Equiprobable mappings in weighted constraint grammars

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

ACL Sigmorphon 2019

Firenze | August 2nd, 2019

FRANCE-STANFORD CENTER FOR INTERDISCIPLINARY STUDIES

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)

[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]

- Both SHG and ME typologies of probabilistic grammars are the result of "smoothing" an underlying typology of **categorical** HG grammars

- 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]

- 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) [Anttila and Magri 2018]

- 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]

- 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) [Anttila and Magri 2018]

- 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, [$\pm$back], but ignores, say, number of syllables: two mappings that only differ for length such as

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

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, [$\pm$back], but ignores, say, number of syllables: two mappings that only differ for length such as

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

  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 $\mathbb{P}([SR] \mid /UR/)$ over the corresponding set of candidate surface representations $[SR]$

- A **probabilistic** phonological grammar assigns to each underlying representation $/UR/$ a probability distribution $\mathbb{P}([SR] \mid /UR/)$ 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

$$\mathbb{P}\big([SR] \,\big|\, /UR/\big) = \mathbb{P}\big([\widehat{SR}] \,\big|\, /\widehat{UR}/\big)$$

- 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
$$\mathbb{P}\big([SR] \,\big|\, /UR/\big) = \mathbb{P}\big([\widehat{SR}] \,\big|\, /\widehat{UR}/\big)$$
- 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

$$\mathbb{P}\big([SR] \,\big|\, /UR/\big) = \mathbb{P}\big([\widehat{SR}] \,\big|\, /\widehat{UR}/\big)$$

- 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

$$\mathbb{P}\big(\text{[maana]} \,\big|\, \text{/maa-nä/}\big) = \mathbb{P}\big(\text{[rakastajana]} \,\big|\, \text{/rakastaja-nä/}\big)$$

2.
ME admits no equiprobable mappings

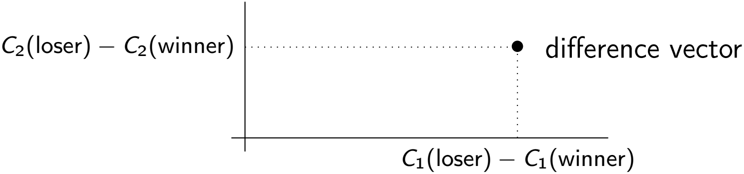
- Given a winner mapping and a competing loser mapping, we call **difference vector** the vector consisting of the constraint violations of the loser discounted of the violations of the winner

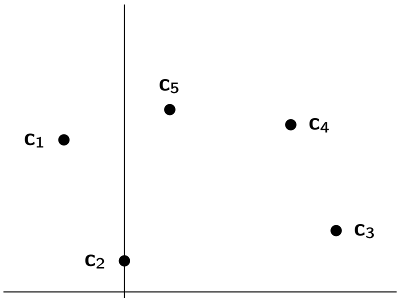
$$\text{difference vector} = \begin{bmatrix} C_1(\text{loser}) - C_1(\text{winner}) \\ C_2(\text{loser}) - C_2(\text{winner}) \\ \vdots \\ C_n(\text{loser}) - C_n(\text{winner}) \end{bmatrix}$$

- 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

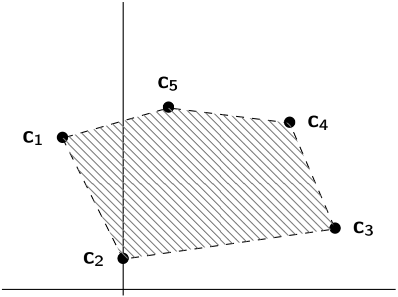
$$\text{difference vector} = \begin{bmatrix} C_1(\text{loser}) - C_1(\text{winner}) \\ C_2(\text{loser}) - C_2(\text{winner}) \\ \vdots \\ C_n(\text{loser}) - C_n(\text{winner}) \end{bmatrix}$$

- When there are only $n = 2$ constraints C_1 and C_2, difference vectors can be plotted as points in the Cartesian plane:

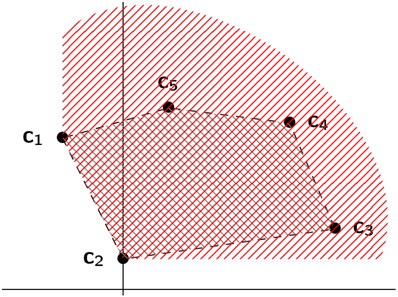




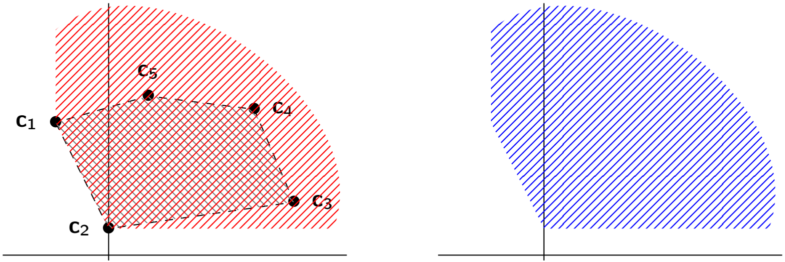
- Suppose the mapping $(/UR/, [SR])$ comes with 5 difference vectors



