies inherit all the linguistic structure predicted by the corresponding categorical HG typologies, ME typologies instead encode **little** structure, when structure is assessed through their universals.

Yet, ME typologies must encode **some** linguistic structure. To try to extract and describe this structure, section 3 develops a new class of implicational universals of probabilistic typologies. To preview, AM's implicational universals have the form "if a grammar realizes /VC/ faithfully, it **must** realize /CVC/ faithfully as well", whose consequent is a "positive" injunction. Our new universals have instead the form "if a grammar realizes /VC/ faithfully, it **cannot** neutralize /CVC/ to [CV]", whose consequent is a "negative" injunction. These new negative universals are weaker than AM's positive universals: the negative injunction that /CVC/ not be neutralized to [CV] leaves open the possibility that /CVC/ is nonetheless neutralized to [CV.CV], which is instead ruled out by the positive injunction that /CVC/ be faithfully realized.

Section 4 starts to develop the theory of these new negative universals by addressing the problem of how to concretely compute them for user-specified HG, SHG, and ME typologies. The main result is a complete characterization of the negative universals of ME typologies in terms of a constraint condition that can be checked easily with any linear programming toolkit.[1] As a result, we have extended the software $\mathbb{C}o\mathbb{G}e\mathbb{T}o$ (`https://cogeto.stanford.edu/home`; Magri & Anttila 2019), that computes AM's positive universals for HG, SHG, and ME typologies, with a new subroutine that computes our new negative universals as well.

Building on this computational result, the rest of the paper starts to use these new negative universals to investigate the linguistic structure encoded by ME typologies. More precisely, section 5 formulates the idea of using the weaker negative universals as a tool to diagnose the failure of AM's stronger positive universals for ME typologies. Sections 6 and 7 then illustrate this program with two text cases: basic syllable structure and t-deletion. On both test cases, AM found that the ME typologies miss many of the expected generalizations, when these generalizations are stated as strong positive universals. We show that ME instead misses no generalizations stated in terms of weaker negative universals, at least when we restrict ourselves to

[1] Section 4 is more technical than the rest of the paper and can be skipped at first reading.

Proceedings of *AMP 2023/2024*

phonological mappings that are phonologically sensible, in the sense of being HG possible. We conclude in section 8 that our weaker negative universals might provide a better toolkit for ME phonology.

2 A brief review of AM's positive implicational universals

To set the stage, we start with a brief review of AM's universals. As usual, we assume that Gen lists all the relevant phonological mappings (x, y), each consisting of an underlying form x (such as /CVC/) and a conceivable surface realization y (such as [CV]). A constraint set $\boldsymbol{C}$ lists a finite number n of constraints $C_1, \ldots, C_n$ for these mappings. The number of violations assigned by a constraint C_k to a mapping (x, y) is denoted by $C_k(\mathsf{x}, \mathsf{y})$. These numbers are collected into the constraint violation vector $\boldsymbol{C}(\mathsf{x}, \mathsf{y}) = (C_1(\mathsf{x}, \mathsf{y}), \ldots, C_n(\mathsf{x}, \mathsf{y}))$ of the mapping (x, y). A weight vector $\boldsymbol{w} = (w_1, \ldots, w_n)$ assigns a non-negative weight w_k to each constraint C_k. The relative size of a weight w_k quantifies the relative importance of the corresponding constraint C_k. The product $\boldsymbol{w} \cdot \boldsymbol{C}(\mathsf{x}, \mathsf{y})$ between the weight vector $\boldsymbol{w}$ and the constraint violation vector $\boldsymbol{C}(\mathsf{x}, \mathsf{y})$ is the sum of the products of the corresponding entries, namely $\boldsymbol{w} \cdot \boldsymbol{C}(\mathsf{x}, \mathsf{y}) = \sum_{k=1}^{n} w_k C_k(\mathsf{x}, \mathsf{y})$. The size of this product quantifies how badly the mapping of the underlying form x to the surface candidate y violates the constraints, taking their relative importance into account.

The categorical **HG grammar** $G_{\boldsymbol{w}}$ realizes an underlying form x as the winner surface form $\mathsf{y} = G_{\boldsymbol{w}}(\mathsf{x})$ from the candidate set $Gen(\mathsf{x})$ that violates the constraints the least because the inequality $\boldsymbol{w} \cdot \boldsymbol{C}(\mathsf{x}, \mathsf{y}) < \boldsymbol{w} \cdot \boldsymbol{C}(\mathsf{x}, \mathsf{z})$ holds for any other loser candidate z from $Gen(\mathsf{x})$. The probabilistic **SHG grammar** $G_{\boldsymbol{w}}$ predicts that a given underlying form x is realized as a surface form y from the candidate set $Gen(\mathsf{x})$ with a probability $G_{\boldsymbol{w}}(\mathsf{y} \,|\, \mathsf{x})$ equal to the probability of sampling according to the exponential distribution[2] $p(v_1 \ldots v_n) = \lambda^n \prod_{k=1}^{n} \exp(-\lambda(w_i - v_i)) \mathbb{I}_{[w_i, +\infty)}(v_i)$ a weight vector $\boldsymbol{v} = (v_1, \ldots, v_n)$ such that the corresponding HG grammar $G_{\boldsymbol{v}}$ realizes that underlying form x as that surface form y. Finally, the probabilistic **ME grammar** $G_{\boldsymbol{w}}$ predicts that a given underlying form x is realized as a surface form y from the candidate set $Gen(\mathsf{x})$ with a probability $G_{\boldsymbol{w}}(\mathsf{y} \,|\, \mathsf{x})$ proportional to $\exp(-\boldsymbol{w} \cdot \boldsymbol{C}(\mathsf{x}, \mathsf{y}))$. The categorical HG typology $\mathfrak{T}_{\text{HG}}(Gen, \boldsymbol{C})$ and the probabilistic SHG and ME typologies $\mathfrak{T}_{\text{SHG}}(Gen, \boldsymbol{C})$ and $\mathfrak{T}_{\text{ME}}(Gen, \boldsymbol{C})$ collect together the HG, SHG, and ME grammars $G_{\boldsymbol{w}}$ corresponding to all vectors $\boldsymbol{w}$ of non-negative constraint weights.

How can we study typologies that can be very large (HG) or infinite (SHG, ME)? Here is a natural answer: instead of inspecting the grammars listed by these typologies, we extract the **universals** they predict, namely the properties that hold true of every single grammar in the typology and thus plausibly explain why those grammars have been bundled together into a typology. To implement this intuition, AM focus on implications $(\mathsf{x}, \mathsf{y}) \rightarrow (\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ between two mappings (x, y) and $(\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ from Gen. They say that this implication is a universal of the categorical HG typology $\mathfrak{T}_{\text{HG}}(Gen, \boldsymbol{C})$ provided every HG grammar that realizes the antecedent underlying form x as the antecedent surface form y, also realizes the consequent underlying form $\widehat{\mathsf{x}}$ as the consequent surface form $\widehat{\mathsf{y}}$, whereby the material implication (1) holds universally for any vector $\boldsymbol{w}$ of non-negative constraint weights. To illustrate, (/VC/, [V]) $\rightarrow$ (/CVC/, [CV]) is an HG universal provided any HG grammar that realizes /VC/ as [V] also realizes /CVC/ as [CV].

(1) If $\overbrace{G_{\boldsymbol{w}}(\mathsf{x}) = \mathsf{y}}^{\mathsf{x} \text{ is realized as } \mathsf{y}}$, then also $\underbrace{G_{\boldsymbol{w}}(\widehat{\mathsf{x}}) = \widehat{\mathsf{y}}}_{\widehat{\mathsf{x}} \text{ is realized as } \widehat{\mathsf{y}}}$

(2) $\overbrace{G_{\boldsymbol{w}}(\mathsf{y} \,|\, \mathsf{x})}^{\text{probability that } \mathsf{x} \text{ is realized as } \mathsf{y}} \leq \underbrace{G_{\boldsymbol{w}}(\widehat{\mathsf{y}} \,|\, \widehat{\mathsf{x}})}_{\text{probability that } \widehat{\mathsf{x}} \text{ is realized as } \widehat{\mathsf{y}}}$

How should this definition be extended from a typology of categorical grammars (HG) to a typology of probabilistic grammars (SHG, ME)? AM propose that the implication $(\mathsf{x}, \mathsf{y}) \rightarrow (\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ is a universal of the probabilistic ME typology $\mathfrak{T}_{\text{ME}}(Gen, \boldsymbol{C})$ provided the ME probability $G_{\boldsymbol{w}}(\mathsf{y} \,|\, \mathsf{x})$ of realizing the antecedent underlying form x as the antecedent surface form y is never larger than the ME probability $G_{\boldsymbol{w}}(\widehat{\mathsf{y}} \,|\, \widehat{\mathsf{x}})$ of realizing the consequent underlying form $\widehat{\mathsf{x}}$ as the consequent surface form $\widehat{\mathsf{y}}$, whereby the inequality (2) holds universally for any vector $\boldsymbol{w}$ of non-negative constraint weights. This definition extends unchanged to the SHG typology $\mathfrak{T}_{\text{SHG}}(Gen, \boldsymbol{C})$. Both conditions (1) and (2) formalize the Greenbergian intuition that the consequent mapping $(\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ is easier to get and thus better than the antecedent mapping (x, y). To illustrate,

[2] The SHG literature traditionally uses a gaussian rather than an exponential distribution. Yet, the choice of the gaussian distribution runs into the problem of negative HG weights (Hayes & Kaplan 2023). The choice of the exponential distribution avoids that problem without compromising SHG's typological predictions (Magri & Anttila 2025).

