Equiprobable mappings
in weighted constraint grammars

Arto Anttila (Stanford), Scott Borgeson (Stanford),
and Giorgio Magri (CNRS)

ACL Sigmorphon 2019

Firenze | August 2nd, 2019
1. From phonological equivalences to equiprobability
This paper is part of a larger project systematically comparing two approaches to probabilistic generative phonology:

- **Stochastic Harmonic Grammar** (SHG)
- **Maximum Entropy** (ME)

[Boersma 1997; Boersma and Hayes 2001; Goldwater and Johnson 2003; Hayes and Wilson 2008; Boersma and Pater 2016; Smith and Pater 2017; Zuraw and Hayes 2017; Hayes 2017]
This paper is part of a larger project systematically comparing two approaches to probabilistic generative phonology:

- **Stochastic Harmonic Grammar** (SHG)
- **Maximum Entropy** (ME)

Both SHG and ME typologies of probabilistic grammars are the result of “smoothing” an underlying typology of **categorical** HG grammars.

[Boersma 1997; Boersma and Hayes 2001; Goldwater and Johnson 2003; Hayes and Wilson 2008; Boersma and Pater 2016; Smith and Pater 2017; Zuraw and Hayes 2017; Hayes 2017]
This paper is part of a larger project systematically comparing two approaches to **probabilistic** generative phonology:

- **Stochastic Harmonic Grammar** (SHG)
- **Maximum Entropy** (ME)

Both SHG and ME typologies of probabilistic grammars are the result of “smoothing” an underlying typology of **categorical** HG grammars.

Yet, our project shows that SHG **retains** all the grammatical properties of the underlying HG categorical typology while ME **loses** most of them (and thus over-generates)
This paper is part of a larger project systematically comparing two approaches to probabilistic generative phonology:

- **Stochastic Harmonic Grammar** (SHG)
- **Maximum Entropy** (ME)

Both SHG and ME typologies of probabilistic grammars are the result of “smoothing” an underlying typology of categorical HG grammars.

Yet, our project shows that SHG retains all the grammatical properties of the underlying HG categorical typology while ME loses most of them (and thus over-generates).

In this paper, we illustrate the gist of this argument by focusing on a specific grammatical property: equivalent mappings.
A phonological process applies uniformly to all forms that share some relevant properties, but ignores the irrelevant ways in which they differ.
A phonological process applies uniformly to all forms that share some relevant properties, but ignores the irrelevant ways in which they differ.

To illustrate, vowel harmony targets, say, \([\pm \text{back}]\), but ignores, say, number of syllables: two mappings that only differ for length such as 

\[
(/\text{maa-nä}/, [\text{maana}]) \sim (/\text{rakastaja-nä}/, [\text{rakastajana}])
\]

are equivalent for categorical vowel harmony.
A phonological process applies uniformly to all forms that share some relevant properties, but ignores the irrelevant ways in which they differ.

To illustrate, vowel harmony targets, say, \([\pm \text{back}]\), but ignores, say, number of syllables: two mappings that only differ for length such as

\[
(\text{/maa-nä/}, [\text{maana}]) \sim (\text{/rakastaja-nä/}, [\text{rakastajana}])
\]

are equivalent for categorical vowel harmony.

These phonological equivalences are a key property of phonological systems: so how should this notion of phonological equivalence be extended from the categorical to the probabilistic setting?
A **probabilistic** phonological grammar assigns to each underlying representation \(/UR/\) a probability distribution \(\mathbb{P}([SR] | /UR/)\) over the corresponding set of candidate surface representations \([SR]\).
□ A **probabilistic** phonological grammar assigns to each underlying representation /UR/ a probability distribution $\mathbb{P}(\text{[SR]} \mid /UR/)$. over the corresponding set of candidate surface representations [SR].

□ Two mappings (/UR/, [SR]) and (/\hat{UR}/, [\hat{SR}]) are **equiprobable** if every grammar in the typology assigns them the same probability

$$\mathbb{P}([SR] \mid /UR/) = \mathbb{P}([\hat{SR}] \mid /\hat{UR}/)$$
A **probabilistic** phonological grammar assigns to each underlying representation \(/UR/\) a probability distribution \(\mathbb{P}([SR] \mid UR)\) over the corresponding set of candidate surface representations \([SR]\).

