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

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

- Intuitively, this implication says that the consequent mapping $(\widehat{x}, \widehat{y})$ 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 $(\widehat{x}, \widehat{y})$ than to the worse antecedent (x, y)

- For instance, the implication

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

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

$$({\color{red}x}, {\color{red}y}) \rightarrow ({\color{blue}\widehat{x}}, {\color{blue}\widehat{y}})$$

- Intuitively, this implication says that the consequent mapping $({\color{blue}\widehat{x}}, {\color{blue}\widehat{y}})$ is "better" than the antecedent mapping $({\color{red}x}, {\color{red}y})$

- 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 $(\widehat{x}, \widehat{y})$ than to the worse antecedent (x, y)

- For instance, the implication

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

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

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

- Intuitively, this implication says that the consequent mapping $(\widehat{x}, \widehat{y})$ 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 $(\widehat{x}, \widehat{y})$ 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

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

- Intuitively, this implication says that the consequent mapping $(\widehat{x}, \widehat{y})$ 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 $(\widehat{x}, \widehat{y})$ than to the worse antecedent (x, y)
- For instance, the implication

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

says that *t*-deletion always has a larger probability before consonants than before vowels [Coetzee & Kawahara 2013]

- We focus on a universal $(x, y) \rightarrow (\widehat{x}, \widehat{y})$ 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 $\widehat{z}_1, \ldots, \widehat{z}_m$
- In earlier work, we have shown that, if this implication $(x, y) \rightarrow (\widehat{x}, \widehat{y})$ is indeed a universal of ME, constraint violations must satisfy the following inequality [Anttila & Magri 2017]

$$\sum_{i=1}^{m} \Big(C(x, z_i) - C(x, y) \Big) \leq \sum_{j=1}^{m} \Big(C(\widehat{x}, \widehat{z}_j) - C(\widehat{x}, \widehat{y}) \Big)$$

- 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 $(x, y) \rightarrow (\widehat{x}, \widehat{y})$ 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 $\widehat{z}_1, \ldots, \widehat{z}_m$

- In earlier work, we have shown that, if this implication $(x, y) \rightarrow (\widehat{x}, \widehat{y})$ is indeed a universal of ME, constraint violations must satisfy the following inequality [Anttila & Magri 2017]

$$\sum_{i=1}^{m} \Big(C(x, z_i) - C(x, y) \Big) \leq \sum_{j=1}^{m} \Big(C(\widehat{x}, \widehat{z}_j) - C(\widehat{x}, \widehat{y}) \Big)$$

- 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 $(x, y) \rightarrow (\widehat{x}, \widehat{y})$ 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 $\widehat{z}_1, \ldots, \widehat{z}_m$
- In earlier work, we have shown that, if this implication $(x, y) \rightarrow (\widehat{x}, \widehat{y})$ is indeed a universal of ME, constraint violations must satisfy the following inequality [Anttila & Magri 2017]

$$\sum_{i=1}^{m} \Big(C(x, z_i) - C(x, y) \Big) \leq \sum_{j=1}^{m} \Big(C(\widehat{x}, \widehat{z}_j) - C(\widehat{x}, \widehat{y}) \Big)$$

- 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 $(x, y) \rightarrow (\widehat{x}, \widehat{y})$ 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 $\widehat{z}_1, \ldots, \widehat{z}_m$

- In earlier work, we have shown that, if this implication $(x, y) \rightarrow (\widehat{x}, \widehat{y})$ is indeed a universal of ME, constraint violations must satisfy the following inequality [Anttila & Magri 2017]

$$\sum_{i=1}^{m} \Big(C(x, z_i) - C(x, y) \Big) \leq \sum_{j=1}^{m} \Big(C(\widehat{x}, \widehat{z}_j) - C(\widehat{x}, \widehat{y}) \Big)$$