(/VC/, [V]) $\rightarrow$ (/CVC/, [CV]) is a ME or SHG universal provided the ME or SHG probability of realizing /VC/ as [V] is never larger than the probability of realizing /CVC/ as [CV].

AM observe that probabilistic SHG and categorical HG share the same implicational universals: $(\mathsf{x}, \mathsf{y}) \rightarrow (\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ is a universal of the probabilistic SHG typology $\mathfrak{T}_{\text{SHG}}(Gen, \boldsymbol{C})$ in the sense of the universal probability inequality (2) if and only if it is also a universal of the categorical HG typology $\mathfrak{T}_{\text{HG}}(Gen, \boldsymbol{C})$ in the sense of the universal material implication (1). AM show that the situation is different for ME. Indeed, if $(\mathsf{x}, \mathsf{y}) \rightarrow (\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ is a universal of the probabilistic ME typology $\mathfrak{T}_{\text{ME}}(Gen, \boldsymbol{C})$, it is also a universal of the categorical HG typology $\mathfrak{T}_{\text{HG}}(Gen, \boldsymbol{C})$—and therefore also of the probabilistic SHG typology $\mathfrak{T}_{\text{SHG}}(Gen, \boldsymbol{C})$, because HG and SHG share the same universals. Yet, the reverse fails spectacularly. Indeed, AM consider a number of simple constraint sets drawn from the literature and observe that the corresponding ME typologies miss a score of universals that are reasonable and indeed predicted by the corresponding HG/SHG typologies. Sections 6 and 7 below review a couple of AM's examples. The conclusion we draw from ME's misses is that AM's universals are too strong to capture whatever linguistic structure is encoded by ME typologies. New weaker universals are thus needed for ME phonology. The next section makes a concrete proposal.

3 Negative implicational universals

How can AM's universals be weakened? AM's universal (/VC/, [V]) $\rightarrow$ (/CVC/, [CV]) says that, whenever deletion is chosen to repair the antecedent underlying coda of /VC/, deletion **must** also be chosen to repair the consequent underlying coda of /CVC/. This injunction to delete the consequent coda of /CVC/ means that all other options are forbidden: the consequent underlying coda **cannot** be allowed to surface faithfully as in [CVC] and **cannot** be repaired in some different way, say by epenthesizing a vowel as in [CV.CV]. This simple observation suggests to weaken AM's universals by replacing the positive injunction about what must be done to the consequent underlying form with a negative injunction about what cannot be done to it.

To implement this intuition, we focus on implications $(\mathsf{x}, \mathsf{y}) \rightarrow \overline{(\widehat{\mathsf{x}}, \widehat{\mathsf{y}})}$ between some antecedent mapping (x, y) and the negation[3] of some consequent mapping $(\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$. We say that this implication is a universal of the categorical HG typology $\mathfrak{T}_{\text{HG}}(Gen, \boldsymbol{C})$ provided every HG grammar that realizes the antecedent underlying form x as the antecedent surface form y does **not** also realize the consequent underlying form $\widehat{\mathsf{x}}$ as the consequent surface form $\widehat{\mathsf{y}}$, whereby the material implication (3) holds universally for any vector $\boldsymbol{w}$ of non-negative constraint weights. To illustrate, (/VC/, [V]) $\rightarrow$ $\overline{(\text{/CVC/, [CV.CV]})}$ is an HG universal provided no HG grammar repairs the coda of /VC/ through deletion while it repairs the coda of /CVC/ by re-syllabifying it as the onset of an epenthetic vowel. An HG grammar that deletes the antecedent coda but allows the consequent coda to surface faithfully would not count as a counterexample to this specific universal.

(3) $\quad \text{If } G_{\boldsymbol{w}}(\mathsf{x}) \overset{\text{x is realized as y}}{=} \mathsf{y}, \text{ then } G_{\boldsymbol{w}}(\widehat{\mathsf{x}}) \underset{\widehat{\text{x}} \text{ is } \textbf{not} \text{ realized as } \widehat{\text{y}}}{\neq} \widehat{\mathsf{y}}$

(4) $\quad \overset{\text{probability that x is realized as y}}{G_{\boldsymbol{w}}(\mathsf{y} \,|\, \mathsf{x})} \leq \underset{\text{probability that } \widehat{\text{x}} \text{ is } \textbf{not} \text{ realized as } \widehat{\text{y}}}{1 - G_{\boldsymbol{w}}(\widehat{\mathsf{y}} \,|\, \widehat{\mathsf{x}})}$

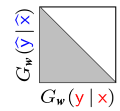
How should this definition (3) be extended from a typology of categorical grammars (HG) to a typology of probabilistic grammars (SHG, ME)? Building on AM, we propose that the implication $(\mathsf{x}, \mathsf{y}) \rightarrow \overline{(\widehat{\mathsf{x}}, \widehat{\mathsf{y}})}$ is a universal of the probabilistic ME typology $\mathfrak{T}_{\text{ME}}(Gen, \boldsymbol{C})$ provided the ME probability $G_{\boldsymbol{w}}(\mathsf{y} \,|\, \mathsf{x})$ of realizing the antecedent underlying form x as the antecedent surface form y is never larger than the probability of **not** realizing the consequent underlying form $\widehat{\mathsf{x}}$ as the consequent surface form $\widehat{\mathsf{y}}$. Since the latter probability of not realizing $\widehat{\mathsf{x}}$ as $\widehat{\mathsf{y}}$ can be expressed as one minus the probability $G_{\boldsymbol{w}}(\widehat{\mathsf{y}} \,|\, \widehat{\mathsf{x}})$ of realizing $\widehat{\mathsf{x}}$ as $\widehat{\mathsf{y}}$, we conclude that the inequality (4) holds universally for any vector $\boldsymbol{w}$ of non-negative constraint weights. This definition extends unchanged to the SHG typology $\mathfrak{T}_{\text{SHG}}(Gen, \boldsymbol{C})$. To illustrate, (/VC/, [V]) $\rightarrow$ $\overline{(\text{/CVC/, [CV.CV]})}$ is a ME or SHG universal provided the ME or SHG probability of realizing /VC/ as [V] through consonant deletion is never larger than the probability of not realizing /CVC/ as [CV.CV] through vowel epenthesis.

From now on, we will say that AM's implicational universals of the form $(\mathsf{x}, \mathsf{y}) \rightarrow (\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ recalled in section 2 are **positive** universals because the consequent is a positive injunction that needs to be satisfied by any grammar that complies with the antecedent. We will say instead that the implicational universals $(\mathsf{x}, \mathsf{y}) \rightarrow \overline{(\widehat{\mathsf{x}}, \widehat{\mathsf{y}})}$ introduced here are **negative** universals because the consequent is a negative injunction.

[3] In some logic formalisms, the negation of a proposition p is indeed notated $\overline{p}$.

The negative universals (contrary to the positive universals) are **symmetrical**: $(\mathsf{x}, \mathsf{y}) \rightarrow \overline{(\widehat{\mathsf{x}}, \widehat{\mathsf{y}})}$ is an HG, ME, or SHG universal if and only if $(\widehat{\mathsf{x}}, \widehat{\mathsf{y}}) \rightarrow \overline{(\mathsf{x}, \mathsf{y})}$ is an HG, ME, or SHG universal as well. Indeed, the material implication (3) that defines HG negative universals says that no non-negative weight vector $\boldsymbol{w}$ satisfies both $G_{\boldsymbol{w}}(\mathsf{x}) = \mathsf{y}$ and $G_{\boldsymbol{w}}(\widehat{\mathsf{x}}) = \widehat{\mathsf{y}}$, which is a symmetric condition in the two mappings. Analogously, the inequality (4) that defines ME and SHG negative universals can be rewritten as the inequality (5) that the sum of the probabilities of the two mappings is never larger than one, which is again a symmetric condition in the two mappings.