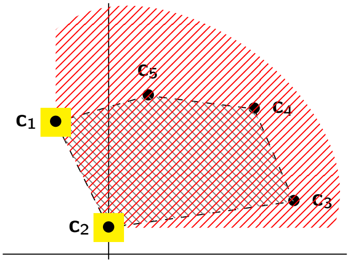
- Suppose the mapping $(/UR/, [SR])$ comes with 5 difference vectors
- The gray region is the **convex hull** generated by these vectors

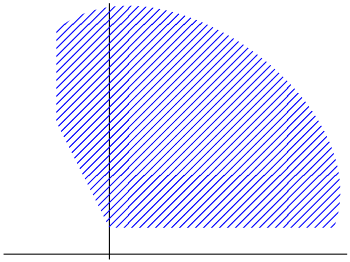


- Suppose the mapping $(/UR/, [SR])$ comes with 5 difference vectors
- The gray region is the **convex hull** generated by these vectors
- The red region consists of points larger than a point in this convex hull

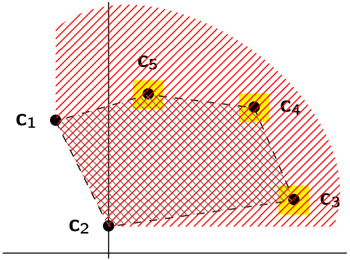


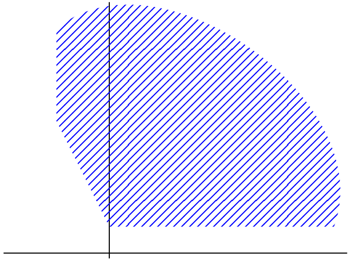
- Two mappings $(/\color{red}{UR}/,[\color{red}{SR}])$ and $(/\color{blue}{\widehat{UR}}/,[\color{blue}{\widehat{SR}}])$ are ME equiprobable if and only if the corresponding red and blue regions coincide



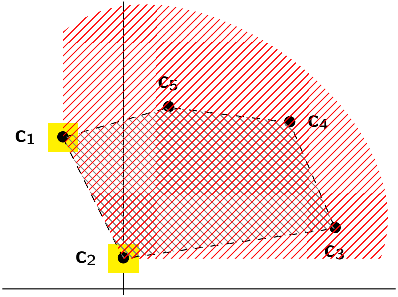


- Two mappings $(/\textcolor{red}{UR}/, [\textcolor{red}{SR}])$ and $(/\textcolor{blue}{\widehat{UR}}/, [\textcolor{blue}{\widehat{SR}}])$ are ME equiprobable if and only if the corresponding red and blue regions coincide
- The difference vectors $\mathbf{c}_1$ and $\mathbf{c}_2$ are **extreme points**: they determine the shape of the region and must be shared by the two mappings

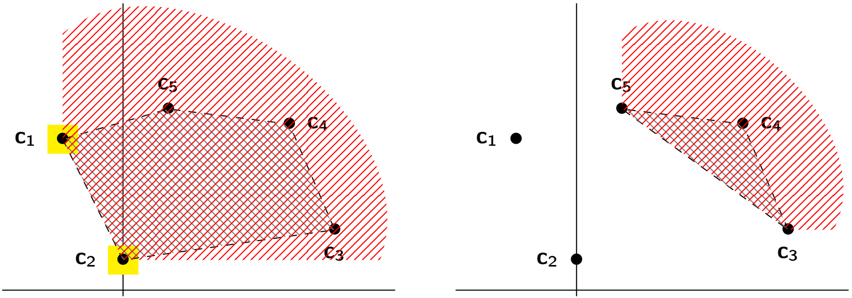




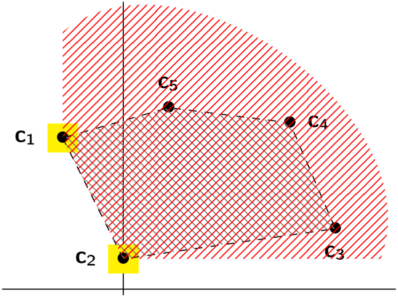
- Two mappings $(/UR/, [SR])$ and $(/\widehat{UR}/, [\widehat{SR}])$ are ME equiprobable if and only if the corresponding red and blue regions coincide
- The difference vectors $\mathbf{c}_1$ and $\mathbf{c}_2$ are **extreme points**: they determine the shape of the region and must be shared by the two mappings
- The difference vectors $\mathbf{c}_3$, $\mathbf{c}_4, \mathbf{c}_5$ are instead **interior points**: they do not contribute to the shape of the region