Two mappings (\(/UR/, [SR]\)) and (\(/\hat{UR}/, [\hat{SR}]\)) are **equiprobable** if every grammar in the typology assigns them the same probability

\[
\mathbb{P}([SR] \mid UR) = \mathbb{P}([\hat{SR}] \mid \hat{UR})
\]

We submit that **equiprobability** is the proper way of extending phonological equivalence from categorical to probabilistic phonology.
A **probabilistic** phonological grammar assigns to each underlying representation /*UR*/ a probability distribution $\mathbb{P}([SR] \mid /UR/) \) over the corresponding set of candidate surface representations [SR]

Two mappings (/UR/, [SR]) and (/\widehat{ UR}/, [\widehat{ SR}]) are **equiprobable** if every grammar in the typology assigns them the same probability

$$\mathbb{P}([SR] \mid /UR/) = \mathbb{P}([\widehat{SR}] \mid /\widehat{UR}/)$$

We submit that **equiprobability** is the proper way of extending phonological equivalence from categorical to probabilistic phonology

To illustrate, the fact that words that only differ for length are equivalent for vowel harmony means that the probability of vowel harmony does not depend on the number of syllables

$$\mathbb{P}([\text{maana}] \mid /\text{maa-nä}/) = \mathbb{P}([\text{rakastajana}] \mid /\text{rakastaja-nä}/)$$
2. ME admits no equiprobable mappings
Given a winner mapping and a competing loser mapping, we call **difference vector** the vector consisting of the constraint violations of the loser discounted of the violations of the winner.

\[
\text{difference vector} = \begin{bmatrix}
C_1(\text{loser}) - C_1(\text{winner}) \\
C_2(\text{loser}) - C_2(\text{winner}) \\
\vdots \\
C_n(\text{loser}) - C_n(\text{winner})
\end{bmatrix}
\]
Given a winner mapping and a competing loser mapping, we call **difference vector** the vector consisting of the constraint violations of the loser discounted of the violations of the winner

\[
\text{difference vector} = \begin{bmatrix}
C_1(\text{loser}) - C_1(\text{winner}) \\
C_2(\text{loser}) - C_2(\text{winner}) \\
\vdots \\
C_n(\text{loser}) - C_n(\text{winner})
\end{bmatrix}
\]

When there are only \( n = 2 \) constraints \( C_1 \) and \( C_2 \), difference vectors can be plotted as points in the Cartesian plane:
Suppose the mapping \((\text{UR}, [SR])\) comes with 5 difference vectors.

- \(c_1\)
- \(c_2\)
- \(c_3\)
- \(c_4\)
- \(c_5\)
Suppose the mapping (/UR/, [SR]) comes with 5 difference vectors.

The gray region is the **convex hull** generated by these vectors.
Suppose the mapping \((/UR/, [SR])\) comes with 5 difference vectors.

- The gray region is the **convex hull** generated by these vectors.
- The red region consists of points larger than a point in this convex hull.
Two mappings $(/UR/, [SR])$ and $(/\hat{UR}/, [\hat{SR}])$ are ME equiprobable if and only if the corresponding red and blue regions coincide.
Two mappings \((\overline{UR}, [SR])\) and \((\overline{\widehat{UR}}, \overline{SR})\) are ME equiprobable if and only if the corresponding red and blue regions coincide.

The difference vectors \(c_1\) and \(c_2\) are extreme points: they determine the shape of the region and must be shared by the two mappings.
Two mappings (/\(UR\)/, [\(SR\)]) and (/\(\hat{UR}\)/, [\(\hat{SR}\)]) are ME equiprobable if and only if the corresponding red and blue regions coincide.

The difference vectors \(c_1\) and \(c_2\) are extreme points: they determine the shape of the region and must be shared by the two mappings.

The difference vectors \(c_3\), \(c_4\), \(c_5\) are instead interior points: they do not contribute to the shape of the region.
Yet, since we have established that $c_1$ and $c_2$ are shared, we can effectively “peel them off” the two sides of the ME probability identity.
Yet, since we have established that $c_1$ and $c_2$ are shared, we can effectively “peel them off” the two sides of the ME probability identity.

In other words, we can ignore $c_1$ and $c_2$ and only consider the remaining difference vectors $c_3, c_4, c_5$. 
Yet, since we have established that $c_1$ and $c_2$ are shared, we can effectively “peel them off” the two sides of the ME probability identity.

In other words, we can ignore $c_1$ and $c_2$ and only consider the remaining difference vectors $c_3, c_4, c_5$.

Now $c_3$ and $c_5$ are extreme points of the new red region and must be shared by the equiprobable mappings $(\hat{UR}/, [\hat{SR}])$ and $(\hat{UR}/, [\hat{SR}])$. 
Yet, since we have established that $c_1$ and $c_2$ are shared, we can effectively “peel them off” the two sides of the ME probability identity.