(5) $\quad G_{\boldsymbol{w}}(\mathsf{y} \,|\, \mathsf{x}) + G_{\boldsymbol{w}}(\widehat{\mathsf{y}} \,|\, \widehat{\mathsf{x}}) \leq 1$

(6) 

This condition (5) says that, when we plot the probabilities of the two mappings (x, y) and $(\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ as a point in the square (6) of side one, the point must sit in the gray zone underneath the diagonal. This geometric interpretation shows that our new negated universals are weak because the condition (5) that defines them is weak, as it only rules out half of the probability square. For another corollary of this reformulation (5), we recall from AM that the ME and SHG probabilities of an **HG impossible** mapping (a mapping that is not selected by any categorical HG grammar) is never larger than 0.5, no matter the vector $\boldsymbol{w}$ of non-negative constraint weights. Hence, when the two mappings (x, y) and $(\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ are both HG impossible, condition (5) holds universally for any weight vector $\boldsymbol{w}$ and $(\mathsf{x}, \mathsf{y}) \rightarrow \overline{(\widehat{\mathsf{x}}, \widehat{\mathsf{y}})}$ is therefore a negative universal of both ME and SHG. We will use this fact in subsection 6.3 below.

A forthcoming longer version of this paper extends to the negative universals the same relationships that hold among HG, SHG, and ME positive universals recalled at the end of section 2. Indeed, $(\mathsf{x}, \mathsf{y}) \rightarrow \overline{(\widehat{\mathsf{x}}, \widehat{\mathsf{y}})}$ is a negative universal of the probabilistic SHG typology $\mathfrak{T}_{\mathrm{SHG}}(Gen, \boldsymbol{C})$ in the sense of the universal probability inequality (4)/(5) if and only if it is also a negative universal of the categorical HG typology $\mathfrak{T}_{\mathrm{HG}}(Gen, \boldsymbol{C})$ in the sense of the universal material implication (3). Furthermore, if $(\mathsf{x}, \mathsf{y}) \rightarrow \overline{(\widehat{\mathsf{x}}, \widehat{\mathsf{y}})}$ is a negative universal of the probabilistic ME typology $\mathfrak{T}_{\mathrm{ME}}(Gen, \boldsymbol{C})$, it is also a negative universal of the categorical HG typology $\mathfrak{T}_{\mathrm{HG}}(Gen, \boldsymbol{C})$—and therefore also of the probabilistic SHG typology $\mathfrak{T}_{\mathrm{SHG}}(Gen, \boldsymbol{C})$. The reverse can fail, as we will see in section 6 below.

4 Computing the negative implicational universals

How can we check whether some negative universal $(\mathsf{x}, \mathsf{y}) \rightarrow \overline{(\widehat{\mathsf{x}}, \widehat{\mathsf{y}})}$ holds or not? For categorical HG, the answer is straightforward: we can use any linear programming library to find a non-negative weight vector $\boldsymbol{w}$ that contradicts the universal material implication (3) or to certify that no such counterexample weights exist and that the negative HG universal therefore holds. This strategy extends to SHG, because HG and SHG share the same negative universals. The case of ME is different. Because of the non-linearity introduced by the exponential, it is not straightforward to find a non-negative weight vector $\boldsymbol{w}$ that contradicts the universal inequality (4) or to certify that no such counterexample weights exist. This section addresses this problem.

We denote by $\mathsf{z}_1, \dots, \mathsf{z}_m$ the **antecedent losers**, namely the surface forms in the candidate set $Gen(\mathsf{x})$ of the antecedent underlying form x other than the antecedent surface form y, that is instead construed as the **antecedent winner**. When describing one of these antecedent losers z through its constraint violations, it proves useful to **discount** the constraint violations of the antecedent winner y. Thus, we denote by $C_k^{\mathsf{y}}(\mathsf{x}, \mathsf{z})$ the difference $C_k(\mathsf{x}, \mathsf{z}) - C_k(\mathsf{x}, \mathsf{y})$ between the number $C_k(\mathsf{x}, \mathsf{z})$ of violations assigned by a constraint C_k to the antecedent loser z minus the number $C_k(\mathsf{x}, \mathsf{y})$ of violations assigned to the antecedent winner y. We collect these discounted numbers of constraint violations into the vector $\boldsymbol{C}^{\mathsf{y}}(\mathsf{x}, \mathsf{z})$. The **consequent loser** candidates $\widehat{\mathsf{z}}_1, \dots, \widehat{\mathsf{z}}_{\widehat{m}}$ and their vector $\boldsymbol{C}^{\widehat{\mathsf{y}}}(\widehat{\mathsf{x}}, \widehat{\mathsf{z}})$ of discounted constraint violations are defined analogously, only with the consequent mapping $(\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ in place of the antecedent mapping (x, y).

The result boxed below says that condition (7) provides a complete constraint characterization of the ME negative universals $(\mathsf{x}, \mathsf{y}) \rightarrow \overline{(\widehat{\mathsf{x}}, \widehat{\mathsf{y}})}$. When this condition fails, the inequality (8) provides an explicit recipe to construct counterexample non-negative weight vectors $\boldsymbol{w}$, namely weight vectors whose corresponding ME grammar defies the universal considered. The proof of this boxed result starts from a classical result of linear programming duality, namely **Motzkin's Transposition Theorem** (Bertsekas, 2009). And it consists of manipulations analogous to those used by AM, as detailed in a forthcoming longer version of this paper.

An implication $(\mathsf{x},\mathsf{y}) \rightarrow \overline{(\widehat{\mathsf{x}},\widehat{\mathsf{y}})}$ between an antecedent mapping (x,y) listed by Gen with m antecedent losers $\mathsf{z}_1,\ldots,\mathsf{z}_m$ and a negated consequent mapping $(\widehat{\mathsf{x}},\widehat{\mathsf{y}})$ listed with $\widehat{m}$ consequent losers $\widehat{\mathsf{z}}_1,\ldots,\widehat{\mathsf{z}}_{\widehat{m}}$ is a negative universal of the ME typology $\mathfrak{T}_{\mathrm{ME}}(Gen,\boldsymbol{C})$ if and only if there exist some non-negative coefficients $\alpha_1,\ldots,\alpha_m$ and $\beta_1,\ldots,\beta_{\widehat{m}}$ that satisfy condition (7).

(7) a. $\displaystyle\sum_{i=1}^{m}\alpha_i\boldsymbol{C}^{\mathsf{y}}(\mathsf{x},\mathsf{z}_i)+\sum_{j=1}^{\widehat{m}}\beta_j\boldsymbol{C}^{\widehat{\mathsf{y}}}(\widehat{\mathsf{x}},\widehat{\mathsf{z}}_j)\leq\mathbf{0}$ b. $\displaystyle\sum_{i=1}^{m}\alpha_i=\sum_{j=1}^{\widehat{m}}\beta_j=1$

If no such coefficients α_i,β_j exist, some nonnegative weight vector $\boldsymbol{w}$ solves the following inequality

(8) $\displaystyle\boldsymbol{w}\cdot\sum_{i,j}\Big(\boldsymbol{C}^{\mathsf{y}}(\mathsf{x},\mathsf{z}_i)+\boldsymbol{C}^{\widehat{\mathsf{y}}}(\widehat{\mathsf{x}},\widehat{\mathsf{z}}_j)\Big) > -\log m\widehat{m}$

and the ME probability inequality $G_{\boldsymbol{w}}(\mathsf{y}\,|\,\mathsf{x})\leq 1-G(\widehat{\mathsf{y}}\,|\,\widehat{\mathsf{x}})$ in (4) fails for any weight vector $\boldsymbol{w}$ that solves (8), certifying the failure of the negative universal $(\mathsf{x},\mathsf{y})\rightarrow\overline{(\widehat{\mathsf{x}},\widehat{\mathsf{y}})}$ in ME.

A sum of vectors rescaled by non-negative coefficients[4] that add up to one is called a **convex combination** of those vectors. Condition (7) thus says that there exists some convex combination of the vectors $\boldsymbol{C}^{\mathsf{y}}(\mathsf{x},\mathsf{z}_i)$ of discounted constraint violations of the antecedent losers and some convex combination of the vectors $\boldsymbol{C}^{\widehat{\mathsf{y}}}(\widehat{\mathsf{x}},\widehat{\mathsf{z}}_j)$ of discounted violations of the consequent losers that add up to less than the zero vector. This condition (7) is a system of linear inequalities in the unknowns α_i and β_j that can be solved with any library for linear programming.[5] We have thus upgraded the software $\mathbb{C}$o$\mathbb{G}$e$\mathbb{T}$o (`https://cogeto.stanford.edu/home`; Magri & Anttila 2019) with a new subroutine that computes the ME and HG (and thus SHG) negative universals predicted by user-specified candidate and constraint sets.

