Sensitivity to string length and feature count subverts MaxEnt universals

Giorgio Magri (CNRS, SFL, Paris 8) and Arto Anttila (Stanford)

MFM 2022

Manchester | 25-57 May 2022
Stochastic (noisy) HG (SHG) and MaxEnt (ME) are often considered minor variants of the weighted formalism that make comparable phonological predictions [Alderete & Finley 2021].

We show this is not so:
- many reasonable implicational universals hold in SHG but fail in ME
- because ME is paradoxically sensitive to spurious properties
  - sheer string length
  - sheer number of occurrences of a marked feature value
Stochastic (noisy) HG (SHG) and MaxEnt (ME) are often considered minor variants of the weighted formalism that make comparable phonological predictions [Alderete & Finley 2021].

We show this is not so:

- many reasonable implicational universals hold in SHG but fail in ME
- because ME is paradoxically sensitive to spurious properties
  - sheer string length
  - sheer number of occurrences of a marked feature value
In order to develop our argument, we look at **implicational universals** that compare two phonological mappings

\[(x, y) \rightarrow (\hat{x}, \hat{y})\]

Intuitively, this implication says that the consequent mapping \((\hat{x}, \hat{y})\) is “better” than the antecedent mapping \((x, y)\).

Formally, we say that this implication is a universal of a typology of probabilistic phonological grammars provided every single grammar in the typology assigns more probability to the better consequent mapping \((\hat{x}, \hat{y})\) than to the worse antecedent \((x, y)\).

For instance, the implication

\[(/cost+us/, [cos.us]) \rightarrow (/cost+me/, [cos.me])\]

says that *t*-deletion always has a larger probability before consonants than before vowels [Coetzee & Kawahara 2013].
In order to develop our argument, we look at **implicational universals** that compare two phonological mappings 

\[(x, y) \rightarrow (\hat{x}, \hat{y})\]

Intuitively, this implication says that the consequent mapping \((\hat{x}, \hat{y})\) is “better” than the antecedent mapping \((x, y)\).

Formally, we say that this implication is a universal of a typology of probabilistic phonological grammars provided every single grammar in the typology assigns more probability to the better consequent mapping \((\hat{x}, \hat{y})\) than to the worse antecedent \((x, y)\).

For instance, the implication 

\[(/cost+us/, [cos.us]) \rightarrow (/cost+me/, [cos.me])\]

says that *t*-deletion always has a larger probability before consonants than before vowels

[Coetzee & Kawahara 2013]
In order to develop our argument, we look at **implicational universals** that compare two phonological mappings

$$(x, y) \rightarrow (\hat{x}, \hat{y})$$

Intuitively, this implication says that the consequent mapping $$(\hat{x}, \hat{y})$$ is “better” than the antecedent mapping $$(x, y)$$.

Formally, we say that this implication is a universal of a typology of probabilistic phonological grammars provided every single grammar in the typology assigns more probability to the better consequent mapping $$(\hat{x}, \hat{y})$$ than to the worse antecedent $$(x, y)$$.

For instance, the implication

$$(/\text{cost+us}/, [\text{cos.us}]) \rightarrow (/\text{cost+me}/, [\text{cos.me}])$$

says that $t$-deletion always has a larger probability before consonants than before vowels [Coetzee & Kawahara 2013]
In order to develop our argument, we look at implicational universals that compare two phonological mappings

\[(x, y) \rightarrow (\hat{x}, \hat{y})\]

Intuitively, this implication says that the consequent mapping \((\hat{x}, \hat{y})\) is “better” than the antecedent mapping \((x, y)\).

Formally, we say that this implication is a universal of a typology of probabilistic phonological grammars provided every single grammar in the typology assigns more probability to the better consequent mapping \((\hat{x}, \hat{y})\) than to the worse antecedent \((x, y)\).

For instance, the implication

\[
(/\text{cost+us}/, [\text{cos.us}]) \rightarrow (/\text{cost+me}/, [\text{cos.me}])
\]

says that \(t\)-deletion always has a larger probability before consonants than before vowels

[Coetzee & Kawahara 2013]
We focus on a universal \((x, y) \rightarrow (\hat{x}, \hat{y})\) such that

- the antecedent comes with \(m\) additional candidates \(z_1, \ldots, z_m\)
- the consequent comes with the same number \(m\) of (possibly different) additional candidates \(\hat{z}_1, \ldots, \hat{z}_m\)

In earlier work, we have shown that, if this implication \((x, y) \rightarrow (\hat{x}, \hat{y})\) is indeed a universal of ME, constraint violations must satisfy the following inequality

\[
\sum_{i=1}^{m} \left( C(x, z_i) - C(x, y) \right) \leq \sum_{j=1}^{m} \left( C(\hat{x}, \hat{z}_j) - C(\hat{x}, \hat{y}) \right)
\]

This necessary condition follows from standard calculus. The hard question is: what the hell does it mean, phonologically?