In other words, we can ignore $c_1$ and $c_2$ and only consider the remaining difference vectors $c_3, c_4, c_5$.

Now $c_3$ and $c_5$ are extreme points of the new red region and must be shared by the equiprobable mappings ($/UR/, [SR]$) and ($/\hat{UR}/, [\hat{SR}]$).

And so on.
The conclusion of this reasoning is that the two mappings $(/UR/, [SR])$ and $(/\hat{UR}/, [\hat{SR}])$ are equiprobable in ME if and only if the corresponding sets of difference vectors exactly coincide.
The conclusion of this reasoning is that the two mappings \((\mathcal{U}R, [S\mathcal{R}])\) and \((\hat{\mathcal{U}}\hat{R}, [\hat{S}\hat{R}])\) are equiprobable in ME if and only if the corresponding sets of difference vectors exactly coincide.

Realistically, the difference vectors coincide only if the two mappings \((\mathcal{U}R, [S\mathcal{R}])\) and \((\hat{\mathcal{U}}\hat{R}, [\hat{S}\hat{R}])\) are actually the same mapping.
The conclusion of this reasoning is that the two mappings \((/UR/, [SR])\) and \((/\hat{UR}/, [\hat{SR}])\) are equiprobable in ME if and only if the corresponding sets of difference vectors exactly coincide.

Realistically, the difference vectors coincide only if the two mappings \((/UR/, [SR])\) and \((/\hat{UR}/, [\hat{SR}])\) are actually the same mapping.

In conclusion, ME effectively admits no equiprobable mappings.
3.

SHG instead does admit equiprobable mappings
Suppose the mapping \((/UR/, [SR])\) comes with 5 difference vectors.
Suppose the mapping \((UR, [SR])\) comes with 5 difference vectors.

The gray region is the convex cone generated by these vectors.
Suppose the mapping \( (\mathcal{UR}, [SR]) \) comes with 5 difference vectors.

- The gray region is the **convex cone** generated by these vectors.
- The red region consists of points larger than a point in this cone.
Suppose the mapping \((/UR/, [SR])\) comes with 5 difference vectors.

The gray region is the **convex cone** generated by these vectors.

The red region consists of points larger than a point in this cone.

Same as for ME with cones (\(\square\)) rather than convex hulls (\(\square\)).
Again, the mappings \((/UR/, [SR])\) and \((/\widehat{UR}/, [\widehat{SR}])\) are SHG equiprobable if and only if the red and blue regions coincide.
Again, the mappings $([UR],[SR])$ and $([\hat{UR}],[\hat{SR}])$ are SHG equiprobable if and only if the red and blue regions coincide.

The difference vector $c_1$ of the mapping $([UR],[SR])$ sits on the border but can be shifted (rescaled) without affecting the red region.
Again, the mappings \((/UR/, [SR])\) and \((/\widehat{UR}/, [\widehat{SR}])\) are SHG equiprobable if and only if the red and blue regions coincide.

The difference vector \(c_1\) of the mapping \((/UR/, [SR])\) sits on the border but can be shifted (rescaled) without affecting the red region.

The equiprobable mapping \((/\widehat{UR}/, [\widehat{SR}])\) thus needs not share this difference vector \(c_1\) but only a rescaling thereof.
Again, the mappings (\(UR, SR\)) and (\(\hat{UR}, \hat{SR}\)) are SHG equiprobable if and only if the red and blue regions coincide.

The difference vector \(c_1\) of the mapping (\(UR, SR\)) sits on the border but can be shifted (rescaled) without affecting the red region.

The equiprobable mapping (\(\hat{UR}, \hat{SR}\)) thus needs not share this difference vector \(c_1\) but only a rescaling thereof.

And nothing can be said about the interior vectors \(c_2, \ldots, c_5\).
The conclusion of this reasoning is that two mappings \((/UR/,[SR])\) and \((/\hat{UR}/,[\hat{SR}])\) are equiprobable in SHG if and only if each of the non-interior difference vectors is a rescaling of a non-interior difference vector of the other.
The conclusion of this reasoning is that two mappings $([UR], [SR])$ and $([\hat{UR}], [\hat{SR}])$ are equiprobable in SHG if and only if each of the non-interior difference vectors is a rescaling of a non-interior difference vector of the other.