- 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:

$$x = y \qquad \widehat{x} = \widehat{y}$$

- Considerations of faithfulness cannot distinguish between antecedent (x, x) and consequent $(\widehat{x}, \widehat{x})$: both are impeccable
- Suppose that faithfulness and markedness are the only two perspectives relevant for phonology
- The only sense in which the consequent $(\widehat{x}, \widehat{x})$ is better than the antecedent (x, x) is that $\widehat{x}$ is **less marked** than x
- Implications $(x, x) \rightarrow (\widehat{x}, \widehat{x})$ 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:

$$x = y \qquad \widehat{x} = \widehat{y}$$

- Considerations of faithfulness cannot distinguish between antecedent (x, x) and consequent $(\widehat{x}, \widehat{x})$: both are impeccable
- Suppose that faithfulness and markedness are the only two perspectives relevant for phonology
- The only sense in which the consequent $(\widehat{x}, \widehat{x})$ is better than the antecedent (x, x) is that $\widehat{x}$ is **less marked** than x
- Implications $(x, x) \rightarrow (\widehat{x}, \widehat{x})$ 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:

$$x = y \qquad \widehat{x} = \widehat{y}$$

- Considerations of faithfulness cannot distinguish between antecedent (x, x) and consequent $(\widehat{x}, \widehat{x})$: both are impeccable
- Suppose that faithfulness and markedness are the only two perspectives relevant for phonology
- The only sense in which the consequent $(\widehat{x}, \widehat{x})$ is better than the antecedent (x, x) is that $\widehat{x}$ is **less marked** than x
- Implications $(x, x) \rightarrow (\widehat{x}, \widehat{x})$ 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:

$$x = y \qquad \widehat{x} = \widehat{y}$$

- Considerations of faithfulness cannot distinguish between antecedent (x, x) and consequent $(\widehat{x}, \widehat{x})$: both are impeccable
- Suppose that faithfulness and markedness are the only two perspectives relevant for phonology
- The only sense in which the consequent $(\widehat{x}, \widehat{x})$ is better than the antecedent (x, x) is that $\widehat{x}$ is **less marked** than x
- Implications $(x, x) \rightarrow (\widehat{x}, \widehat{x})$ 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:

$$x = y \qquad \widehat{x} = \widehat{y}$$

- Considerations of faithfulness cannot distinguish between antecedent (x, x) and consequent $(\widehat{x}, \widehat{x})$: both are impeccable
- Suppose that faithfulness and markedness are the only two perspectives relevant for phonology
- The only sense in which the consequent $(\widehat{x}, \widehat{x})$ is better than the antecedent (x, x) is that $\widehat{x}$ is **less marked** than x
- Implications $(x, x) \rightarrow (\widehat{x}, \widehat{x})$ 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:

$$\begin{aligned}\overline{\textsc{Max}}(/CV/) &= \frac{\textsc{Max}(/CV/, [CV]) + \textsc{Max}(/CV/, [CVC]) + \textsc{Max}(/CV/, [V]) + \textsc{Max}(/CV/, [VC])}{4} \\ &= \frac{0+0+1+1}{4} = 0.5\end{aligned}$$

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:

$$\begin{aligned}\overline{\textsc{Max}}(/CV/) &= \frac{\begin{array}{c}\textsc{Max}(/CV/, [CV]) + \textsc{Max}(/CV/, [CVC]) + \\ \textsc{Max}(/CV/, [V]) + \textsc{Max}(/CV/, [VC])\end{array}}{4} \\ &= \frac{0+0+1+1}{4} = 0.5\end{aligned}$$

- A straightforward manipulation of our previous uninterpretable result yields the following corollary:

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

$$\overline{F}(\widehat{x}) \geq \overline{F}(x)$$

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* $(x, x) \rightarrow (\widehat{x}, \widehat{x})$ *is a universal of ME, the consequent* $\widehat{\mathrm{x}}$ *has average faithfulness violations at least as large as the antecedent* x
>
> $$\overline{F}(\widehat{x}) \geq \overline{F}(x)$$
>
> *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 $\overline{\text{Max}}$ of deletions (when all strings share the same candidate set):

  more underlying segments ≡ more stuff to delete

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

  [Prince & Smolensky 2004]

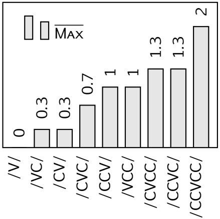
- The faithfulness average inequality $\overline{F}(\widehat{x}) \geq \overline{F}(x)$ for $F = \text{Max}$ entails that the consequent form $\widehat{x}$ of a ME markedness implication **cannot be shorter** than the antecedent form x

- Longer underlying strings have a larger average number $\overline{\text{MAX}}$ of deletions (when all strings share the same candidate set):

  more underlying segments ≡ more stuff to delete

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

  [Prince & Smolensky 2004]