Having pondered this hard question for a few years, we think we finally have an answer to share.
We focus on a universal \((x, y) \rightarrow (\hat{x}, \hat{y})\) such that

- the antecedent comes with \(m\) additional candidates \(z_1, \ldots, z_m\)
- the consequent comes with the same number \(m\) of (possibly different) additional candidates \(\hat{z}_1, \ldots, \hat{z}_m\)

In earlier work, we have shown that, if this implication \((x, y) \rightarrow (\hat{x}, \hat{y})\) is indeed a universal of ME, constraint violations must satisfy the following inequality

\[
\sum_{i=1}^{m} \left( C(x, z_i) - C(x, y) \right) \leq \sum_{j=1}^{m} \left( C(\hat{x}, \hat{z}_j) - C(\hat{x}, \hat{y}) \right)
\]

This necessary condition follows from standard calculus. The hard question is: what the hell does it mean, phonologically?

Having pondered this hard question for a few years, we think we finally have an answer to share.
We focus on a universal \((x, y) \rightarrow (\hat{x}, \hat{y})\) such that
- the antecedent comes with \(m\) additional candidates \(z_1, \ldots, z_m\)
- the consequent comes with the same number \(m\) of (possibly different) additional candidates \(\hat{z}_1, \ldots, \hat{z}_m\)

In earlier work, we have shown that, if this implication \((x, y) \rightarrow (\hat{x}, \hat{y})\) is indeed a universal of ME, constraint violations must satisfy the following inequality

\[
\sum_{i=1}^{m} \left( C(x, z_i) - C(x, y) \right) \leq \sum_{j=1}^{m} \left( C(\hat{x}, \hat{z}_j) - C(\hat{x}, \hat{y}) \right)
\]

This necessary condition follows from standard calculus. The hard question is: what the hell does it mean, phonologically?

Having pondered this hard question for a few years, we think we finally have an answer to share.
We focus on a universal \((x, y) \rightarrow (\hat{x}, \hat{y})\) such that

- the antecedent comes with \(m\) additional candidates \(z_1, \ldots, z_m\)
- the consequent comes with the same number \(m\) of (possibly different) additional candidates \(\hat{z}_1, \ldots, \hat{z}_m\)

In earlier work, we have shown that, if this implication \((x, y) \rightarrow (\hat{x}, \hat{y})\) is indeed a universal of ME, constraint violations must satisfy the following inequality

\[
\sum_{i=1}^{m} \left( C(x, z_i) - C(x, y) \right) \leq \sum_{j=1}^{m} \left( C(\hat{x}, \hat{z}_j) - C(\hat{x}, \hat{y}) \right)
\]  

This necessary condition follows from standard calculus. The hard question is: what the hell does it mean, phonologically?

Having pondered this hard question for a few years, we think we finally have an answer to share.
First ingredient:

- We focus on cases where the underlying and surface forms coincide in both the antecedent and the consequent mapping:
  \[ x = y \quad \hat{x} = \hat{y} \]

- Considerations of faithfulness cannot distinguish between antecedent \((x, x)\) and consequent \((\hat{x}, \hat{x})\): both are impeccable.

- Suppose that faithfulness and markedness are the only two perspectives relevant for phonology.

- The only sense in which the consequent \((\hat{x}, \hat{x})\) is better than the antecedent \((x, x)\) is that \(\hat{x}\) is less marked than \(x\).