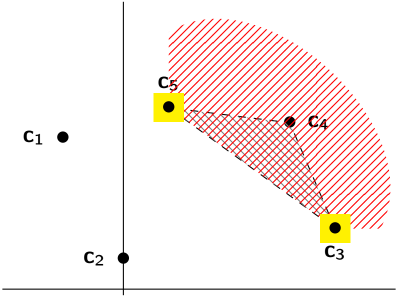


- Yet, since we have established that $\mathbf{c}_1$ and $\mathbf{c}_2$ are shared, we can effectively "peel them off" the two sides of the ME probability identity

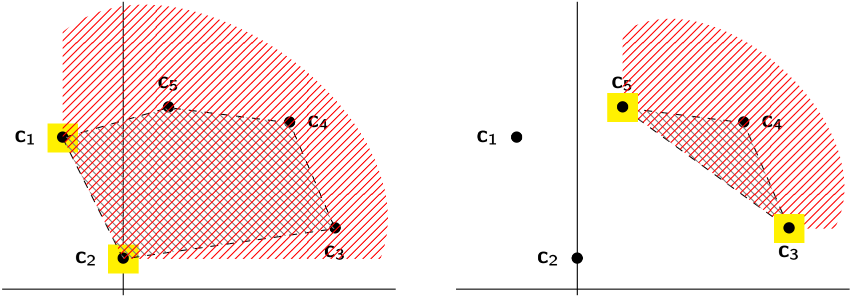


- Yet, since we have established that $\mathbf{c}_1$ and $\mathbf{c}_2$ are shared, we can effectively "peel them off" the two sides of the ME probability identity
- In other words, we can ignore $\mathbf{c}_1$ and $\mathbf{c}_2$ and only consider the remaining difference vectors $\mathbf{c}_3, \mathbf{c}_4, \mathbf{c}_5$





- Yet, since we have established that $\mathbf{c}_1$ and $\mathbf{c}_2$ are shared, we can effectively "peel them off" the two sides of the ME probability identity
- In other words, we can ignore $\mathbf{c}_1$ and $\mathbf{c}_2$ and only consider the remaining difference vectors $\mathbf{c}_3, \mathbf{c}_4, \mathbf{c}_5$
- Now $\mathbf{c}_3$ and $\mathbf{c}_5$ are extreme points of the new red region and must be shared by the equiprobable mappings $(/UR/, [SR])$ and $(/\widehat{UR}/, [\widehat{SR}])$



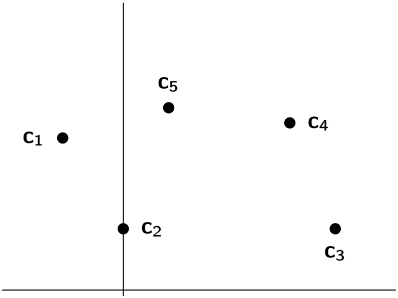
- Yet, since we have established that $\mathbf{c}_1$ and $\mathbf{c}_2$ are shared, we can effectively "peel them off" the two sides of the ME probability identity
- In other words, we can ignore $\mathbf{c}_1$ and $\mathbf{c}_2$ and only consider the remaining difference vectors $\mathbf{c}_3, \mathbf{c}_4, \mathbf{c}_5$
- Now $\mathbf{c}_3$ and $\mathbf{c}_5$ are extreme points of the new red region and must be shared by the equiprobable mappings $(/UR/, [SR])$ and $(/\widehat{UR}/, [\widehat{SR}])$
- And so on

- The conclusion of this reasoning is that the two mappings $(/\textcolor{red}{UR}/, [\textcolor{red}{SR}])$ and $(/\widehat{\textcolor{blue}{UR}}/, [\widehat{\textcolor{blue}{SR}}])$ are equiprobable in ME if and only if the corresponding sets of difference vectors exactly coincide

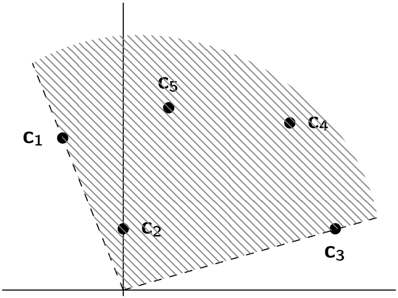
- The conclusion of this reasoning is that the two mappings $(/UR/, [SR])$ and $(/\widehat{UR}/, [\widehat{SR}])$ are equiprobable in ME if and only if the corresponding sets of difference vectors exactly coincide
- Realistically, the difference vectors coincide only if the two mappings $(/UR/, [SR])$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