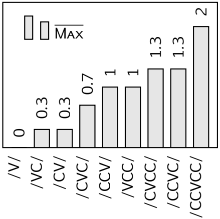
- The faithfulness average inequality $\overline{F}(\widehat{x}) \geq \overline{F}(x)$ for $F = \text{MAX}$ entails that the consequent form $\widehat{x}$ of a ME markedness implication **cannot be shorter** than the antecedent form x

- Longer underlying strings have a larger average number $\overline{\text{MAX}}$ of deletions (when all strings share the same candidate set):

more underlying segments ≡ more stuff to delete

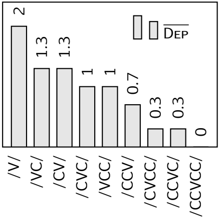
- To illustrate, here is the average number $\overline{\text{MAX}}$ of deletions of the nine underlying strings of the Extended Syllable System

[Prince & Smolensky 2004]



- The faithfulness average inequality $\overline{F}(\widehat{x}) \geq \overline{F}(x)$ for $F = \text{MAX}$ entails that the consequent form $\widehat{x}$ 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 $\overline{\text{DEP}}$ of epentheses, as illustrated again with the Extended Syllable System



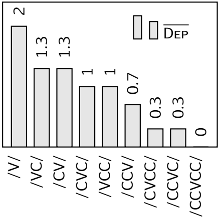
- The faithfulness average inequality $\overline{F}(\widehat{x}) \geq \overline{F}(x)$ for $F = \text{DEP}$ entails that the consequent form $\widehat{x}$ of a ME markedness implication **cannot be longer** than the antecedent form x

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

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

- 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 $\overline{\text{DEP}}$ of epentheses, as illustrated again with the Extended Syllable System



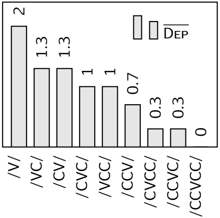
- The faithfulness average inequality $\overline{F}(\widehat{x}) \geq \overline{F}(x)$ for $F = $ DEP entails that the consequent form $\widehat{x}$ of a ME markedness implication **cannot be longer** than the antecedent form x

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

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

- 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 $\overline{\text{DEP}}$ of epentheses, as illustrated again with the Extended Syllable System



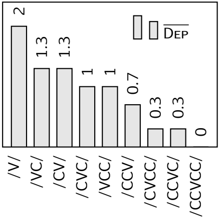
- The faithfulness average inequality $\overline{F}(\widehat{x}) \geq \overline{F}(x)$ for $F = \text{DEP}$ entails that the consequent form $\widehat{x}$ of a ME markedness implication **cannot be longer** than the antecedent form x

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

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

- 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 $\overline{\text{DEP}}$ of epentheses, as illustrated again with the Extended Syllable System



- The faithfulness average inequality $\overline{F}(\widehat{x}) \geq \overline{F}(x)$ for $F = \text{DEP}$ entails that the consequent form $\widehat{x}$ 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 $(x, x) \rightarrow (\widehat{x}, \widehat{x})$ is a universal of ME, the strings compared share same sheer length $|\widehat{x}| = |x|$*

- Needless to say, this generalization about markedness in ME is phonologically paradoxical!