- Implications \((x, x) \rightarrow (\hat{x}, \hat{x})\) between faithful mappings are markedness implications that capture markedness asymmetries.
First ingredient:

- We focus on cases where the underlying and surface forms coincide in both the antecedent and the consequent mapping:
  \[ x = y \quad \hat{x} = \hat{y} \]

- Considerations of faithfulness cannot distinguish between antecedent \((x, x)\) and consequent \((\hat{x}, \hat{x})\): both are impeccable

- Suppose that faithfulness and markedness are the only two perspectives relevant for phonology

- The only sense in which the consequent \((\hat{x}, \hat{x})\) is better than the antecedent \((x, x)\) is that \(\hat{x}\) is less marked than \(x\)

- Implications \((x, x) \rightarrow (\hat{x}, \hat{x})\) between faithful mappings are markedness implications that capture markedness asymmetries
First ingredient:

- We focus on cases where the underlying and surface forms coincide in both the antecedent and the consequent mapping:

\[ x = y \quad \hat{x} = \hat{y} \]

- Considerations of faithfulness cannot distinguish between antecedent \((x, x)\) and consequent \((\hat{x}, \hat{x})\): both are impeccable

- Suppose that faithfulness and markedness are the only two perspectives relevant for phonology

- The only sense in which the consequent \((\hat{x}, \hat{x})\) is better than the antecedent \((x, x)\) is that \(\hat{x}\) is less marked than \(x\)

- Implications \((x, x) \rightarrow (\hat{x}, \hat{x})\) between faithful mappings are markedness implications that capture markedness asymmetries
First ingredient:

- We focus on cases where the underlying and surface forms coincide in both the antecedent and the consequent mapping:

  \[ x = y \quad \hat{x} = \hat{y} \]

- Considerations of faithfulness cannot distinguish between antecedent \((x, x)\) and consequent \((\hat{x}, \hat{x})\): both are impeccable.

- Suppose that faithfulness and markedness are the only two perspectives relevant for phonology.

- The only sense in which the consequent \((\hat{x}, \hat{x})\) is better than the antecedent \((x, x)\) is that \(\hat{x}\) is less marked than \(x\).

- Implications \((x, x) \rightarrow (\hat{x}, \hat{x})\) between faithful mappings are markedness implications that capture markedness asymmetries.
First ingredient:

- We focus on cases where the underlying and surface forms coincide in both the antecedent and the consequent mapping:
  \[ x = y \quad \hat{x} = \hat{y} \]

- Considerations of faithfulness cannot distinguish between antecedent \((x, x)\) and consequent \((\hat{x}, \hat{x})\): both are impeccable.

- Suppose that faithfulness and markedness are the only two perspectives relevant for phonology.

- The only sense in which the consequent \((\hat{x}, \hat{x})\) is better than the antecedent \((x, x)\) is that \(\hat{x}\) is less marked than \(x\).

- Implications \((x, x) \rightarrow (\hat{x}, \hat{x})\) between faithful mappings are markedness implications that capture markedness asymmetries.
Second ingredient:

- We denote by $\overline{C}(x)$ the **average** number of violations assigned by a constraint $C$ to the candidates of the underlying form $x$:

  $$\overline{C}(x) = \frac{1}{|Gen(x)|} \sum_{u \in Gen(x)} C(x, u)$$

- To illustrate, if /CV/ comes with four candidates:

  $$\overline{\text{Max}(/CV/)} = \frac{\text{Max}(/CV, [CV]) + \text{Max}(/CV, [CVC]) + \text{Max}(/CV, [V]) + \text{Max}(/CV, [VC])}{4}$$

  $$= \frac{0 + 0 + 1 + 1}{4} = 0.5$$
Second ingredient:

- We denote by $\overline{C}(x)$ the **average** number of violations assigned by a constraint $C$ to the candidates of the underlying form $x$:

$$\overline{C}(x) = \frac{1}{|\text{Gen}(x)|} \sum_{u \in \text{Gen}(x)} C(x,u)$$

- To illustrate, if /CV/ comes with four candidates:

$$\overline{\text{Max}}(/CV/) = \frac{\text{Max}(/CV/, [CV]) + \text{Max}(/CV/, [CVC]) + \text{Max}(/CV/, [V]) + \text{Max}(/CV/, [VC])}{4} = \frac{0 + 0 + 1 + 1}{4} = 0.5$$
A straightforward manipulation of our previous uninterpretable result yields the following corollary:

If a markedness implication \((x, x) \rightarrow (\hat{x}, \hat{x})\) is a universal of ME, the consequent \(\hat{x}\) has average faithfulness violations at least as large as the antecedent \(x\)

\[
\bar{F}(\hat{x}) \geq \bar{F}(x)
\]

for every faithfulness constraint \(F\)

This corollary is more insightful because average faithfulness violations have straightforward phonological interpretations.
A straightforward manipulation of our previous uninterpretable result yields the following corollary:

If a markedness implication \((x, x) \rightarrow (\tilde{x}, \tilde{x})\) is a universal of ME, the consequent \(\tilde{x}\) has average faithfulness violations at least as large as the antecedent \(x\)

\[ \bar{F}(\tilde{x}) \geq \bar{F}(x) \]

for every faithfulness constraint \(F\)