- The conclusion of this reasoning is that the two mappings $(/UR/, [SR])$ and $(/\widehat{UR}/, [\widehat{SR}])$ 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 $(/\widehat{UR}/, [\widehat{SR}])$ are actually the same mapping
- $\boxed{\text{In conclusion, ME effectively admits no equiprobable mappings}}$

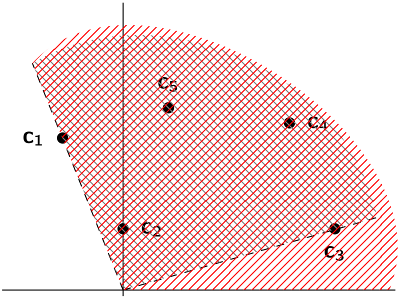
3.
SHG instead does admit equiprobable mappings



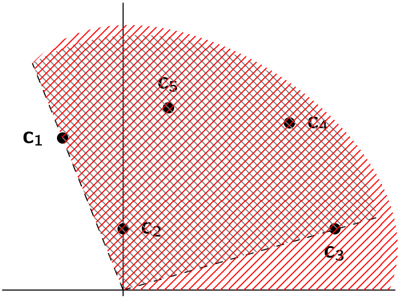
- Suppose the mapping $(/\textcolor{red}{UR}/, [\textcolor{red}{SR}])$ comes with 5 difference vectors



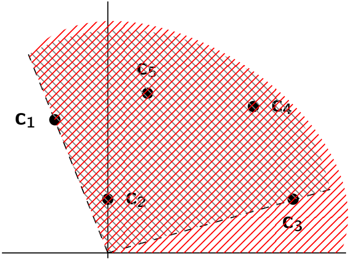
- Suppose the mapping $(/UR/, [SR])$ comes with 5 difference vectors
- The gray region is the **convex cone** generated by these vectors

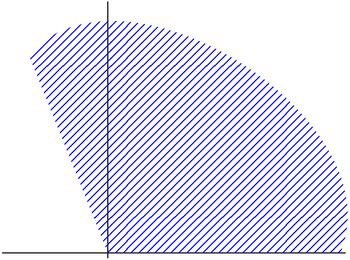


- Suppose the mapping $(/UR/, [SR])$ comes with 5 difference vectors
- The gray region is the **convex cone** generated by these vectors
- The red region consists of points larger than a point in this cone

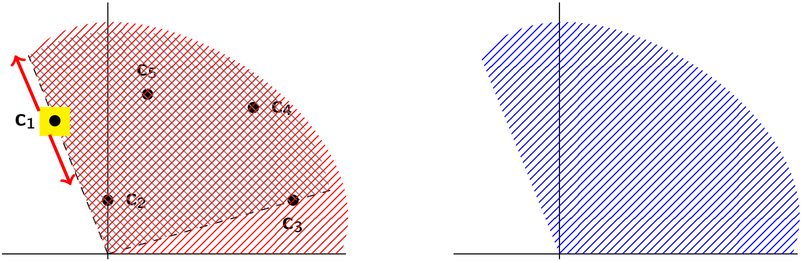


- Suppose the mapping ($/UR/, [SR]$) comes with 5 difference vectors
- The gray region is the **convex cone** generated by these vectors
- The red region consists of points larger than a point in this cone
- Same as for ME with cones ( ) rather than convex hulls ( )

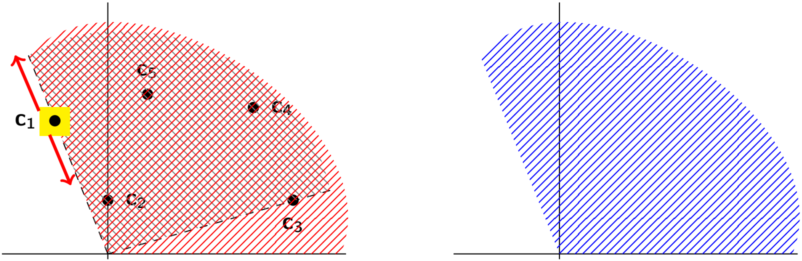




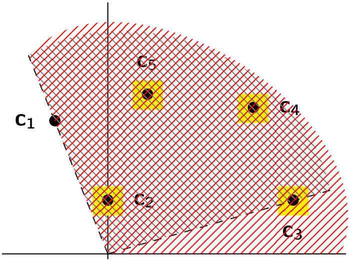
- Again, the mappings $(/UR/, [SR])$ and $(/\widehat{UR}/, [\widehat{SR}])$ are SHG equiprobable if and only if the red and blue regions coincide