This SHG condition is way weaker than than the ME condition above:
- ME condition requires identity of difference vectors...
  ...while this SHG condition only requires rescaling
- ME condition applies to all difference vectors...
  ...while this SHG condition ignores interior difference vectors
The conclusion of this reasoning is that two mappings \((/UR/, [SR])\) and \((/\widehat{UR}/, [\widehat{SR}])\) are equiprobable in SHG if and only if each of the **non-interior** difference vectors is a **rescaling** of a non-interior difference vector of the other.

This SHG condition is way weaker than than the ME condition above:

- ME condition requires **identity** of difference vectors...
  
  ...while this SHG condition only requires **rescaling**

- ME condition applies to **all** difference vectors...
  
  ...while this SHG condition ignores **interior** difference vectors

In conclusion, SHG does admit equiprobable mappings.
4.

A test case: equiprobability in Finnish secondary stress
We have seen that ME is so much richer than SHG that any two different mappings can be assigned different ME probabilities.

Is this richness of ME relative to SHG an empirical advantage or a case of unmotivated over-generation?

In this final part of the talk, we would like to offer (an admittedly very preliminary) argument that the latter might be the case.

We will look at Finnish secondary stress and argue that the equiprobable mappings predicted by SHG are (by and large) attested at least in Finnish while ME over-generates.
- We have seen that ME is so much richer than SHG that any two different mappings can be assigned different ME probabilities.

- Is this richness of ME relative to SHG an empirical advantage or a case of unmotivated over-generation?
We have seen that ME is so much richer than SHG that any two different mappings can be assigned different ME probabilities.

Is this richness of ME relative to SHG an empirical advantage or a case of unmotivated over-generation?

In this final part of the talk, we would like to offer (an admittedly very preliminary) argument that the latter might be the case.
We have seen that ME is so much richer than SHG that any two different mappings can be assigned different ME probabilities.

Is this richness of ME relative to SHG an empirical advantage or a case of unmotivated over-generation?

In this final part of the talk, we would like to offer (an admittedly very preliminary) argument that the latter might be the case.

We will look at Finnish secondary stress and argue that the equiprobable mappings predicted by SHG are (by and large) attested at least in Finnish while ME over-generates.
Finnish stress is governed by four core rules:

- Primary stress falls on the initial syllable
- Secondary stress falls on every other syllable after that...
- except that a light syllable is skipped if the syllable after that is heavy
- unless that heavy syllable is final

Finnish stress is governed by four core rules:

- Primary stress falls on the initial syllable.
- Secondary stress falls on every other syllable after that... except that a light syllable is skipped if the syllable after that is heavy.
- unless that heavy syllable is final.


We focus on the “skipping clause” and note that it leaks in long words: instead of being categorical it exhibits probabilistic variation:

\[
\text{(pró.fes.so)(ril.la)} \sim \text{(pró.fes)(sò.ril)la} \quad \text{‘professor-ADESSIVE’}
\]

- The “low” vowels /a, ä, o, ö/ attract stress.
- The “high” vowels /e, i, u, y/ repel stress.
- Stress is avoided next to a heavy syllable.
Finnish stress is governed by four core rules:

- Primary stress falls on the initial syllable
- Secondary stress falls on every other syllable after that...
- except that a light syllable is skipped if the syllable after that is heavy
- unless that heavy syllable is final


We focus on the “skipping clause” and note that it leaks in long words: instead of being categorical it exhibits probabilistic variation

\[(\text{pró.fes.so})(\text{ril}.la) \sim (\text{pró}.fes)(\text{sò}.ril)la\]

‘professor-ADESSIVE’

Skipping depends on vowel quality and preceding syllable weight in a non-categorical/probabilistic manner

- the “low” vowels /a, ä, o, ö/ attract stress
- the “high” vowels /e, i, u, y/ repel stress
- stress is avoided next to a heavy syllable

[Anttila 2012]
But how can we measure the effect of these additional conditions on the frequency of skipping given that secondary stress is hard to hear?
But how can we measure the effect of these additional conditions on the frequency of skipping given that secondary stress is hard to hear?

Fortunately there is a segmental alternation that we can use as a stress diagnostic and that is present even in the written standard language:

\[
\text{a short underlying } t \text{ is deleted if it is extrametrical}
\]

[Keyser and Kiparsky 1984; Anttila 2012]
But how can we measure the effect of these additional conditions on the frequency of skipping given that secondary stress is hard to hear?