This corollary is more insightful because average faithfulness violations have straightforward phonological interpretations.
Longer underlying strings have a larger average number $\text{MAX}$ of deletions (when all strings share the same candidate set):

more underlying segments $\equiv$ more stuff to delete

To illustrate, here is the average number $\text{MAX}$ of deletions of the nine underlying strings of the Extended Syllable System [Prince & Smolensky 2004]

The faithfulness average inequality $\overline{F}(\hat{x}) \geq \overline{F}(x)$ for $F = \text{MAX}$ entails that the consequent form $\hat{x}$ of a ME markedness implication cannot be shorter than the antecedent form $x$
Longer underlying strings have a larger average number $\text{MAX}$ of deletions (when all strings share the same candidate set):

more underlying segments $\equiv$ more stuff to delete

To illustrate, here is the average number $\text{MAX}$ of deletions of the nine underlying strings of the Extended Syllable System:

$\text{MAX}$

The faithfulness average inequality $\overline{F(\hat{x})} \geq \overline{F(x)}$ for $F = \text{MAX}$ entails that the consequent form $\hat{x}$ of a ME markedness implication cannot be shorter than the antecedent form $x$.
Longer underlying strings have a larger average number $\text{MAX}$ of deletions (when all strings share the same candidate set): 

*more underlying segments $\equiv$ more stuff to delete*

To illustrate, here is the average number $\text{MAX}$ of deletions of the nine underlying strings of the Extended Syllable System.

[Prince & Smolensky 2004]

The faithfulness average inequality $\overline{F(\hat{x})} \geq \overline{F(x)}$ for $F = \text{MAX}$ entails that the consequent form $\hat{x}$ of a ME markedness implication *cannot be shorter* than the antecedent form $x$. 
Reverse considerations hold for Dep: shorter underlying strings have a larger average number of epentheses, as illustrated again with the Extended Syllable System:

\[
\begin{array}{c|c|c|c|c|c|c|c}
/V/ & /VC/ & /CV/ & /CVC/ & /CVCC/ & /CCVC/ & /CCVCC/ \\
\hline
2 & 1.3 & 1.3 & 1 & 1 & 0.7 & 0.3 & 0.3 & 0
\end{array}
\]

The faithfulness average inequality \( \bar{F}(\hat{x}) \geq \bar{F}(x) \) for \( F = \text{Dep} \) entails that the consequent form \( \hat{x} \) of a ME markedness implication cannot be longer than the antecedent form \( x \).

In conclusion, we obtain the equi-length generalization:

If a markedness implication \((x, x) \rightarrow (\hat{x}, \hat{x})\) is a universal of ME, the strings compared share same sheer length \(|\hat{x}| = |x|\).

Needless to say, this generalization about markedness in ME is phonologically paradoxical!
Reverse considerations hold for Dep: shorter underlying strings have a larger average number of epentheses, as illustrated again with the Extended Syllable System.

The faithfulness average inequality $\bar{F}(\hat{x}) \geq \bar{F}(x)$ for $F = \text{Dep}$ entails that the consequent form $\hat{x}$ of a ME markedness implication cannot be longer than the antecedent form $x$.

In conclusion, we obtain the equi-length generalization:

\[
\text{If a markedness implication } (x, x) \rightarrow (\hat{x}, \hat{x}) \text{ is a universal of ME, the strings compared share same sheer length } |\hat{x}| = |x|.
\]

Needless to say, this generalization about markedness in ME is phonologically paradoxical!
□ Reverse considerations hold for Dep: shorter underlying strings have a larger average number of epentheses, as illustrated again with the Extended Syllable System

□ The faithfulness average inequality \( \overline{F}(\hat{x}) \geq \overline{F}(x) \) for \( F = \text{Dep} \) entails that the consequent form \( \hat{x} \) of a ME markedness implication cannot be longer than the antecedent form \( x \)

□ In conclusion, we obtain the equi-length generalization:

\[
\text{If a markedness implication } (x, x) \to (\hat{x}, \hat{x}) \text{ is a universal of ME, the strings compared share same sheer length } |\hat{x}| = |x|
\]

□ Needless to say, this generalization about markedness in ME is phonologically paradoxical!
Reverse considerations hold for \( \text{Dep} \): shorter underlying strings have a larger average number of epentheses, as illustrated again with the Extended Syllable System:

\[
\begin{array}{cccccccc}
2 & 1.3 & 1.3 & 1 & 0.7 & 0.3 & 0.3 & 0 \\
\end{array}
\]

The faithfulness average inequality \( \overline{F}(\hat{x}) \geq \overline{F}(x) \) for \( F = \text{Dep} \) entails that the consequent form \( \hat{x} \) of a ME markedness implication cannot be longer than the antecedent form \( x \).