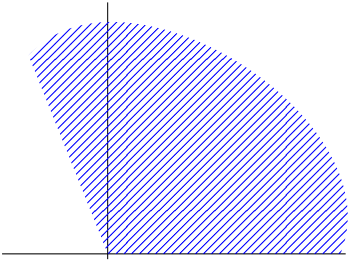


- Again, the mappings $(/\textcolor{red}{UR}/, [\textcolor{red}{SR}])$ and $(/\widehat{\textcolor{blue}{UR}}/, [\widehat{\textcolor{blue}{SR}}])$ are SHG equiprobable if and only if the red and blue regions coincide
- The difference vector $\mathbf{c}_1$ of the mapping $(/\textcolor{red}{UR}/, [\textcolor{red}{SR}])$ 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 $\mathbf{c}_1$ 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 $\mathbf{c}_1$ but only a rescaling thereof





- Again, the mappings $(/\color{red}{UR}/, [\color{red}{SR}])$ and $(/\color{blue}{\widehat{UR}}/, [\color{blue}{\widehat{SR}}])$ are SHG equiprobable if and only if the red and blue regions coincide
- The difference vector $\mathbf{c}_1$ of the mapping $(/\color{red}{UR}/, [\color{red}{SR}])$ sits on the border but can be shifted (rescaled) without affecting the red region
- The equiprobable mapping $(/\color{blue}{\widehat{UR}}/, [\color{blue}{\widehat{SR}}])$ thus needs *not* share this difference vector $\mathbf{c}_1$ but only a rescaling thereof
- And nothing can be said about the interior vectors $\mathbf{c}_2, \ldots, \mathbf{c}_5$

- 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

- 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

- 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

- $\boxed{\text{In conclusion, SHG does admit equiprobable mappings}}$

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

- 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

[Carlson 1978; Hanson and Kiparsky 1996; Elenbaas and Kager 1999; Kiparsky 2003; Karvonen 2005; Karttunen 2006]

- 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

[?Carlson 1978; Hanson and Kiparsky 1996; Elenbaas and Kager 1999; Kiparsky 2003; Karvonen 2005; Karttunen 2006]

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

*(**pró**.fes.so)(**rìl**.la)*	∼	*(**pró**.fes)(**sò**.ril)la*	'professor-ADESSIVE'
red syllable skipped		red syllable not skipped	

- 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

[?Carlson 1978; Hanson and Kiparsky 1996; Elenbaas and Kager 1999; Kiparsky 2003; Karvonen 2005; Karttunen 2006]

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

*(**pró**.fes.so)(**rìl**.la)*	∼	*(**pró**.fes)(**sò**.ril)la*	'professor-ADESSIVE'
red syllable skipped		red syllable not skipped	

- Skipping depends on vowel quality and preceding syllable weight in a non-categorical/probabilistic manner [Anttila 2012]
  - 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

- 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:

  a short underlying *t* is deleted if it is extrametrical

- 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:

		/professori-i-ta/ 'professor-PL-PART'
no skipping $\Longleftrightarrow$ *t*-deletion		(**pró**.fes)(**sò**.re)ja
skipping $\Longleftrightarrow$ *t*-retention		(**pró**.fes.so)(**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 $\mathbb{C}$o$\mathbb{G}$e$\mathbb{T}$o: a new suite of $\mathbb{T}$ools for studying categorical and probabilistic constraint-based typologies based on their rich underlying $\mathbb{C}$onvex $\mathbb{G}$eometry [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 $\leq$):

(kon.sul)(taa.ti.o)ja (kom.mu)(ni.ke.o)ja (o.pe)(raa.ti.o)ja (al.le)(go.ri.o)ja	$\leq$	(sym.po)(si.u.me)ja (po.ly)(a.mi.de)ja (in.ku)(naa.be.le)ja (lii.rum)(laa.ru.me)ja	$\leq$	(pro.pa)(gan.dis.te)ja (ak.va)(rel.lis.te)ja	$\leq$	(ter.mos)(taat.te)ja (mar.ga)(rii.ne)ja (af.fri)(kaat.to)ja
(pro.pa)(gan.dis)(tei.ta) (ak.va)(rel.lis)(tei.ta)	$\leq$	(po.ly)(a.mi)(dei.ta) (lii.rum)(laa.ru)(mei.ta) (sym.po)(si.u)(mei.ta) (in.ku)(naa.be)(lei.ta)	$\leq$	(al.le)(go.ri)(oi.ta) (kom.mu)(ni.ke)(oi.ta) (kon.sul)(taa.ti)(oi.ta) (o.pe)(raa.ti)(oi.ta)	$\leq$	