To illustrate, we consider the antecedent mapping $(\mathsf{x},\mathsf{y})=$ (/VC/, [CV.CV]) with $m=2$ loser candidates $\mathsf{z}_1=$ [CVC] and $\mathsf{z}_2=$ [V] and the consequent mapping $(\widehat{\mathsf{x}},\widehat{\mathsf{y}})=$ (/VCC/, [VC]) with $\widehat{m}=3$ loser candidates $\widehat{\mathsf{z}}_1=$ [CV.CV.CV], $\widehat{\mathsf{z}}_2=$ [CVC], and $\widehat{\mathsf{z}}_3=$ [VC.CV]. We assume that the constraint set $\boldsymbol{C}$ consists of the $n=5$ constraints ONSET, NOCODA, MAX, DEPV, and DEPC. The inequality (7a) thus becomes (9). The latter inequality holds with coefficients $\alpha_1=1/3$, $\alpha_2=2/3$, and $\beta_1=\beta_2=\beta_3=1/3$, which are all non-negative and add up to one as required by (7b). We conclude that (/VC/, [CV.CV]) $\rightarrow$ $\overline{\text{(/VCC/, [VC])}}$ is a ME negative universal because it satisfies the **sufficient** condition (7).

(9) $$\begin{array}{c} \boldsymbol{C}^{\mathsf{y}}(\mathsf{x},\mathsf{z}_1) \\ \alpha_1\begin{bmatrix}0\\1\\0\\-1\\0\end{bmatrix}\end{array} + \begin{array}{c} \boldsymbol{C}^{\mathsf{y}}(\mathsf{x},\mathsf{z}_2) \\ \alpha_2\begin{bmatrix}1\\0\\1\\-1\\-1\end{bmatrix}\end{array} + \begin{array}{c} \boldsymbol{C}^{\widehat{\mathsf{y}}}(\widehat{\mathsf{x}},\widehat{\mathsf{z}}_1) \\ \beta_1\begin{bmatrix}-1\\-1\\-1\\2\\1\end{bmatrix}\end{array} + \begin{array}{c} \boldsymbol{C}^{\widehat{\mathsf{y}}}(\widehat{\mathsf{x}},\widehat{\mathsf{z}}_2) \\ \beta_2\begin{bmatrix}-1\\0\\0\\0\\1\end{bmatrix}\end{array} + \begin{array}{c} \boldsymbol{C}^{\widehat{\mathsf{y}}}(\widehat{\mathsf{x}},\widehat{\mathsf{z}}_3) \\ \beta_3\begin{bmatrix}0\\0\\-1\\1\\0\end{bmatrix}\end{array} \leq \begin{bmatrix}0\\0\\0\\0\\0\end{bmatrix}\begin{array}{l}\textsc{Onset}\\\textsc{Coda}\\\textsc{Max}\\\textsc{DepV}\\\textsc{DepC}\end{array}$$

To illustrate further, we consider the antecedent mapping $(\mathsf{x},\mathsf{y})=$ (/VC/, [CVC]) with $m=3$ loser candidates $\mathsf{z}_1=$ [CV], $\mathsf{z}_2=$ [V], and $\mathsf{z}_3=$ [VC] and the consequent mapping $(\widehat{\mathsf{x}},\widehat{\mathsf{y}})=$ (/V/, [CVC]) with the same $\widehat{m}=3$ loser candidates $\mathsf{z}_1=$ [CV], $\mathsf{z}_1=$ [V], and $\mathsf{z}_1=$ [VC]. We assume the same constraints of the preceding example, although we conflate together DEPV and DEPC into DEP. The inequality (7a) thus becomes (10). The first and third entries (highlighted) corresponding to ONSET and MAX are all non-negative in each vector. The coefficients of the vectors that have positive first or third entry must therefore be equal to zero in order for the sum of the first components and the sum of the third components to be smaller than zero. Thus in particular $\alpha_1=\alpha_2=\alpha_3=0$, whereby condition (7b) fails because $\sum_i\alpha_i\neq 1$. We conclude that (/VC/, [CVC]) $\rightarrow$ $\overline{\text{(/V/, [CVC])}}$ is not a ME negative universal because it flouts the **necessary** condition (7).

[4] Inequalities and the operations of sum and rescaling are all intended to apply to vectors component-wisely.

[5] The recipe (8) to construct counterexample weights is also a linear inequality in the weights, that can therefore be solved with any library for linear programming.

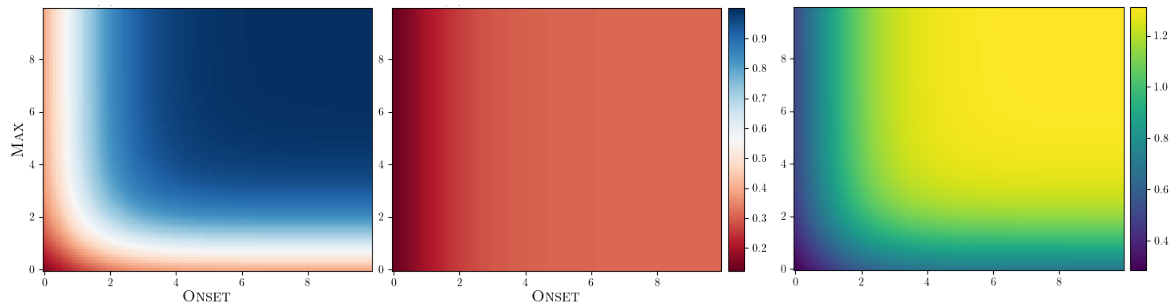


Figure 1: ME probabilities of (/VC/, [CVC]) (left) and (/V/, [CVC]) (center) and their sum (right) when CODA and DEP have weight 0.4 while the weights of ONSET and MAX grow along the horizontal and vertical axes.

(10)
$$\alpha_1 \overset{\boldsymbol{C}^{\mathsf{y}}(\mathsf{x},\mathsf{z}_1)}{\begin{bmatrix}0\\-1\\1\\0\end{bmatrix}} + \alpha_2 \overset{\boldsymbol{C}^{\mathsf{y}}(\mathsf{x},\mathsf{z}_2)}{\begin{bmatrix}1\\-1\\1\\-1\end{bmatrix}} + \alpha_3 \overset{\boldsymbol{C}^{\mathsf{y}}(\mathsf{x},\mathsf{z}_3)}{\begin{bmatrix}1\\0\\0\\-1\end{bmatrix}} + \beta_1 \overset{\boldsymbol{C}^{\widehat{\mathsf{y}}}(\widehat{\mathsf{x}},\widehat{\mathsf{z}}_1)}{\begin{bmatrix}0\\-1\\0\\-1\end{bmatrix}} + \beta_2 \overset{\boldsymbol{C}^{\widehat{\mathsf{y}}}(\widehat{\mathsf{x}},\widehat{\mathsf{z}}_2)}{\begin{bmatrix}1\\-1\\0\\-2\end{bmatrix}} + \beta_3 \overset{\boldsymbol{C}^{\widehat{\mathsf{y}}}(\widehat{\mathsf{x}},\widehat{\mathsf{z}}_3)}{\begin{bmatrix}1\\0\\0\\-1\end{bmatrix}} \leq \begin{bmatrix}0\\0\\0\\0\end{bmatrix}\begin{array}{l}\textsc{Onset}\\\textsc{Coda}\\\textsc{Max}\\\textsc{Dep}\end{array}$$

To appreciate the failure of this ME negative universal, we construct counterexample weights by solving the inequality (8), which becomes (11a) in our case. By computing the sums component-wisely, we obtain (11b). The latter can be rewritten as in (11c) because the product of two vectors is the sum of the products of corresponding components. The resulting inequality holds when the weights of CODA and DEP (that are multiplied by negative coefficients) are small while the weights of ONSET and MAX (that are multiplied by positive coefficients) are large. Thus, we choose a small positive weight for CODA and DEP, say 0.4. Furthermore, we let the weights of ONSET and MAX grow up to 10. Figure 1 shows that the sum of the ME probabilities of the mappings (/VC/, [CVC]) and (/V/, [CVC]) grows larger than one, defying condition (5).