In conclusion, we obtain the equi-length generalization:

*If a markedness implication \((x, x) \rightarrow (\hat{x}, \hat{x})\) is a universal of ME, the strings compared share same sheer length \( |\hat{x}| = |x| \).*

Needless to say, this generalization about markedness in ME is phonologically paradoxical!
To illustrate, we consider again the nine faithful mappings in the Extended Syllable System: they are ordered by SHG into 16 reasonable markedness implications.

We know that, if an implication is a universal of ME, it is also a universal of SHG: so how many of these 16 reasonable markedness implications carry over from SHG to ME?

15 of these markedness implications compare forms with different lengths, such as (/CCVCC/, [CCVCC]) → (/CV/, [CV]): they fail in ME because they flout the equi-length generalization.

Only the implication (/VC/, [VC]) → (/CV/, [CV]) satisfies the equi-length generalization... but it fails in ME for independent reasons (that we can detail, but in a different talk).
To illustrate, we consider again the nine faithful mappings in the Extended Syllable System: they are ordered by SHG into 16 reasonable markedness implications.

We know that, if an implication is a universal of ME, it is also a universal of SHG: so how many of these 16 reasonable markedness implications carry over from SHG to ME?

15 of these markedness implications compare forms with different lengths, such as (/CCVCC/, [CCVCC]) → (/CV/, [CV]): they fail in ME because they flout the equi-length generalization.

Only the implication (/VC/, [VC]) → (/CV/, [CV]) satisfies the equi-length generalization... but it fails in ME for independent reasons (that we can detail, but in a different talk).
To illustrate, we consider again the nine faithful mappings in the Extended Syllable System: they are ordered by SHG into 16 reasonable markedness implications.

We know that, if an implication is a universal of ME, it is also a universal of SHG: so how many of these 16 reasonable markedness implications carry over from SHG to ME?

15 of these markedness implications compare forms with different lengths, such as (/CCVCC/,[CCVCC]) → (/CV/,[CV]): they fail in ME because they flout the equi-length generalization.

Only the implication (/VC/,[VC]) → (/CV/,[CV]) satisfies the equi-length generalization... but it fails in ME for independent reasons (that we can detail, but in a different talk).
To illustrate, we consider again the nine faithful mappings in the Extended Syllable System: they are ordered by SHG into 16 reasonable markedness implications.

We know that, if an implication is a universal of ME, it is also a universal of SHG: so how many of these 16 reasonable markedness implications carry over from SHG to ME?

15 of these markedness implications compare forms with different lengths, such as (/CCVCC/, [CCVCC]) → (/CV/, [CV]): they fail in ME because they flout the equi-length generalization.

Only the implication (/VC/, [VC]) → (/CV/, [CV]) satisfies the equi-length generalization... but it fails in ME for independent reasons (that we can detail, but in a different talk).
In conclusion, ME predicts no markedness implications for the Extended Syllable System:

- no syllable counts as more marked than any other syllable
- any syllable can have a larger probability than any other syllable

When Max has some positive weight while the other constraints have weights equal or close to zero, ME probabilities track length rather than markedness, yielding complete markedness reversals.
To uncover further paradoxes of this average faithfulness inequality, let us dig deeper into the formalism of ME.

Another straightforward manipulation of our previous uninterpretable formal result yields the following corollary:

\[
\text{If a markedness implication } (x, x) \rightarrow (\hat{x}, \hat{x}) \text{ is a universal of ME, the consequent form } \hat{x} \text{ cannot violate any markedness constraint } M \text{ more than the antecedent form } x.
\]

Consider a markedness constraint \( M = \ast [+\varphi] \) that penalizes the marked value \(+\) of some feature \( \varphi \), such as \( M = \ast [+\text{nasal}] \).

The corollary ensures that the consequent string \( \hat{x} \) cannot have more segments with the marked value \([+\varphi]\) than the antecedent string \( x \), say it cannot have more nasals.
To uncover further paradoxes of this average faithfulness inequality, let us dig deeper into the formalism of ME.