- 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 $\leq$):

(kon.sul)(taa.ti.o)ja
(kom.mu)(ni.ke.o)ja
(o.pe)(raa.ti.o)ja
(al.le)(go.ri.o)ja
$\leq$
(sym.po)(si.u.me)ja
(po.ly)(a.mi.de)ja
(in.ku)(naa.be.le)ja
(lii.rum)(laa.ru.me)ja
$\leq$
(pro.pa)(gan.dis.te)ja
(ak.va)(rel.lis.te)ja
$\leq$
(ter.mos)(taat.te)ja
(mar.ga)(rii.ne)ja
(af.fri)(kaat.to)ja

(pro.pa)(gan.dis)(tei.ta)
(ak.va)(rel.lis)(tei.ta)
$\leq$
(po.ly)(a.mi)(dei.ta)
(lii.rum)(laa.ru)(mei.ta)
(sym.po)(si.u)(mei.ta)
(in.ku)(naa.be)(lei.ta)
$\leq$
(al.le)(go.ri)(oi.ta)
(kom.mu)(ni.ke)(oi.ta)
(kon.sul)(taa.ti)(oi.ta)
(o.pe)(raa.ti)(oi.ta)
$\leq$
(ter.mos)(taat.te)ja
(mar.ga)(rii.ne)ja
(af.fri)(kaat.to)ja

- This confirms on a naturalistic example the formal result in the first pat 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)

(kon.sul)(taa.ti.o)ja 0.5%
(kom.mu)(ni.ke.o)ja 0.3%
(o.pe)(raa.ti.o)ja 0.0%
(al.le)(go.ri.o)ja 0.0%

$\leq$

(sym.po)(si.u.me)ja 98.6%
(po.ly)(a.mi.de)ja 95.7%
(in.ku)(naa.be.le)ja 9.5%
(lii.rum)(laa.ru.me)ja 18.6%

$\leq$

(pro.pa)(gan.dis.te)ja 100%
(ak.va)(rel.lis.te)ja 100%

$\leq$

(ter.mos)(taat.te)ja 100%
(mar.ga)(rii.ne)ja 100%
(af.fri)(kaat.to)ja 99.7%

$\geq$

(al.le)(go.ri)(oi.ta) 100%
(kom.mu)(ni.ke)(oi.ta) 99.7%
(kon.sul)(taa.ti)(oi.ta) 99.5%
(o.pe)(raa.ti)(oi.ta) 100%

$\geq$

(po.ly)(a.mi)(dei.ta) 4.3%
(lii.rum)(laa.ru)(mei.ta) 81.4%
(sym.po)(si.u)(mei.ta) 1.4%
(in.ku)(naa.be)(lei.ta) 90.5%

$\geq$

(pro.pa)(gan.dis)(tei.ta) 0.0%
(ak.va)(rel.lis)(tei.ta) 0.0%

- 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)

(kon.sul)(taa.ti.o)ja 0.5% (kom.mu)(ni.ke.o)ja 0.3% (o.pe)(raa.ti.o)ja 0.0% (al.le)(go.ri.o)ja 0.0%	$\leq$	(sym.po)(si.u.me)ja 98.6% (po.ly)(a.mi.de)ja 95.7% (in.ku)(naa.be.le)ja 9.5% (lii.rum)(laa.ru.me)ja 18.6%	$\leq$	(pro.pa)(gan.dis.te)ja 100% (ak.va)(rel.lis.te)ja 100%	$\leq$	(ter.mos)(taat.te)ja 100% (mar.ga)(rii.ne)ja 100% (af.fri)(kaat.to)ja 99.7%
(pro.pa)(gan.dis)(tei.ta) 0.0% (ak.va)(rel.lis)(tei.ta) 0.0%	$\leq$	(po.ly)(a.mi)(dei.ta) 4.3% (lii.rum)(laa.ru)(mei.ta) 81.4% (sym.po)(si.u)(mei.ta) 1.4% (in.ku)(naa.be)(lei.ta) 90.5%	$\leq$	(al.le)(go.ri)(oi.ta) 100% (kom.mu)(ni.ke)(oi.ta) 99.7% (kon.sul)(taa.ti)(oi.ta) 99.5% (o.pe)(raa.ti)(oi.ta) 100%	$\leq$	

- **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)

(kon.sul)(taa.ti.o)ja 0.5%
(kom.mu)(ni.ke.o)ja 0.3%
(o.pe)(raa.ti.o)ja 0.0%
(al.le)(go.ri.o)ja 0.0%