(11)
$$\begin{aligned}-\log m\widehat{m} &\overset{(a)}{<} \boldsymbol{w}\cdot\Bigg(\overset{\boldsymbol{C}^{\mathsf{y}}(\mathsf{x},\mathsf{z}_1)\times\boldsymbol{C}^{\widehat{\mathsf{y}}}(\widehat{\mathsf{x}},\widehat{\mathsf{z}}_1)}{\begin{bmatrix}0\\-2\\1\\-1\end{bmatrix}} + \overset{\boldsymbol{C}^{\mathsf{y}}(\mathsf{x},\mathsf{z}_1)\times\boldsymbol{C}^{\widehat{\mathsf{y}}}(\widehat{\mathsf{x}},\widehat{\mathsf{z}}_2)}{\begin{bmatrix}1\\-2\\1\\-2\end{bmatrix}} + \overset{\boldsymbol{C}^{\mathsf{y}}(\mathsf{x},\mathsf{z}_1)\times\boldsymbol{C}^{\widehat{\mathsf{y}}}(\widehat{\mathsf{x}},\widehat{\mathsf{z}}_3)}{\begin{bmatrix}1\\-1\\1\\-1\end{bmatrix}} + \overset{\boldsymbol{C}^{\mathsf{y}}(\mathsf{x},\mathsf{z}_2)\times\boldsymbol{C}^{\widehat{\mathsf{y}}}(\widehat{\mathsf{x}},\widehat{\mathsf{z}}_1)}{\begin{bmatrix}1\\-2\\1\\-2\end{bmatrix}} + \overset{\boldsymbol{C}^{\mathsf{y}}(\mathsf{x},\mathsf{z}_2)\times\boldsymbol{C}^{\widehat{\mathsf{y}}}(\widehat{\mathsf{x}},\widehat{\mathsf{z}}_2)}{\begin{bmatrix}2\\-2\\1\\-3\end{bmatrix}} + \overset{\boldsymbol{C}^{\mathsf{y}}(\mathsf{x},\mathsf{z}_2)\times\boldsymbol{C}^{\widehat{\mathsf{y}}}(\widehat{\mathsf{x}},\widehat{\mathsf{z}}_3)}{\begin{bmatrix}2\\-1\\1\\-2\end{bmatrix}} + \overset{\boldsymbol{C}^{\mathsf{y}}(\mathsf{x},\mathsf{z}_3)\times\boldsymbol{C}^{\widehat{\mathsf{y}}}(\widehat{\mathsf{x}},\widehat{\mathsf{z}}_1)}{\begin{bmatrix}1\\-1\\0\\-2\end{bmatrix}} + \overset{\boldsymbol{C}^{\mathsf{y}}(\mathsf{x},\mathsf{z}_3)\times\boldsymbol{C}^{\widehat{\mathsf{y}}}(\widehat{\mathsf{x}},\widehat{\mathsf{z}}_2)}{\begin{bmatrix}2\\-1\\0\\-3\end{bmatrix}} + \overset{\boldsymbol{C}^{\mathsf{y}}(\mathsf{x},\mathsf{z}_3)\times\boldsymbol{C}^{\widehat{\mathsf{y}}}(\widehat{\mathsf{x}},\widehat{\mathsf{z}}_3)}{\begin{bmatrix}2\\0\\0\\-2\end{bmatrix}} \Bigg)\\ &\overset{(b)}{=} \boldsymbol{w}\cdot\begin{bmatrix}12\\-12\\6\\-17\end{bmatrix} \overset{(c)}{=} 12w_{\textsc{Onset}} - 12w_{\textsc{Coda}} + 6w_{\textsc{Max}} - 17w_{\textsc{Dep}}\end{aligned}$$

5 How to use the negative universals to diagnose the failure of ME positive universals

Having understood how to compute ME negative universals, we now discuss how they can be used to understand the failure of ME positive universals observed by AM. To this end, we start from one of AM's positive implications (12a), that we call the **parent** of the **children** negative implications (12b) obtained by replacing the consequent winner $\widehat{\mathsf{y}}$ with the negation of the consequent losers $\widehat{\mathsf{z}}_1, \widehat{\mathsf{z}}_2, \ldots$. To illustrate, the positive parent (/VC/, [V]) $\rightarrow$ (/CVC/, [CV]) has negative children such as (/VC/, [V]) $\rightarrow$ $\overline{\text{(/CVC/, [CV.CV])}}$ and (/VC/, [V]) $\rightarrow$ $\overline{\text{(/CVC/, [CVC])}}$ obtained by replacing the consequent winner [CV] with the negation of the consequent losers [CV.CV] and [CVC] (more examples are discussed in sections 6 and 7 below).

(12) a. $(\mathsf{x}, \mathsf{y}) \longrightarrow (\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ b. $(\mathsf{x}, \mathsf{y}) \begin{array}{l} \nearrow \overline{(\widehat{\mathsf{x}}, \widehat{\mathsf{z}}_1)} \\ \longrightarrow \overline{(\widehat{\mathsf{x}}, \widehat{\mathsf{z}}_2)} \\ \searrow \ldots \end{array}$

If the positive parent (12a) is an HG universal, the negative children (12b) are all HG universals as well, under the assumption that HG grammars are **strict** because each selects a unique surface realization per underlying form. In fact, if the positive parent (12a) is an HG universal, every HG grammar that complies with the antecedent mapping (x, y) realizes the consequent underlying form $\widehat{\mathsf{x}}$ as the consequent winner $\widehat{\mathsf{y}}$. The negative children (12b) are then all HG universals as well because no HG grammar that complies with the antecedent mapping (x, y) can realize the consequent underlying form $\widehat{\mathsf{x}}$ as a consequent loser $\widehat{\mathsf{z}}_i$, lest that grammar specifies two surface realizations $\widehat{\mathsf{y}}$ and $\widehat{\mathsf{z}}_i$ for the consequent underlying form $\widehat{\mathsf{x}}$, flouting strictness.

This observation extends from the categorical to the probabilistic setting: if the positive parent (12a) is a ME or SHG universal, the negative children (12b) are all ME or SHG universals as well. In fact, let us suppose that the positive parent (12a) satisfies the probability inequality (2) repeated in (13a) universally for any non-negative weight vector $\boldsymbol{w}$. Since the probabilities of the candidates of the underlying form $\widehat{\mathsf{x}}$ add up to one, then $G_{\boldsymbol{w}}(\widehat{\mathsf{y}} \,|\, \widehat{\mathsf{x}}) + G_{\boldsymbol{w}}(\widehat{\mathsf{z}}_i \,|\, \widehat{\mathsf{x}}) \leq 1$, whereby (13b). The inequality $G_{\boldsymbol{w}}(\mathsf{y} \,|\, \mathsf{x}) \leq 1 - G_{\boldsymbol{w}}(\widehat{\mathsf{z}}_i \,|\, \widehat{\mathsf{x}})$ obtained in (13) thus holds for any weight vector $\boldsymbol{w}$, whereby the negative children in (12b) are all universals.

(13) $$G_{\boldsymbol{w}}(\mathsf{y} \,|\, \mathsf{x}) \overset{(a)}{\leq} G_{\boldsymbol{w}}(\widehat{\mathsf{y}} \,|\, \widehat{\mathsf{x}}) \overset{(b)}{\leq} 1 - G_{\boldsymbol{w}}(\widehat{\mathsf{z}}_i \,|\, \widehat{\mathsf{x}})$$

We now turn to the vice versa: what can be inferred about the positive parent by looking at the negative children? If the negative children (12b) are all HG universals, the positive parent (12a) is an HG universal as well, under the assumption that HG grammars are **total** because each specifies a surface realization for each underlying form. In fact, if the negative children (12b) are all HG universals, no HG grammar that complies with the antecedent mapping (x, y) realizes the consequent underlying form $\widehat{\mathsf{x}}$ as any consequent loser $\widehat{\mathsf{z}}_i$. The positive parent (12a) is then an HG universal because no HG grammar that complies with the antecedent mapping (x, y) can fail to realize the consequent underlying form $\widehat{\mathsf{x}}$ as the consequent winner $\widehat{\mathsf{y}}$, lest that grammar fails to specify a surface realization for the consequent underlying form $\widehat{\mathsf{x}}$, flouting totality.

This observation extends from the categorical to the probabilistic SHG setting simply because HG and SHG share the same positive universals (as recalled in section 2) as well as the same negative universals (as claimed in section 3). The situation for ME is more delicate. To start, let us consider the simplest case where Gen lists the consequent underlying form $\widehat{\mathsf{x}}$ with a unique loser $\widehat{\mathsf{z}}$ besides the winner $\widehat{\mathsf{y}}$. Then, if the negative only child (12b) is a ME universal, the positive parent (12a) is a ME universal as well. In fact, let us suppose that the negative only child (12b) satisfies the probability inequality (4) repeated in (14a) universally for any non-negative weight vector $\boldsymbol{w}$. Since the probabilities of the two candidates $\widehat{\mathsf{y}}$ and $\widehat{\mathsf{z}}$ of the underlying form $\widehat{\mathsf{x}}$ add up to one, then $G_{\boldsymbol{w}}(\widehat{\mathsf{y}} \,|\, \widehat{\mathsf{x}}) + G_{\boldsymbol{w}}(\widehat{\mathsf{z}} \,|\, \widehat{\mathsf{x}}) = 1$, whereby (14b). The inequality $G_{\boldsymbol{w}}(\mathsf{y} \,|\, \mathsf{x}) \leq G_{\boldsymbol{w}}(\widehat{\mathsf{y}} \,|\, \widehat{\mathsf{x}})$ obtained in (14) thus holds for any non-negative weight vector $\boldsymbol{w}$, whereby the positive parent (12a) is a ME universal.