Another straightforward manipulation of our previous uninterpretable formal result yields the following corollary:

\[
\text{If a markedness implication } (x, x) \rightarrow (\widehat{x}, \widehat{x}) \text{ is a universal of ME, the consequent form } \widehat{x} \text{ cannot violate any markedness constraint } M \text{ more than the antecedent form } x.
\]

Consider a markedness constraint \( M = \ast [+\varphi] \) that penalizes the marked value \(+\) of some feature \( \varphi \), such as \( M = \ast [+\text{nasal}] \).

The corollary ensures that the consequent string \( \widehat{x} \) cannot have more segments with the marked value \(+\varphi\) than the antecedent string \( x \), say it cannot have more nasals.
To uncover further paradoxes of this average faithfulness inequality, let us dig deeper into the formalism of ME.

Another straightforward manipulation of our previous uninterpretable formal result yields the following corollary:

**If a markedness implication \((x, x) \rightarrow (\hat{x}, \hat{x})\) is a universal of ME, the consequent form \(\hat{x}\) cannot violate any markedness constraint \(M\) more than the antecedent form \(x\).**

Consider a markedness constraint \(M = [+\varphi]\) that penalizes the marked value \(+\) of some feature \(\varphi\), such as \(M = [+\text{nasal}]\). The corollary ensures that the consequent string \(\hat{x}\) cannot have more segments with the marked value \([+\varphi]\) than the antecedent string \(x\), say it cannot have more nasals.
To uncover further paradoxes of this average faithfulness inequality, let us dig deeper into the formalism of ME.

Another straightforward manipulation of our previous uninterpretable formal result yields the following corollary:

If a markedness implication \((x, x) \rightarrow (\hat{x}, \hat{x})\) is a universal of ME, the consequent form \(\hat{x}\) cannot violate any markedness constraint \(M\) more than the antecedent form \(x\).

Consider a markedness constraint \(M = \ast[+\varphi]\) that penalizes the marked value \(+\) of some feature \(\varphi\), such as \(M = \ast[+\text{nasal}]\).

The corollary ensures that the consequent string \(\hat{x}\) cannot have more segments with the marked value \([+\varphi]\) than the antecedent string \(x\), say it cannot have more nasals.
□ The marked feature value \([+\varphi]\) has been argued to be protected by a dedicated featural constraint \(\text{MAX}[+\varphi]\), such as \(\text{MAX}[+\text{nasal}]\) that penalizes only de-nasalization, not nasalization.[Pater 1999]

□ Underlying strings with more nasals have a larger average number \(\text{MAX}[+\text{nasal}]\) of de-nasalizations (when all strings share the same candidate set):

\[
\text{more underlying nasals} \equiv \text{more stuff to de-nasalize}
\]

□ The faithfulness average inequality \(\bar{F}(\hat{x}) \geq \bar{F}(x)\) for \(F = \text{MAX}[+\varphi]\) entails that the consequent form \(\hat{x}\) of a ME markedness implication cannot have fewer segments with the marked value \([+\varphi]\) than the antecedent string \(x\), say it cannot have fewer nasals.
The marked feature value \([+\varphi]\) has been argued to be protected by a dedicated featural constraint \(\text{Max}[+\varphi]\), such as \(\text{Max}[+[\text{nasal}]\) that penalizes only de-nasalization, not nasalization [Pater 1999].

Underlying strings with more nasals have a larger average number \(\text{Max}_{-[\text{nasal}]}\) of de-nasalizations (when all strings share the same candidate set):

\[
\text{more underlying nasals} \equiv \text{more stuff to de-nasalize}
\]

The faithfulness average inequality \(\overline{F}(\hat{x}) \geq \overline{F}(x)\) for \(F = \text{Max}[+\varphi]\) entails that the consequent form \(\hat{x}\) of a ME markedness implication cannot have fewer segments with the marked value \([+\varphi]\) than the antecedent string \(x\), say it cannot have fewer nasals.
The marked feature value $[+\varphi]$ has been argued to be protected by a dedicated featural constraint $\text{Max}[+\varphi]$, such as $\text{Max}[+\text{nasal}]$ that penalizes only de-nasalization, not nasalization [Pater 1999].