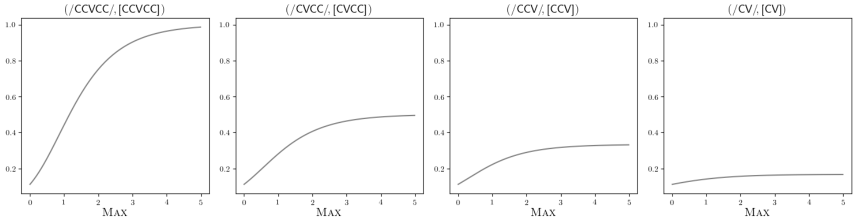
- To illustrate, we consider again the nine faithful mappings in the Extended Syllable System: they are ordered by SHG into 16 reasonable markedness implications
- We know that, if an implication is a universal of ME, it is also a universal of SHG: so how many of these 16 reasonable markedness implications carry over from SHG to ME?
- 15 of these markedness implications compare forms with different lengths, such as $(/\mathrm{CCVCC}/, [\mathrm{CCVCC}]) \rightarrow (/\mathrm{CV}/, [\mathrm{CV}])$: they fail in ME because they flout the equi-length generalization
- Only the implication $(/\mathrm{VC}/, [\mathrm{VC}]) \rightarrow (/\mathrm{CV}/, [\mathrm{CV}])$ 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 $(/\mathrm{CCVCC}/, [\mathrm{CCVCC}]) \rightarrow (/\mathrm{CV}/, [\mathrm{CV}])$: they fail in ME because they flout the equi-length generalization
- Only the implication $(/\mathrm{VC}/, [\mathrm{VC}]) \rightarrow (/\mathrm{CV}/, [\mathrm{CV}])$ 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 $(/\mathrm{CCVCC}/, [\mathrm{CCVCC}]) \rightarrow (/\mathrm{CV}/, [\mathrm{CV}])$: they fail in ME because they flout the equi-length generalization
- Only the implication $(/\mathrm{VC}/, [\mathrm{VC}]) \rightarrow (/\mathrm{CV}/, [\mathrm{CV}])$ 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 $(/\mathrm{CCVCC}/, [\mathrm{CCVCC}]) \rightarrow (/\mathrm{CV}/, [\mathrm{CV}])$: they fail in ME because they flout the equi-length generalization
- Only the implication $(/\mathrm{VC}/, [\mathrm{VC}]) \rightarrow (/\mathrm{CV}/, [\mathrm{CV}])$ 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:

> *If a markedness implication $(x, x) \rightarrow (\widehat{x}, \widehat{x})$ is a universal of ME, the consequent form $\widehat{x}$ 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 φ, such as $M = *[+\text{nasal}]$
- 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 $(x, x) \rightarrow (\widehat{x}, \widehat{x})$ is a universal of ME, the consequent form $\widehat{x}$ 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 φ, such as $M = *[+\text{nasal}]$
- 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* $(x, x) \to (\widehat{x}, \widehat{x})$ *is a universal of ME, the consequent form* $\widehat{x}$ *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 φ, such as $M = *[+\text{nasal}]$
- 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* $(x, x) \rightarrow (\widehat{x}, \widehat{x})$ *is a universal of ME, the consequent form* $\widehat{x}$ *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 φ, such as $M = *[+\text{nasal}]$
- 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

- The marked feature value $[+\varphi]$ has been argued to be protected by a dedicated featural constraint $\textsc{Max}_{[+\varphi]}$, such as $\textsc{Max}_{[+\text{nasal}]}$ that penalizes only de-nasalization, not nasalization [Pater 1999]
- Underlying strings with more nasals have a larger average number $\overline{\textsc{Max}_{[+\text{nasal}]}}$ of de-nasalizations (when all strings share the same candidate set):

more underlying nasals ≡ more stuff to de-nasalize

- The faithfulness average inequality $\overline{F}(\widehat{x}) \geq \overline{F}(x)$ for $F = \textsc{Max}_{[+\varphi]}$ entails that the consequent form $\widehat{x}$ 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 $\textsc{Max}_{[+\varphi]}$, such as $\textsc{Max}_{[+\text{nasal}]}$ that penalizes only de-nasalization, not nasalization [Pater 1999]
- Underlying strings with more nasals have a larger average number $\overline{\textsc{Max}_{[+\text{nasal}]}}$ of de-nasalizations (when all strings share the same candidate set):

more underlying nasals ≡ more stuff to de-nasalize

- The faithfulness average inequality $\overline{F}(\widehat{x}) \geq \overline{F}(x)$ for $F = \textsc{Max}_{[+\varphi]}$ entails that the consequent form $\widehat{x}$ 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 $\textsc{Max}_{[+\varphi]}$, such as $\textsc{Max}_{[+\text{nasal}]}$ that penalizes only de-nasalization, not nasalization [Pater 1999]
- Underlying strings with more nasals have a larger average number $\overline{\textsc{Max}_{[+\text{nasal}]}}$ of de-nasalizations (when all strings share the same candidate set):

  more underlying nasals ≡ more stuff to de-nasalize

- The faithfulness average inequality $\overline{F}(\widehat{x}) \geq \overline{F}(x)$ for $F = \textsc{Max}_{[+\varphi]}$ entails that the consequent form $\widehat{x}$ 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* $(x, x) \to (\widehat{x}, \widehat{x})$ *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* $(x, x) \rightarrow (\widehat{x}, \widehat{x})$ *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 $*[+\text{nasal}]$ and $\textsc{Max}_{[+\text{nasal}]}$ with:
  - *NC̥, that penalizes [anta] [Pater 1999]
  - SyllableContact, that penalizes [atna]
- SHG predicts five reasonable markedness implications

$$
\begin{tikzcd}
(/\text{anta}/, [\text{anta}]) \arrow[dr] \arrow[ddr, bend right] & & (/\text{atna}/, [\text{atna}]) \arrow[dl] \arrow[ddl, bend left] \\
& (/\text{anna}/, [\text{anna}]) \arrow[d] & \\
& (/\text{atta}/, [\text{atta}]) &
\end{tikzcd}
$$