(14) $$G_{\boldsymbol{w}}(\mathsf{y} \,|\, \mathsf{x}) \overset{(a)}{\leq} 1 - G_{\boldsymbol{w}}(\widehat{\mathsf{z}} \,|\, \widehat{\mathsf{x}}) \overset{(b)}{=} G_{\boldsymbol{w}}(\widehat{\mathsf{y}} \,|\, \widehat{\mathsf{x}})$$

This reasoning breaks down when Gen lists more than a single consequent loser candidate: even if all the negative children (12b) corresponding to all the consequent losers $\widehat{\mathsf{z}}_1, \widehat{\mathsf{z}}_2, \ldots$ are ME universals, the positive parent (12a) with the consequent winner $\widehat{\mathsf{y}}$ can fail in ME. We will see some counterexamples in section 6.

We conclude that the new negative universals are indeed weaker than AM's positive universals because the entailment from positive parents to negative children is not reversible in ME. Being weaker, the new negative universals might thus provide a better tool to investigate ME typologies. We can indeed use them as a tool to investigate the failure of the stronger positive universals in ME as follows. We consider each positive parent (12a) that is an HG/SHG universal. If it is also a ME universal, we know that its negative children are ME universals as well, by reasoning as in (13). Hence, we are interested in positive parents (12a) that are HG/SHG universals but not ME universals. What about their negative children (12b)? If these children are all ME universals, they vindicate the failed positive parent: although the ME typology does not encode enough linguistic structure to satisfy the stronger parent positive universal, it does encode some structure after all, which is successfully captured by the weaker children negative universals. If instead only some negative children are ME universals, the children universals that fail might help us to precisely diagnose which aspects of the parent universal turn out to be really problematic for ME. Sections 6 and 7 illustrate this program.

6 Positive and negative ME universals of syllable structure

We suppose that Gen consists of the 36 mappings constructed out of the strings CV, CVC, V, VC, VCV, and CVCV construed as both underlying and candidate surface forms of each other. When VCV and CVCV are construed as surface strings, they are syllabified as [V.CV], and [CV.CV]. Next, we consider the constraint set $\boldsymbol{C}$ consisting of the markedness constraints ONSET and NOCODA against empty onsets and filled codas and the faithfulness constraints MAXC and MAXV against consonant and vowel deletion and DEPC and DEPV against consonant and vowel epenthesis. When evaluating these faithfulness constraints, we assume that onsets and codas are never in correspondence. To illustrate, when /VC/ is realized as [CV], both MAXC and DEPC are violated because the underlying coda is not in correspondence with the surface onset.

AM report that the resulting HG and SHG typologies $\mathfrak{T}_{\mathrm{HG}}(Gen, \boldsymbol{C})$ and $\mathfrak{T}_{\mathrm{SHG}}(Gen, \boldsymbol{C})$ satisfy 744 positive universals. Only 130 of these are also universals of the ME typology $\mathfrak{T}_{\mathrm{ME}}(Gen, \boldsymbol{C})$. AM conclude that ME misses most generalizations of syllable structure, when these generalizations are formulated as positive universals. What about generalizations on syllable structure formulated instead as weaker negative universals? To address this question, we start by discussing implicational universals between two mappings that are both HG possible (subsection 6.1); only later, we bring the HG impossible mappings into the picture, first as consequents (subsections 6.2) and then as antecedents (subsection 6.3). It makes sense to start from the HG possible mappings because those are the mappings sensible enough to be chosen by some HG grammar.

6.1 *Possible antecedents and possible consequents* The HG and SHG typologies satisfy 44 positive universals $(\mathsf{x}, \mathsf{y}) \rightarrow (\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ with an HG possible antecedent mapping (x, y). The consequent mapping $(\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ must be HG possible as well—otherwise the material implication (3) that defines the HG universal would have a contingent antecedent but a contradictory consequent. Of these 44 HG/SHG positive universals, only 13 survive in ME. We refer to AM for a taxonomy of these ME failures. Here, we zoom in on two examples.

As our **first example**, we observe that the HG and SHG typologies satisfy both positive universals in (15). Focusing on SHG for ease of comparison with ME, the universal (15a) says that the SHG probability of faithfully realizing /VC/ is never larger than the SHG probability of faithfully realizing /V/, no matter how we choose the non-negative constraint weights. This universal thus captures the markedness of filled versus empty codas, enshrined into the constraint NOCODA. Analogously, the universal (15b) says that the SHG probability of faithfully realizing /VC/ is never larger than the SHG probability of faithfully realizing /CVC/. This universal thus captures the markedness of empty versus filled onsets, enshrined into the constraint ONSET. Both positive universals fail in ME (as represented by the dotted arrows): we can construct non-negative constraint weights such that the ME probability of faithfully realizing /VC/ is larger than the ME probability of faithfully realizing /V/ or of faithfully realizing /CVC/. In conclusion, the markedness of empty onsets and filled codas is lost in ME when it is formulated in terms of these positive universals.

(15) a. (/VC/, [VC]) $\dashrightarrow$ (/V/, [V]) b. (/VC/, [VC]) $\dashrightarrow$ (/CVC/, [CVC])

Now we reason as in section 5 and look at the children negative universals of these failed parent positive universals. The consequent underlying form /V/ in (15a) is listed by Gen with the faithful consequent winner [V] plus five loser candidates [CV], [CVC], [VC], [V.CV], and [CV.CV]. Because of our initial focus on the HG possible mappings, we only consider the consequent loser [CV], as the other four consequent losers are instead HG impossible. We thus only consider the child negative universal (16a) obtained from the failed parent positive universal (15a) by replacing the consequent winner [V] with the negation of the consequent loser [CV]. Analogously, we consider the two negative children (16b) obtained from the failed parent positive universal (15b) by replacing the consequent winner [CVC] with the negation of the two HG possible consequent losers [CV] and [CV.CV], as the other three consequent losers [V], [VC], and [V.CV] are instead HG impossible. Together, these weaker negative universals (16) say that, if /VC/ is faithfully realized, the onset of /V/ is not filled and the coda of /CVC/ is neither deleted nor re-syllabified as the onset of an epenthetic vowel. Although the parent positive universals in (15) fail in ME, the result boxed in section 4 can be used to easily verify that the three children negative universals in (16) instead all hold in ME.

(16) a. (/VC/, [VC]) $\longrightarrow$ $\overline{(\text{/V/, [CV]})}$ b. (/VC/, [VC]) $\longrightarrow$ $\overline{(\text{/CVC/, [CV]})}$, $\overline{(\text{/CVC/, [CV.CV]})}$

As our **second example**, we observe that the HG and SHG typologies satisfy both positive universals in (17). The universal (17a) compares the two underlying forms /VC/ and /V/ that only differ for their codas and it says that, if the empty onset of the former is filled, the empty onset of the latter is filled as well, because differences in codas do not matter for what happens to onsets. The mirror-image universal (17b) compares the underlying forms /VC/ and /CVC/ that only differ for their onsets and it says that, if the coda of the former is deleted, the coda of the latter is deleted as well, because differences in onsets do not matter for what happens to codas. The ME typology satisfies only the positive universal (17a) but not (17b) (whereby the former is represented by a solid arrow while the latter by a dotted arrow). In conclusion, ME encodes a spurious asymmetry between codas and onsets: differences in codas do not matter for what happens to onsets in ME, because (17a) holds; but differences in onsets do matter for what happens to codas, because (17b) fails.

(17) a. (/VC/, [CV]) $\longrightarrow$ (/V/, [CV]) b. (/VC/, [V]) $\dashrightarrow$ (/CVC/, [CV])

Now we reason again as in section 5 and look at the children negative universals. Because of our initial focus on the HG possible mappings, we only consider the three children negative universals in (18) that feature the negation of the three consequent losers that are HG possible. The negative universal (18a) compares the two underlying forms /VC/ and /V/ that only differ for their codas and it says that, if the empty onset of the former is filled, the empty onset of the latter is not allowed to surface empty. Furthermore, the negative universals (18b) compare the two underlying forms /VC/ and /CVC/ that only differ for their onsets and they say that, if the coda of the former is deleted, the coda of the latter is not allowed to surface faithfully nor to be re-syllabified. The negative universal (18a) holds in ME because it is a child of the parent positive universal (17a) that holds in ME. Furthermore, although the parent positive universal (17b) fails in ME, the result boxed in section 4 can be used to verify that its two children negative universals (18b) both hold in ME. The symmetry between onsets and codas is therefore restored in ME at the level of the negative universals.