Underlying strings with more nasals have a larger average number $\text{Max}[+\text{nasal}]$ of de-nasalizations (when all strings share the same candidate set):

$$\text{more underlying nasals} \equiv \text{more stuff to de-nasalize}$$

The faithfulness average inequality $\overline{F}(\hat{x}) \geq \overline{F}(x)$ for $F = \text{Max}[+\varphi]$ entails that the consequent form $\hat{x}$ of a ME markedness implication cannot have fewer segments with the marked value $[+\varphi]$ than the antecedent string $x$, say it cannot have fewer nasals.
Putting the two corollaries together, we obtain the following **equi-count** generalization:

If a ME markedness implication \((x, x) \rightarrow (\hat{x}, \hat{x})\) is a universal of ME, the strings compared have the same number of occurrences of the marked feature value \([+\varphi]\), say the same number of nasals.

Needless to say, this generalization about markedness in ME is phonologically paradoxical!
Putting the two corollaries together, we obtain the following **equi-count** generalization:

*If a ME markedness implication \((x, x) \rightarrow (\bar{x}, \bar{x})\) is a universal of ME, the strings compared have the same number of occurrences of the marked feature value \([+\varphi]\), say the same number of nasals.*

Needless to say, this generalization about markedness in ME is phonologically paradoxical!
First example:

- Four forms atta, atna, anta, and anna, each a candidate of the other.
- We supplement the constraints $\ast[+\text{nasal}]$ and $\text{MAX}[+\text{nasal}]$ with:
  - $\ast\text{N}_C$, that penalizes [anta] [Pater 1999]
  - SYLLABLECONTACT, that penalizes [atna]
- SHG predicts five reasonable markedness implications:
  - (/anta/, [anta])  (/atna/, [atna])
  - (/anna/, [anna])
  - (/atta/, [atta])
- They all fail in ME because they all compare forms with different numbers of nasals: ME predicts no markedness asymmetries.
First example:

- Four forms atta, atna, anta, and anna, each a candidate of the other
- We supplement the constraints $\ast[+\text{nasal}]$ and $\text{MAX}[+\text{nasal}]$ with:
  - $\ast\text{NC}$, that penalizes [anta] [Pater 1999]
  - SYLLABLECONTACT, that penalizes [atna]
- SHG predicts five reasonable markedness implications
  
  $$(/\text{anta}/, [\text{anta}]) \\ (/\text{atna}/, [\text{atna}])$$

- They all fail in ME because they all compare forms with different numbers of nasals: ME predicts no markedness asymmetries
First example:

- Four forms atta, atna, anta, and anna, each a candidate of the other
- We supplement the constraints $\ast[+\text{nasal}]$ and $\text{MAX}[+\text{nasal}]$ with:
  - $\ast\text{NC}$, that penalizes [anta] [Pater 1999]
  - SyllableContact, that penalizes [atna]
- SHG predicts five reasonable markedness implications

\[
\begin{align*}
\text{(anta, [anta])} & \quad \text{(atna, [atna])} \\
\text{(anna, [anna])} & \quad \text{(atta, [atta])}
\end{align*}
\]

- They all fail in ME because they all compare forms with different numbers of nasals: ME predicts no markedness asymmetries
First example:

- Four forms atta, atna, anta, and anna, each a candidate of the other
- We supplement the constraints $\star [+nasal]$ and $\text{Max}[+nasal]$ with:
  - $\star \text{NC}$, that penalizes [anta]
  - $\text{SyllableContact}$, that penalizes [atna]

- SHG predicts five reasonable markedness implications

  $$(/\text{anta}/, [\text{anta}]) \quad (/\text{atna}/, [\text{atna}])$$

- They all fail in ME because they all compare forms with different numbers of nasals: ME predicts no markedness asymmetries
Second example:

- Four forms $an$, $\tilde{an}$, $ad$, and $\tilde{ad}$, each a candidate of the other
- We supplement the constraints $*[+++\text{nasal}]$ and $\text{MAX}[++\text{nasal}]$ with:
  - $*[-\text{nasal}, +\text{syllabic}][++\text{nasal}]$, that penalizes $[an]$ [Kager 1999]
  - $\text{MAX}_{\text{consonant}}[++\text{nasal}]$, that penalizes consonant denasalization
- SHG predicts four reasonable markedness implications
  \[
  \begin{align*}
  (/\tilde{ad}/, [\tilde{ad}]) \\
  (/an/, [an]) & \quad (/\tilde{an}/, [\tilde{an}]) \\
  (/ad/, [ad])
  \end{align*}
  \]
- They all fail in ME because they all compare forms with different numbers of nasals: ME predicts no markedness asymmetries.
Second example:

- Four forms \( an, \tilde{an}, ad, \) and \( \tilde{ad} \), each a candidate of the other