Fortunately there is a segmental alternation that we can use as a stress diagnostic and that is present even in the written standard language:

- A short underlying $t$ is deleted if it is extrametrical

In other words, we have the following correspondences:

- no skipping $\iff t$-deletion $(\text{pró.fes})(\text{sò.re})ja$
- skipping $\iff t$-retention $(\text{pró.fes.so})(\text{rèi.ta})$

[Keyser and Kiparsky 1984; Anttila 2012]
To model this distribution of Finnish secondary stress, we considered an input space consisting of 48 types of nouns obtained by varying systematically stem length, syllable weight, and vowel sonority.
To model this distribution of Finnish secondary stress, we considered an input space consisting of 48 types of nouns obtained by varying systematically stem length, syllable weight, and vowel sonority.

These phonological forms are evaluated by a set of eight constraints:

- **WSP**: No unstressed heavy syllables
- **WSP/VV**: No unstressed heavy syllables with a long vowel
- **FTBIN**: Feet are disyllabic
- **PKPROM**: No stressed light syllables
- **ALIGN-L**: All feet left
- ***REV**: No trochees with sonority reversal = *(Í.A)
- ***FLAT**: No trochees with imperfect sonority = *(Í.I, *(Á.A, *(Í.A)
- ***H.X**: No stress next to a heavy syllable
We computed SHG/ME uniform probability inequalities for this model using CoGeTo: a new suite of tools for studying categorical and probabilistic constraint-based typologies based on their rich underlying Convex Geometry

[Magri and Anttila 2019: https://cogeto.stanford.edu]
We computed SHG/ME uniform probability inequalities for this model using CoGeTo: a new suite of tools for studying categorical and probabilistic constraint-based typologies based on their rich underlying Convex Geometry [Magri and Anttila 2019: https://cogeto.stanford.edu]

SGH predicts seven blocks of equiprobable mappings ordered through uniform probability inequalities (denoted $\leq$):

- $(\text{kon.sul})(\text{taa.ti.o})ja$ $\leq$ $(\text{sym.po})(\text{si.u.me})ja$
- $(\text{kom.mu})(\text{ni.ke.o})ja$ $\leq$ $(\text{po.ly})(\text{a.mi.de})ja$
- $(\text{o.pe})(\text{raa.ti.o})ja$ $\leq$ $(\text{in.ku})(\text{naa.be.le})ja$
- $(\text{al.le})(\text{go.ri.o})ja$ $\leq$ $(\text{lii.rum})(\text{laa.ru.me})ja$
- $(\text{pro.pa})(\text{gan.dis.te})ja$ $\leq$ $(\text{pro.pa})(\text{gan.dis})(\text{tei.ta})$
- $(\text{ak.va})(\text{rel.lis.te})ja$ $\leq$ $(\text{ak.va})(\text{rel.lis})(\text{tei.ta})$
- $(\text{ter.mos})(\text{taat.te})ja$ $\leq$ $(\text{mar.ga})(\text{rri.pe})ja$
- $(\text{pro.pa})(\text{gan.dis})(\text{tei.ta})$ $\leq$ $(\text{pro.pa})(\text{gan.dis})(\text{tei.ta})$
- $(\text{ak.va})(\text{rel.lis})(\text{tei.ta})$ $\leq$ $(\text{ak.va})(\text{rel.lis})(\text{tei.ta})$
- $(\text{mar.ga})(\text{rri.pe})(\text{kaat.to})ja$ $\leq$ $(\text{mar.ga})(\text{rri.pe})(\text{kaat.to})ja$
- $(\text{ter.mos})(\text{taat.te})ja$ $\leq$ $(\text{mar.ga})(\text{rri.pe})(\text{kaat.to})ja$
□ We computed SHG/ME uniform probability inequalities for this model using **CoGeTo**: a new suite of **To**ols for studying categorical and probabilistic constraint-based typologies based on their rich underlying Convex Geometry  

[Magri and Anttila 2019: https://cogeto.stanford.edu]

□ SGH predicts seven blocks of equiprobable mappings ordered through uniform probability inequalities (denoted \( \leq \)):

\[
\begin{align*}
(kon.sul)(taa.ti.o)ja & \leq (sym.po)(si.u.me)ja \\
(kom.mu)(ni.ke.o)ja & \leq (po.ly)(a.mi.de)ja \\
(o.pe)(raa.ti.o)ja & \leq (in.ku)(naa.be.le)ja \\
(al.le)(go.r.i.o)ja & \leq (lii.rum)(laa.ru.me)ja
\end{align*}
\]

□ This confirms on a naturalistic example the formal result in the first part of the talk, that SHG indeed allows for equiprobable mappings
To start evaluating empirically these predicted equiprobable blocks, we examined Finnish t-deletion in a corpus of approximately 9 million nouns (tokens) harvested from Finnish internet sites on April 12, 2005.
To start evaluating empirically these predicted equiprobable blocks, we examined Finnish *t*-deletion in a corpus of approximately 9 million nouns (tokens) harvested from Finnish internet sites on April 12, 2005.