(18) a. (/VC/, [CV]) $\longrightarrow$ $\overline{(\text{/V/, [V]})}$

b. (/VC/, [V]) $\longrightarrow$ $\overline{(\text{/CVC/, [CVC]})}$
(/VC/, [V]) $\longrightarrow$ $\overline{(\text{/CVC/, [CV.CV]})}$

These observations extend beyond these two examples (15) and (17). Indeed, let us consider each one of the $44 - 13 = 31$ positive universals $(\mathsf{x}, \mathsf{y}) \rightarrow (\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ that hold in HG/SHG but not in ME and have an HG possible antecedent mapping (x, y). For each of them, we replace the consequent winner $\widehat{\mathsf{y}}$ with the negation of each consequent loser $\widehat{\mathsf{z}}$ that is HG possible. We observe that the children negative universals $(\mathsf{x}, \mathsf{y}) \rightarrow \overline{(\widehat{\mathsf{x}}, \widehat{\mathsf{z}})}$ thus obtained all hold in ME. We conclude that the ME typology captures the expected generalizations about syllable structure, when they are formulated in terms of weaker negative universals, at least when we restrict ourselves to the phonological mappings that are sensible, namely HG possible.

6.2 *Possible antecedents but impossible consequents* Now we progressively bring the HG impossible mappings into the picture. To start, we stick with the $44 - 13 = 31$ positive universals $(\mathsf{x}, \mathsf{y}) \rightarrow (\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ that hold in HG/SHG but not in ME and have an HG possible antecedent mapping (x, y). But we now replace the consequent winner $\widehat{\mathsf{y}}$ with the negation of a consequent loser $\widehat{\mathsf{z}}$ that is HG *im*possible, rather than HG possible. We observe that, for some positive parents $(\mathsf{x}, \mathsf{y}) \rightarrow (\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$, the children negative universals $(\mathsf{x}, \mathsf{y}) \rightarrow \overline{(\widehat{\mathsf{x}}, \widehat{\mathsf{z}})}$ thus obtained all hold in ME. For some other parents instead, the choice of an HG impossible consequent loser $\widehat{\mathsf{z}}$ yields a child negative universal $(\mathsf{x}, \mathsf{y}) \rightarrow \overline{(\widehat{\mathsf{x}}, \widehat{\mathsf{z}})}$ that fails in ME. Let us illustrate both cases.

We consider again the two positive universals in (15). Their children negative universals with HG possible consequent losers have been listed in (16). The children negative universals with HG *im*possible consequent losers are listed in (19). They all hold in ME (as represented by the solid arrows). In conclusion, although the parent positive universals (15) both fail in ME, all their children negative universals hold in ME, irrespectively of whether the negated consequent losers in these negative universals are HG possible or not.

(19) a. (/VC/, [VC]) $\longrightarrow$ $\overline{(\text{/V/, [CVC]})}$
(/VC/, [VC]) $\longrightarrow$ $\overline{(\text{/V/, [VC]})}$
(/VC/, [VC]) $\longrightarrow$ $\overline{(\text{/V/, [V.CV]})}$
(/VC/, [VC]) $\longrightarrow$ $\overline{(\text{/V/, [CV.CV]})}$

b. (/VC/, [VC]) $\longrightarrow$ $\overline{(\text{/CVC/, [V]})}$
(/VC/, [VC]) $\longrightarrow$ $\overline{(\text{/CVC/, [VC]})}$
(/VC/, [VC]) $\longrightarrow$ $\overline{(\text{/CVC/, [V.CV]})}$

The situation is different for the positive universals in (17). Their children negative universals with HG possible consequent losers have been listed in (18). The children negative universals with HG *im*possible consequent losers are listed in (20). The negative universals in (20a) all hold in ME (as represented by the solid arrows), because they are children of the parent positive universal (17a) which holds in ME (as represented by the solid arrow). Of the negative universals in (20b) instead, only two hold in ME while the one represented by the dotted arrow fails. We can perhaps make sense of this failure as follows. The consequent mapping (/CVC/, [V]) is HG impossible because of the gratuitous and thus impossible onset deletion. The HG impossibility of the consequent mapping suffices for this dotted negative universal to hold in HG: the material implication (3) that defines HG universals holds trivially because of a tautological consequent. But if we ignore the onset, this dotted negative universal that fails in ME indeed makes little sense because it provides contradictory instructions for the codas: the antecedent says that the coda of /VC/ is deleted while the negated consequent says that the coda of /CVC/ is not deleted!

(20) a. (/VC/, [CV]) $\to$ $\overline{\text{(/V/, [CVC])}}$
$\to$ $\overline{\text{(/V/, [VC])}}$
$\to$ $\overline{\text{(/V/, [V.CV])}}$
$\to$ $\overline{\text{(/V/, [CV.CV])}}$

b. (/VC/, [V]) $\dashrightarrow$ $\overline{\text{(/CVC/, [V])}}$
$\to$ $\overline{\text{(/CVC/, [VC])}}$
$\to$ $\overline{\text{(/CVC/, [V.CV])}}$

6.3 *Impossible antecedents but possible consequents* So far, we have focused on the 44 HG/SHG positive universals $(x, y) \to (\widehat{x}, \widehat{y})$ with an HG possible antecedent mapping (x, y). Now we turn to the remaining 700 HG/SHG positive universals with an HG *im*possible antecedent (x, y). The large number of these universals is due to the fact that the HG impossibility of the antecedent mapping (x, y) in and by itself suffices to yield a valid HG positive universal $(x, y) \to (\widehat{x}, \widehat{y})$, no matter the consequent mapping $(\widehat{x}, \widehat{y})$: the material implication (1) that defines the HG universal holds trivially because of a false antecedent. Yet, AM observe that 583 of these 700 HG/SHG positive universals with HG impossible mappings fail in ME.

Again as in section 5, we thus turn from these parent positive universals $(x, y) \to (\widehat{x}, \widehat{y})$ to their children negative universals $(x, y) \to \overline{(\widehat{x}, \widehat{z})}$ obtained by replacing the consequent winner $\widehat{y}$ with the negation of a consequent loser $\widehat{z}$. Since the antecedent mapping (x, y) is HG impossible, the negative universal $(x, y) \to \overline{(\widehat{x}, \widehat{z})}$ always holds in ME when the consequent mapping $(\widehat{x}, \widehat{z})$ is HG impossible as well, because negative universals between two HG impossible mappings always hold in ME, as observed in section 3. We therefore restrict ourselves to negative children $(x, y) \to \overline{(\widehat{x}, \widehat{z})}$ obtained by replacing the consequent winner $\widehat{y}$ with the negation of a consequent loser $\widehat{z}$ that is HG possible. Again as in subsection 6.2, we observe that in some cases, the negative children $(x, y) \to \overline{(\widehat{x}, \widehat{z})}$ thus obtained all hold in ME. In some other cases instead, some HG possible consequent loser $\widehat{z}$ yields a child negative universal $(x, y) \to \overline{(\widehat{x}, \widehat{z})}$ that fails in ME.

To illustrate, the positive universal (21a) fails in ME despite the HG impossible mapping (/CV/, [VC]). The consequent mapping (/VCV/, [CV.CV]) has five loser candidates [CV], [V.CV], [V], [CVC], and [VC]. We only consider the first two of these consequent losers, because they are the only two that are HG possible. By replacing the consequent winner [CV.CV] with these two HG possible consequent losers [CV] and [V.CV], we obtain the two children negative universals (21b) that both hold in ME (as represented by the solid arrows). In conclusion, all the children negative universals of the parent positive universal (21a) hold in ME.

(21) a. (/CV/, [VC]) $\dashrightarrow$ (/VCV/, [CV.CV])

b. (/CV/, [VC]) $\to$ $\overline{\text{(/VCV/, [CV])}}$
$\to$ $\overline{\text{(/VCV/, [V.CV])}}$

We consider next the positive universal (22a), that fails in ME despite the HG impossible mapping (/CVCV/, [CVC]). The consequent mapping (/CV/, [V.CV]) has five loser candidates [CV], [CVC], [V], [VC], and [CV.CV] . We only consider the first of these consequent losers, because it is the only one that is HG possible. By replacing the consequent winner [V.CV] with this HG possible consequent loser [CV], we obtain the child negative universal (22b) that fails in ME (as represented by the dotted arrow). This makes sense phonologically: the negation of the consequent mapping (/CV/, [CV]) is expected to have a low probability because it is a faithful mapping with an unmarked output ([CV]). In conclusion, in the case of the positive parent (22a), only the HG impossible consequent losers yield children negative universals that hold in ME.

(22) a. $(\text{/CVCV/}, [\text{CVC}]) \dashrightarrow (\text{/CV/}, [\text{V.CV}])$ b. $(\text{/CVCV/}, [\text{CVC}]) \dashrightarrow \overline{(\text{/CV/}, [\text{CV}])}$

To summarize, we have seen that ME does indeed capture many generalizations about syllable structure when they are formulated as weaker negative (rather than stronger positive) universals and restricted to the HG possible mappings. For the negative universals that have HG impossible antecedents or consequents and fail in ME, we can oftentimes find phonological motivations for their failure, unlike the positive universals.