- 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 $*$[+nasal] and MAX$_{[+\text{nasal}]}$ with:
  - $*$NC̥, that penalizes [anta] [Pater 1999]
  - SYLLABLECONTACT, that penalizes [atna]

- SHG predicts five reasonable markedness implications

- 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 $*$[+nasal] and MAX$_{[+\text{nasal}]}$ with:
  - $*$N$\mathring{\text{C}}$, that penalizes [anta] [Pater 1999]
  - SYLLABLECONTACT, that penalizes [atna]
- SHG predicts five reasonable markedness implications

$$
\begin{tikzcd}
(/\text{anta}/, [\text{anta}]) \arrow[dr] \arrow[ddr, bend right] & & (/\text{atna}/, [\text{atna}]) \arrow[dl] \arrow[ddl, bend left] \\
& (/\text{anna}/, [\text{anna}]) \arrow[d] & \\
& (/\text{atta}/, [\text{atta}]) &
\end{tikzcd}
$$

- 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 $*$[+nasal] and MAX$_{[+\text{nasal}]}$ with:
  - $*$NC̥, that penalizes [anta] [Pater 1999]
  - SYLLABLECONTACT, that penalizes [atna]
- SHG predicts five reasonable markedness implications

$$
\begin{tikzcd}
(/\text{anta}/, [\text{anta}]) \arrow[dr] \arrow[ddr, bend right] & & (/\text{atna}/, [\text{atna}]) \arrow[dl] \arrow[ddl, bend left] \\
& (/\text{anna}/, [\text{anna}]) \arrow[d] & \\
& (/\text{atta}/, [\text{atta}]) &
\end{tikzcd}
$$

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

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

- 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, ãn, ad, and ãd, each a candidate of the other
- We supplement the constraints $*$[+nasal] and $\textsc{Max}_{[+\text{nasal}]}$ with:
  - $*$[−nasal, +syllabic][+nasal], that penalizes [an] [Kager 1999]
  - $\textsc{Max}^{\text{consonant}}_{[+\text{nasal}]}$, that penalizes consonant denasalization
- SHG predicts four reasonable markedness implications

$$\begin{tikzcd}
 & & (/\tilde{a}d/, [\tilde{a}d]) \arrow[d] \arrow[ddl, bend right] \\
(/an/, [an]) \arrow[dr] & & (/\tilde{a}n/, [\tilde{a}n]) \arrow[dl] \\
 & (/ad/, [ad]) &
\end{tikzcd}$$

- 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, ãn, ad, and ãd, each a candidate of the other
- We supplement the constraints $*[+\text{nasal}]$ and $\textsc{Max}_{[+\text{nasal}]}$ with:
  - $*[-\text{nasal}, +\text{syllabic}][+\text{nasal}]$, that penalizes [an] [Kager 1999]
  - $\textsc{Max}^{\text{consonant}}_{[+\text{nasal}]}$, that penalizes consonant denasalization
- SHG predicts four reasonable markedness implications

$$
\begin{tikzcd}
 & & (/\tilde{\text{a}}\text{d}/, [\tilde{\text{a}}\text{d}]) \arrow[d] \arrow[dl, bend right] \\
(/\text{an}/, [\text{an}]) \arrow[dr] & & (/\tilde{\text{a}}\text{n}/, [\tilde{\text{a}}\text{n}]) \arrow[dl] \\
 & (/\text{ad}/, [\text{ad}]) &
\end{tikzcd}
$$

- 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 $(x, x) \rightarrow (\widehat{x}, \widehat{x})$ 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 $(x, x) \rightarrow (\widehat{x}, \widehat{x})$ 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 $(x, x) \rightarrow (\widehat{x}, \widehat{x})$ 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 $(x, x) \rightarrow (\widehat{x}, \widehat{x})$ 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!