We repeat the equiprobable blocks annotated with token frequency of *t*-retention/deletion for each stem type (type frequencies are similar).

<table>
<thead>
<tr>
<th>Block</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>(sym.po)(si.u.me)ja</td>
<td>98.6%</td>
</tr>
<tr>
<td>(po.ly)(a.mi.de)ja</td>
<td>95.7%</td>
</tr>
<tr>
<td>(in.ku)(naa.be.le)ja</td>
<td>9.5%</td>
</tr>
<tr>
<td>(lii.rum)(laa.ru.me)ja</td>
<td>18.6%</td>
</tr>
<tr>
<td>(pro.pa)(gan.dis.te)ja</td>
<td>100%</td>
</tr>
<tr>
<td>(ak.va)(rel.lis.te)ja</td>
<td>100%</td>
</tr>
<tr>
<td>(ter.mos)(taat.te)ja</td>
<td>100%</td>
</tr>
<tr>
<td>(mar.ga)(rii.ne)ja</td>
<td>100%</td>
</tr>
<tr>
<td>(af.fri)(kaat.to)ja</td>
<td>99.7%</td>
</tr>
<tr>
<td>(al.le)(go.ri)(oi)ta</td>
<td>100%</td>
</tr>
<tr>
<td>(kom.mu)(ni.ke)(oi)ta</td>
<td>99.7%</td>
</tr>
<tr>
<td>(kon.sul)(taa.ti)(oi)ta</td>
<td>99.5%</td>
</tr>
<tr>
<td>(o.pe)(raa.ti)(oi)ta</td>
<td>100%</td>
</tr>
<tr>
<td>(sym.po)(si.u)(mei)ta</td>
<td>1.4%</td>
</tr>
<tr>
<td>(in.ku)(naa.be)(lei)ta</td>
<td>90.5%</td>
</tr>
<tr>
<td>(pro.pa)(gan.dis)(tei)ta</td>
<td>0.0%</td>
</tr>
<tr>
<td>(ak.va)(rel.lis)(tei)ta</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Observation 1: the SHG equiprobability prediction is consistent with the data in the 5 black blocks (all stems are nearly categorical).

Observation 2: the SHG equiprobability prediction is challenged in the 2 red blocks, that we now focus on.
To start evaluating empirically these predicted equiprobable blocks, we examined Finnish $t$-deletion in a corpus of approximately 9 million nouns (tokens) harvested from Finnish internet sites on April 12, 2005.

We repeat the equiprobable blocks annotated with token frequency of $t$-retention/deletion for each stem type (type frequencies are similar).

Observation 1: the SHG equiprobability prediction is consistent with the data in the 5 black blocks (all stems are nearly categorical).
To start evaluating empirically these predicted equiprobable blocks, we examined Finnish t-deletion in a corpus of approximately 9 million nouns (tokens) harvested from Finnish internet sites on April 12, 2005.

We repeat the equiprobable blocks annotated with token frequency of t-retention/deletion for each stem type (type frequencies are similar).

**Observation 1**: the SHG equiprobability prediction is consistent with the data in the 5 black blocks (all stems are nearly categorical).

**Observation 2**: the SHG equiprobability prediction is challenged in the 2 red blocks, that we now focus on.
These two problematic red blocks involve four stem types:

\textit{lii.rum.laa.ru.mi \ in.ku.naa.be.li \ sym.po.si.u.mi \ po.ly.a.mi.di}

that differ in their preantepenultimate and antepenultimate syllables (heavy vs. light, high vs. low) and yet SHG predicts they should undergo \textit{t}-deletion/retention at identical rates.
These two problematic red blocks involve four stem types

\textit{lii.rum.laa.ru.mi} \textit{in.ku.naa.be.li} \textit{sym.po.si.u.mi} \textit{po.ly.a.mi.di}

that differ in their preantepenultimate and antepenultimate syllables (heavy vs. light, high vs. low) and yet SHG predicts they should undergo \textit{t}-deletion/retention at identical rates