7 Positive and negative ME universals of t-deletion

We move on to a second test case, concerning the classical variable process of t-deletion. We assume the set of underlying and surface candidate forms listed in figure 2 (left). Furthermore, we use the following constraints (from Kiparsky 1994; see Coetzee & Kawahara 2013 for an alternative): ONSET, MAX, NOCOMPEDGE (one violation for every complex onset or complex coda), ALIGNLEFTWORD (one violation for every syllable that straddles a word boundary), and ALIGNRIGHTPHRASE (one violation for every deletion of a phrase-final segment). For reasons of space, we limit ourselves to HG possible mappings.

The arrow (23a) is solid because ME captures the generalization that pre-vocalic deletion has a lower probability than pre-consonantal deletion. The arrow (23b) is dotted instead because ME fails to capture the generalization that also pre-pausal deletion has lower probability than pre-consonantal deletion. Yet, both generalizations are equally borne out by the empirical rates of t-deletion reported in figure 2 (right) for a variety of English dialects (Coetzee 2004). These parent positive universals (23) have the negative children universals (24) that both hold in ME (as represented by the solid arrows). We conclude that the generalizations on t-deletion are rescued in ME when formulated as negative rather than positive universals.

(23) $(\text{/cost\#us/}, [\text{cos.us}]) \xrightarrow{a} (\text{/cost\#me/}, [\text{cos.me}])$
$(\text{/cost\#\#/}, [\text{cos}]) \overset{b}{\dashrightarrow} (\text{/cost\#me/}, [\text{cos.me}])$

(24) $(\text{/cost\#us/}, [\text{cos.us}]) \xrightarrow{a} \overline{(\text{/cost\#me/}, [\text{cost.me}])}$
$(\text{/cost\#\#/}, [\text{cos}]) \xrightarrow{b} \overline{(\text{/cost\#me/}, [\text{cost.me}])}$

8 Conclusions

AM extend the Greenbergian implicational universals from categorical to probabilistic phonology, as recalled in section 2. Their universals $(\mathsf{x}, \mathsf{y}) \rightarrow (\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ hold of a probabilistic SHG or ME typology provided every SHG or ME grammar realizes the antecedent underlying form x as the antecedent surface form y with a probability never larger than the probability with which it realizes the consequent underlying form $\widehat{\mathsf{x}}$ as the consequent surface form $\widehat{\mathsf{y}}$. AM argue that the resulting universals are well suited to the analysis of SHG typologies but are too strong to extract much structure from probabilistic ME typologies.

Section 3 has thus introduced new universals $(\mathsf{x}, \mathsf{y}) \rightarrow \overline{(\widehat{\mathsf{x}}, \widehat{\mathsf{y}})}$ with a **negated** consequent mapping. This universal holds of an SHG or ME typology provided every SHG or ME grammar realizes the antecedent underlying form x as the antecedent surface form y with a probability never larger than the probability with which it does *not* realize the consequent underlying form $\widehat{\mathsf{x}}$ as the consequent surface form $\widehat{\mathsf{y}}$. Section 4 has developed a complete constraint characterization of these new negative universals in ME. This condition can be checked straightforwardly and has therefore been implemented into the software $\mathbb{C}o\mathbb{G}e\mathbb{T}o$ (`https://cogeto.stanford.edu/home`), that now computes both AM's positive and our new negative universals for OT, HG, SHG, and ME typologies corresponding to user-specified candidates and constraints.

Section 5 has sketched the general idea of using the new weaker negative universals to investigate the failure of AM's stronger positive universals in ME. For each **parent** positive universal $(\mathsf{x}, \mathsf{y}) \rightarrow (\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ that holds in SHG but fails in ME, we consider all **children** negative universals $(\mathsf{x}, \mathsf{y}) \rightarrow \overline{(\widehat{\mathsf{x}}, \widehat{\mathsf{z}})}$ obtained by replacing the consequent winner $\widehat{\mathsf{y}}$ with all consequent losers $\widehat{\mathsf{z}}$. If these children universals all hold in ME, they vindicate the failure of the parent. If some of them fails, these failures might help us pinpoint the exact reason for the parent's failure. Sections 6 and 7 have illustrated this program with two test cases concerning syllables and t-deletion. Our findings on these two test cases can be summarized into two observations.

We start by focusing on mappings that are all HG possible, namely sensible enough that some categorical HG grammars select them. Our **first observation** is that, for each parent positive universal $(\mathsf{x}, \mathsf{y}) \rightarrow (\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ with an HG possible antecedent mapping (x, y) and for each HG possible consequent loser mapping $(\widehat{\mathsf{x}}, \widehat{\mathsf{z}})$, the negative child $(\mathsf{x}, \mathsf{y}) \rightarrow \overline{(\widehat{\mathsf{x}}, \widehat{\mathsf{z}})}$ is indeed a negative universal of ME. In other words, when we restrict

(/cost#me/, [cos.me])	(/cost##/, [cos])	(/cost#us/, [cos.us])
(/cost#me/, [cost.me])	(/cost##/, [cost])	(/cost#us/, [cost.us])
(/cost#me/, [cos.tme])		(/cost#us/, [cos.tus])

	_C	_#	_V
AAVE, Washington, DC	.76	.29	.73
Jamaican English	.85	.63	.71
Tejano English	.62	.25	.46
Trinidadian English	.81	.21	.31
Chicano English	.62	.45	.37
Columbus English	.80	.76	.63

Figure 2: Underlying and candidate forms for the analysis of t-deletion (left) and empirical deletion rates.

ourselves to HG possible mappings, the new negative universals vindicate the failure of the positive parents in ME. ME captures the expected phonological generalizations about syllable structure and t-deletion when these generalizations are formulated as weaker negative universals.

When instead the antecedent mapping (x, y) is HG impossible or the consequent loser mapping $(\widehat{\mathsf{x}}, \widehat{\mathsf{z}})$ is HG impossible, our **second observation** is that the child negative universal $(\mathsf{x}, \mathsf{y}) \rightarrow \overline{(\widehat{\mathsf{x}}, \widehat{\mathsf{z}})}$ can fail in ME. Yet, although AM argue that the failures of the ME positive universals are phonologically paradoxical, we have oftentimes been able to find phonological motivations for the failures of the negative children with impossible mappings, as discussed above for the dotted arrows in (20b) and (22b). We conclude that our new negative implicational universals might provide a better tool to analyze ME typologies.

References

Anttila, Arto & Giorgio Magri (2018). Does MaxEnt overgenerate? Implicational universals in Maximum Entropy grammar. Gallagher, Gillian, Maria Gouskova & Yin Sora (eds.), *AMP 2017: Proceedings of the 2017 Annual Meeting on Phonology*, Linguistic Society of America, Washington, DC.

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.

Anttila, Arto, Kushal Thaman, Emiyare Cyril Ikwut-Ukwa, Annalisa Welinder & Giorgio Magri (2024). Probabilistic syllable structure. *WCCFL 42: Proceedings of the 42nd meeting of the West Coast Conference in Linguistics.*

Bertsekas, Dimitri P. (2009). *Convex Optimization Theory*. Athena Scientific, Belmont, MA, USA.

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.

Coetzee, Andries W. (2004). *What it Means to be a Loser: Non-Optimal Candidates in Optimality Theory*. Ph.D. thesis, University of Massachusetts, Amherst.

Coetzee, Andries W. & Shigeto Kawahara (2013). Frequency biases in phonological variation. *Natural Language and Linguistic Theory* 31:1, 47–89.

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

Kiparsky, Paul (1994). An OT perspective on phonological variation, URL `http://www.stanford.edu/~kiparsky/Papers/nwave94`. Paper presented at NWAVE 23, Stanford University.

Magri, Giorgio & Arto Anttila (2019). ℂ𝕠𝔾𝕖𝕋𝕠: Convex geometry tools for typological analysis in categorical and probabilistic constraint-based phonology (version 1.0). Available at `https://cogeto.stanford.edu`.

Magri, Giorgio & Arto Anttila (2023). Paradoxes of MaxEnt markedness. Elkins, Noah, Bruce Hayes, Jinyoung Jo & Jian-Leat Siah (eds.), *AMP 2022: Supplemental Proceedings of the 2022 Annual Meeting on Phonology*, Linguistic Society of America.

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

Potts, Christopher, Joe Pater, Karen Jesney, Rajesh Bhatt & Michael Becker (2010). Harmonic Grammar with linear programming: from linear systems to linguistic typology. *Phonology* 27(1), 1–41.

Prince, Alan & Paul Smolensky (2004). *Optimality Theory: constraint interaction in generative grammar*. Blackwell, Oxford.

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