- We supplement the constraints \(*[+\text{nasal}]\) and \( \text{MAX}_{[+\text{nasal}]} \) with:
  - \(*[-\text{nasal}, +\text{syllabic}][+\text{nasal}], \) that penalizes \( [an] \) [Kager 1999]
  - \( \text{MAX}_{\text{consonant}}^{[+\text{nasal}]}, \) that penalizes consonant denasalization

- SHG predicts four reasonable markedness implications

\[
\begin{align*}
(/\tilde{ad}/,[\tilde{ad}]) & \quad \Downarrow \\
(/an/,[an]) & \quad \Downarrow \\
(/\tilde{an}/,[\tilde{an}]) & \quad \Downarrow \\
(/ad/,[ad]) & \quad \Downarrow \\
\end{align*}
\]

- They all fail in ME because they all compare forms with different numbers of nasals: ME predicts no markedness asymmetries
Second example:

- Four forms an, ān, ad, and ād, each a candidate of the other
- We supplement the constraints *[+nasal] and \( \text{MAX}_{[+\text{nasal}]} \) with:
  - \(*[-\text{nasal}, +\text{syllabic}][+\text{nasal}]\), that penalizes [an] [Kager 1999]
  - \( \text{MAX}_{\text{consonant}}^{[+\text{nasal}]} \), that penalizes consonant denasalization
- SHG predicts four reasonable markedness implications

\[
\begin{align*}
(\text{/ād/}, [\text{ad}]) \\
(\text{/ān/}, [\text{ān}]) \\
(\text{/an/}, [\text{an}]) \\
(\text{/ad/}, [\text{ad}])
\end{align*}
\]

- They all fail in ME because they all compare forms with different numbers of nasals: ME predicts no markedness asymmetries
Second example:

- Four forms an, ān, ad, and ād, each a candidate of the other
- We supplement the constraints *[+nasal] and \( \text{Max}[+\text{nasal}] \) with:
  - \( *[−\text{nasal}, +\text{syllabic}][+\text{nasal}] \), that penalizes [an] \[\text{[Kager 1999]}\]
  - \( \text{Max}^{\text{consonant}}[+\text{nasal}] \), that penalizes consonant denasalization
- SHG predicts four reasonable markedness implications

\[\begin{align*}
\langle /\tilde{a}d/, [\tilde{a}d]\rangle \\
\langle /a\ddot{n}/, [a\ddot{n}]\rangle \\
\langle /a\ddot{n}/, [\tilde{a}n]\rangle \\
\langle /ad/, [ad]\rangle
\end{align*}\]

- They all fail in ME because they all compare forms with different numbers of nasals: ME predicts no markedness asymmetries
Summary:

- We have used markedness implicational universals of the form $(x, x) \rightarrow (\bar{x}, \bar{x})$ to evaluate typologies of probabilistic grammars.

- We have seen that SHG predicts plausible markedness implicational universals.

- ME does not because it requires the antecedent and consequent forms to have
  - the same sheer length
  - the same sheer numbers of marked segments

- Neither ME requirement makes phonological sense: our analysis thus casts doubt on ME as a model of phonological knowledge.
Summary:

- We have used markedness implicational universals of the form $(x, x) \rightarrow (\tilde{x}, \tilde{x})$ to evaluate typologies of probabilistic grammars.
- We have seen that SHG predicts plausible markedness implicational universals.
- ME does not because it requires the antecedent and consequent forms to have:
  - the same sheer length
  - the same sheer numbers of marked segments
- Neither ME requirement makes phonological sense: our analysis thus casts doubt on ME as a model of phonological knowledge.
Summary:

- We have used markedness implicational universals of the form $(x, x) \rightarrow (\bar{x}, \bar{x})$ to evaluate typologies of probabilistic grammars.
- We have seen that SHG predicts plausible markedness implicational universals.
- ME does not because it requires the antecedent and consequent forms to have
  - the same sheer length
  - the same sheer numbers of marked segments
- Neither ME requirement makes phonological sense: our analysis thus casts doubt on ME as a model of phonological knowledge.
Summary:

- We have used markedness implicational universals of the form $(x, x) \rightarrow (\tilde{x}, \tilde{x})$ to evaluate typologies of probabilistic grammars.
- We have seen that SHG predicts plausible markedness implicational universals.
- ME does not because it requires the antecedent and consequent forms to have:
  - the same sheer length
  - the same sheer numbers of marked segments
- Neither ME requirement makes phonological sense: our analysis thus casts doubt on ME as a model of phonological knowledge.
Thank you!