But the difference between \textit{t}-deletion/retention rates for stems of the \textit{liirumlaarumi}-type and \textit{inkunaabeli}-type is not statistically significant ($\chi^2 = 2.9849$, $df = 1$, $p = 0.08404$)
These two problematic red blocks involve four stem types

\[
\text{lii.rum.laa.ru.mi} \quad \text{in.ku.naa.be.li} \quad \text{sym.po.si.u.mi} \quad \text{po.ly.a.mi.di}
\]

that differ in their preantepenultimate and antepenultimate syllables (heavy vs. light, high vs. low) and yet SHG predicts they should undergo \( t \)-deletion/retention at identical rates

But the difference between \( t \)-deletion/retention rates for stems of the \text{liirumlaarumi}-type and \text{inkunaabeli}-type is not statistically significant (\( \chi^2 = 2.9849, \ df = 1, \ p = 0.08404 \))

Furthermore, there are only two stems in the \text{symposiumi}-type (\text{symposiumi, imperiumi}) and both could be re-analyzed as 4-syllable stems, consistently with high \( t \)-deletion rate

---

Admittedly, we have no plausible explanation for the unexpectedly high \( t \)-deletion rate for stem of the \text{polyamidi}-type (\( N = 69 \))

We conclude that the Finnish data largely (although admittedly not completely) support SHG's equiprobability predictions
These two problematic red blocks involve four stem types

\[
\text{lii.rum.laa.ru.mi} \quad \text{in.ku.naa.be.li} \quad \text{sym.po.si.u.mi} \quad \text{po.ly.a.mi.di}
\]

that differ in their preantepenultimate and antepenultimate syllables (heavy vs. light, high vs. low) and yet SHG predicts they should undergo \( t \)-deletion/retention at identical rates

But the difference between \( t \)-deletion/retention rates for stems of the \text{liirumlaarum}i-type and \text{inkunaabeli}-type is not statistically significant \((\chi^2 = 2.9849, df = 1, p = 0.08404)\)

Furthermore, there are only two stems in the \textit{symposiumi}-type \((\text{symposiumi, imperiumi})\) and both could be re-analyzed as 4-syllable stems, consistently with high \( t \)-deletion rate

Admittedly, we have no plausible explanation for the unexpectedly high \( t \)-deletion rate for stem of the \text{polyamidi}-type \((N = 69)\)
These two problematic red blocks involve four stem types:

\[ \text{lli.rum.laa.ru.mi, in.ku.naa.be.li, sym.po.si.u.mi, po.ly.a.mi.di} \]

that differ in their preantepenultimate and antepenultimate syllables (heavy vs. light, high vs. low) and yet SHG predicts they should undergo \( t \)-deletion/retention at identical rates.

But the difference between \( t \)-deletion/retention rates for stems of the \( \text{lliirumlaarumi} \)-type and \( \text{inkunaabeli} \)-type is not statistically significant \((\chi^2 = 2.9849, df = 1, p = 0.08404)\).

Furthermore, there are only two stems in the \( \text{symposiumi} \)-type \((\text{symposiumi, imperiumi})\) and both could be re-analyzed as 4-syllable stems, consistently with high \( t \)-deletion rate [Anttila and Shapiro 2017].

Admittedly, we have no plausible explanation for the unexpectedly high \( t \)-deletion rate for stem of the \( \text{polyamidi} \)-type \((N = 69)\).

We conclude that the Finnish data largely (although admittedly not completely) support SHG’s equiprobability predictions.
Does ME with its ability to make fine-grained distinctions offer a more principled solution to the difficulties just encountered with the two red blocks? This turns out not to be the case.

On the retention side (top), ME's predictions seem promising: corpus frequencies abide by the predicted probability inequalities. On the deletion side (bottom) though, ME reverses the probabilities, yielding exactly the opposite of what we observe in the data. We submit that there is simply no way to reconcile ME's counterintuitive probability reversals with the corpus data.
Does ME with its ability to make fine-grained distinctions offer a more principled solution to the difficulties just encountered with the two red blocks? This turns out not to be the case.

In fact, as expected given the formal result in the first part of the talk, ME breaks up these two red equiprobable blocks and orders their stem types through uniform probability inequalities:

\[
\begin{align*}
\text{(sym.po)(si.u)(mei.ta)} & \leq \text{(po.ly)(a.mi)(dei.ta)} \leq \text{(lii.rum)(laa.ru)(mei.ta)} \leq \text{(in.ku)(naa.be)(lei.ta)}, \\
\text{(sym.po)(si.u.me)ja} & \leq \text{(po.ly)(a.mi.de)ja} \leq \text{(lii.rum)(laa.ru.me)ja} \leq \text{(in.ku)(naa.be.le)ja}.
\end{align*}
\]
Does ME with its ability to make fine-grained distinctions offer a more principled solution to the difficulties just encountered with the two red blocks? This turns out not to be the case.