$\leq$

(sym.po)(si.u.me)ja 98.6%
(po.ly)(a.mi.de)ja 95.7%
(in.ku)(naa.be.le)ja 9.5%
(lii.rum)(laa.ru.me)ja 18.6%

$\leq$

(pro.pa)(gan.dis.te)ja 100%
(ak.va)(rel.lis.te)ja 100%

$\leq$

(ter.mos)(taat.te)ja 100%
(mar.ga)(rii.ne)ja 100%
(af.fri)(kaat.to)ja 99.7%

(pro.pa)(gan.dis)(tei.ta) 0.0%
(ak.va)(rel.lis)(tei.ta) 0.0%

$\leq$

(po.ly)(a.mi)(dei.ta) 4.3%
(lii.rum)(laa.ru)(mei.ta) 81.4%
(sym.po)(si.u)(mei.ta) 1.4%
(in.ku)(naa.be)(lei.ta) 90.5%

$\leq$

(al.le)(go.ri)(oi.ta) 100%
(kom.mu)(ni.ke)(oi.ta) 99.7%
(kon.sul)(taa.ti)(oi.ta) 99.5%
(o.pe)(raa.ti)(oi.ta) 100%

$\leq$

(ter.mos)(taat.te)ja 100% …

- **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
  *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 *t*-deletion/retention at identical rates

- These two problematic red blocks involve four stem types

  *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 *t*-deletion/retention at identical rates

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

- These two problematic red blocks involve four stem types

  *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 *t*-deletion/retention at identical rates
- But the difference between *t*-deletion/retention rates for stems of the *liirumlaarumi*-type and *inkunaabeli*-type is not statistically significant ($\chi^2 = 2.9849$, $df = 1$, $p = 0.08404$)
- Furthermore, there are only two stems in the *symposiumi*-type (*symposiumi*, *imperiumi*) and both could be re-analyzed as 4-syllable stems, consistently with high *t*-deletion rate [Anttila and Shapiro 2017]

- These two problematic red blocks involve four stem types

  *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 *t*-deletion/retention at identical rates
- But the difference between *t*-deletion/retention rates for stems of the *liirumlaarumi*-type and *inkunaabeli*-type is not statistically significant ($\chi^2 = 2.9849$, $df = 1$, $p = 0.08404$)
- Furthermore, there are only two stems in the *symposiumi*-type (*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 *polyamidi*-type ($N = 69$)

- These two problematic red blocks involve four stem types

  *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 *t*-deletion/retention at identical rates
- But the difference between *t*-deletion/retention rates for stems of the *liirumlaarumi*-type and *inkunaabeli*-type is not statistically significant ($\chi^2 = 2.9849$, $df = 1$, $p = 0.08404$)
- Furthermore, there are only two stems in the *symposiumi*-type (*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 *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

- 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:

(sym.po)(si.u)(mei.ta) 1.4%	$\leq$	(po.ly)(a.mi)(dei.ta) 4.3%	$\leq$	(lii.rum)(laa.ru)(mei.ta) 81.4%	$\leq$	(in.ku)(naa.be)(lei.ta) 90.5%
(sym.po)(si.u.me)ja 98.6%	$\leq$	(po.ly)(a.mi.de)ja 95.7%	$\leq$	(lii.rum)(laa.ru.me)ja 18.6%	$\leq$	(in.ku)(naa.be.le)ja 9.5%

- 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:

$\boxed{\text{(sym.po)(si.u)(mei.ta) } 1.4\%} \leq \boxed{\text{(po.ly)(a.mi)(dei.ta) } 4.3\%} \leq \boxed{\text{(lii.rum)(laa.ru)(mei.ta) } 81.4\%} \leq \boxed{\text{(in.ku)(naa.be)(lei.ta) } 90.5\%}$

$\boxed{\text{(sym.po)(si.u.me)ja } 98.6\%} \leq \boxed{\text{(po.ly)(a.mi.de)ja } 95.7\%} \leq \boxed{\text{(lii.rum)(laa.ru.me)ja } 18.6\%} \leq \boxed{\text{(in.ku)(naa.be.le)ja } 9.5\%}$

- 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:

(sym.po)(si.u)(mei.ta) 1.4% ≤ (po.ly)(a.mi)(dei.ta) 4.3% ≤ (lii.rum)(laa.ru)(mei.ta) 81.4% ≤ (in.ku)(naa.be)(lei.ta) 90.5%

(sym.po)(si.u.me)ja 98.6% ≤ (po.ly)(a.mi.de)ja 95.7% ≤ (lii.rum)(laa.ru.me)ja 18.6% ≤ (in.ku)(naa.be.le)ja 9.5%

- 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:

(sym.po)(si.u)(mei.ta) 1.4%	$\leq$	(po.ly)(a.mi)(dei.ta) 4.3%	$\leq$	(lii.rum)(laa.ru)(mei.ta) 81.4%	$\leq$	(in.ku)(naa.be)(lei.ta) 90.5%
(sym.po)(si.u.me)ja 98.6%	$\leq$	(po.ly)(a.mi.de)ja 95.7%	$\leq$	(lii.rum)(laa.ru.me)ja 18.6%	$\leq$	(in.ku)(naa.be.le)ja 9.5%

- 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

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 $\mathbb{C}\mathrm{o}\mathbb{G}\mathrm{e}\mathbb{T}\mathrm{o}$

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 $\mathbb{C}\mathrm{o}\mathbb{G}\mathrm{e}\mathbb{T}\mathrm{o}$
- 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 $\mathbb{C}\mathrm{o}\mathbb{G}\mathrm{e}\mathbb{T}\mathrm{o}$
- Our Finnish corpus data provide (admittedly only preliminary) evidence in favor of SHG's equiprobability predictions against ME

Grazie!/Thank you!

References I

Anttila, Arto. 2012. Modeling phonological variation. In *The oxford handbook of laboratory phonology*, ed. Abigail C. Cohn, Cécile Fougeron, and Marie Huffman, 76–91. Oxford: Oxford University Press.

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

Anttila, Arto, and Naomi Tachikawa Shapiro. 2017. The interaction of stress and syllabification: Serial or parallel? In *Proceedings of the 34th West Coast Conference on Formal Linguistics*, ed. Aaron Kaplan, Abby Kaplan, Miranda K. McCarvel, and Edward J. Rubin, 52–61. Somerville, MA, USA: Cascadilla Proceedings Project.

Boersma, Paul. 1997. How we learn variation, optionality and probability. In *Proceedings of the Institute of Phonetic Sciences (IFA) 21*, ed. Rob van Son, 43–58. University of Amsterdam: Institute of Phonetic Sciences.

Boersma, Paul, and Bruce Hayes. 2001. Empirical tests for the Gradual Learning Algorithm. *Linguistic Inquiry* 32:45–86.

Boersma, Paul, and Joe Pater. 2016. Convergence properties of a gradual learning algorithm for Harmonic Grammar. In *Harmonic Grammar and Harmonic Serialism*, ed. John McCarthy and Joe Pater. London: Equinox Press.

Carlson, Lauri. 1978. Word stress in Finnish. Massachusetts Institute of Technology, Cambridge, Massachusetts.

Elenbaas, Nine, and René Kager. 1999. Ternary rhythm and the lapse constraint. *Phonology* 16:273–329.

Goldwater, Sharon, and Mark Johnson. 2003. Learning OT constraint rankings using a Maximum Entropy model. In *Proceedings of the Stockholm Workshop on Variation Within Optimality Theory*, ed. Jennifer Spenader, Anders Eriksson, and Östen Dahl, 111–120. Stockholm University.

Hanson, Kristin, and Paul Kiparsky. 1996. A parametric theory of poetic meter. *Language* 72:287–335.

Hayes, Bruce. 2017. Varieties of Noisy Harmonic Grammar. In *Proceedings of the 2016 Annual Meeting in Phonology*, ed. Karen Jesney, Charlie O'Hara, Caitlin Smith, and Rachel Walker. Washington, DC: Linguistic Society of America.

Hayes, Bruce, and Colin Wilson. 2008. A Maximum Entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39:379–440.

Karttunen, Lauri. 2006. Word play. *Computational Linguistics* 33.4.

References II

Karvonen, Dan. 2005. Word prosody in finnish. Doctoral Dissertation, University of California, Santa Cruz.

Keyser, Samuel Jay, and Paul Kiparsky. 1984. Syllable structure in Finnish phonology. In *Language sound structure*, ed. Mark Aronoff and Richard Oehrle, 7–31. Cambridge, Massachusetts: MIT Press.

Kiparsky, Paul. 2003. Finnish noun inflection. In *Generative approaches to finnic and saami linguistics*, ed. Satu Manninen and Diane Nelson, 109–161. Stanford, California: CSLI Publications.

Magri, Giorgio, and Arto Anttila. 2019. $\mathbb{C}$o$\mathbb{G}$e$\mathbb{T}$o: Convex geometry tools for typological analysis in categorical and probabilistic constraint-based phonology (version 1.0). Available at `https://cogeto.stanford.edu`.

Smith, Brian W., and Joe Pater. 2017. French schwa and gradient cumulativity. Manuscript. University of California, Berkeley and University of Massachusetts, Amherst.

Zuraw, Kie, and Bruce Hayes. 2017. Intersecting constraint families: an argument for Harmonic Grammar. *Language* 93.3:497–546.