In fact, as expected given the formal result in the first part of the talk, ME breaks up these two red equiprobable blocks and orders their stem types through uniform probability inequalities:

\[
\begin{align*}
\text{(sym.po)(si.u)(mei.ta)} & \leq \text{(po.ly)(a.mi)(dei.ta)} & \leq \text{(lii.rum)(laa.ru)(mei.ta)} & \leq \text{(in.ku)(naa.be)(lei.ta)} \\
\text{(sym.po)(si.u.me)ja} & \leq \text{(po.ly)(a.mi.de)ja} & \leq \text{(lii.rum)(laa.ru.me)ja} & \leq \text{(in.ku)(naa.be.le)ja}
\end{align*}
\]

On the retention side (top), ME’s predictions seem promising: corpus frequencies abide by the predicted probability inequalities.
Does ME with its ability to make fine-grained distinctions offer a more principled solution to the difficulties just encountered with the two red blocks? This turns out not to be the case.

In fact, as expected given the formal result in the first part of the talk, ME breaks up these two red equiprobable blocks and orders their stem types through uniform probability inequalities:

\[
\begin{align*}
(sym.po)(si.u)(mei.ta) & \leq (po.ly)(a.mi)(dei.ta) & \leq (lii.rum)(laa.ru)(mei.ta) & \leq (in.ku)(naa.be)(lei.ta) \\
(sym.po)(si.u.me)ja & \leq (po.ly)(a.mi.de)ja & \leq (lii.rum)(laa.ru.me)ja & \leq (in.ku)(naa.be.le)ja
\end{align*}
\]

On the **retention side** (top), ME’s predictions seem promising: corpus frequencies abide by the predicted probability inequalities.

On the **deletion side** (bottom) though, ME reverses the probabilities, yielding exactly the opposite of what we observe in the data.
Does ME with its ability to make fine-grained distinctions offer a more principled solution to the difficulties just encountered with the two red blocks? This turns out not to be the case.

In fact, as expected given the formal result in the first part of the talk, ME breaks up these two red equiprobable blocks and orders their stem types through uniform probability inequalities:

\[
\begin{align*}
\text{(sym.po)(si.u)(mei.ta)} & \leq \text{(po.ly)(a.mi)(dei.ta)} & \leq \text{(lii.rum)(laa.ru)(mei.ta)} & \leq \text{(in.ku)(naa.be)(lei.ta)} \\
\text{(sym.po)(si.u.me)ja} & \leq \text{(po.ly)(a.mi.de)ja} & \leq \text{(lii.rum)(laa.ru.me)ja} & \leq \text{(in.ku)(naa.be.le)ja}
\end{align*}
\]

On the retention side (top), ME’s predictions seem promising: corpus frequencies abide by the predicted probability inequalities.

On the deletion side (bottom) though, ME reverses the probabilities, yielding exactly the opposite of what we observe in the data.

We submit that there is simply no way to reconcile ME’s counterintuitive probability reversals with the corpus data.
Wrapping up:

- We have shown that ME can distinguish between any two different mappings and thus admits no equiprobable mappings.

In contrast, SHG admits equiprobable mappings, that we can explicitly characterize and efficiently compute through...

Our Finnish corpus data provide (admittedly only preliminary) evidence in favor of SHG's equiprobability predictions against ME.
Wrapping up:

- We have shown that ME can distinguish between any two different mappings and thus admits no equiprobable mappings.

- In contrast, SHG admits equiprobable mappings, that we can explicitly characterize and efficiently compute through CoGeTo.
Wrapping up:

- We have shown that ME can distinguish between any two different mappings and thus admits no equiprobable mappings.

- In contrast, SHG admits equiprobable mappings, that we can explicitly characterize and efficiently compute through CoGeTo.

- Our Finnish corpus data provide (admittedly only preliminary) evidence in favor of SHG’s equiprobability predictions against ME.
Wrapping up:

- We have shown that ME can distinguish between any two different mappings and thus admits no equiprobable mappings.
- In contrast, SHG admits equiprobable mappings, that we can explicitly characterize and efficiently compute through CoGeTo.
- Our Finnish corpus data provide (admittedly only preliminary) evidence in favor of SHG’s equiprobability predictions against ME.

Grazie!/Thank you!


