Learning prevents MaxEnt from giving probability to harmonically bounded candidates

Charlie O’Hara
University of Southern California

AMP 2019
October 13, 2019
A major goal of phonological theory is to develop a model that can capture the attested phonological patterns while not vastly over-predicting.

- Constraint based grammars (Optimality Theory\(^1\), Harmonic Grammar\(^2\), etc.) make strong typological predictions through **Factorial Typology**

- Recently, an abundance of work\(^3\) has investigated the hypothesis that learnability affects both categorical and soft typology.

---

\(^1\)Prince & Smolensky (1993/2004); McCarthy & Prince (1995)

\(^2\)Legendre *et al.* (1990); Pater (2016)

\(^3\)Boersma (2003); Pater & Moreton (2012); Staubs (2014); Hughto (2018); O’Hara (2017, in prep, 2018, 2019)
Introduction

A major goal of phonological theory is to develop a model that can capture the attested phonological patterns while not vastly over-predicting.

- Constraint based grammars (Optimality Theory\(^1\), Harmonic Grammar\(^2\), etc.) make strong typological predictions through **Factorial Typology**
- Recently, an abundance of work\(^3\) has investigated the hypothesis that learnability affects both categorical and soft typology.

---

\(^1\) Prince & Smolensky (1993/2004); McCarthy & Prince (1995)
\(^2\) Legendre *et al.* (1990); Pater (2016)
\(^3\) Boersma (2003); Pater & Moreton (2012); Staubs (2014); Hughto (2018); O’Hara (2017, in prep, 2018, 2019)
Restrictiveness and Learning

Grammatically Possible Patterns given $G$
Restrictiveness and Learning

Grammatically Possible Patterns given $G$
Restrictiveness and Learning

Grammatically Possible Patterns given $G_2$

Grammatically Possible Patterns given $G$
Restrictiveness and Learning
Harmonic Bounding

Simply Harmonically Bounded Candidates

A candidate is **SIMPLY HARMONICALLY BOUNDED** by another candidate if it has a proper superset of the violations of that candidate.

- /CV/ → [V] is simply harmonically bounded by /CV/ → [CV]

<table>
<thead>
<tr>
<th>/CV/</th>
<th>DEP</th>
<th>MAX</th>
<th>ONSET</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- a. CV
- b. V

- In Classic OT, Categorical HG, and Noisy HG, a harmonically bounded candidate will never surface.

- With these constraints, no ranking/weighting is able to find a pattern where onsets delete.
MaxEnt and Harmonically Bounded Candidates

But in MaxEnt, simply harmonically bounded candidates can receive probability (Jesney, 2007).

- As a result, MaxEnt can over-generate categorical (and noisy) HG (see also Anttila & Magri (2018)).
- As an example, MaxEnt generates a pattern where onsets variably delete.

\[
/\text{pa}/ \rightarrow \begin{array}{c} [\text{pa}] \\ [\text{a}] \end{array} \quad 50\% \quad /\text{u}/ \rightarrow \quad [\text{u}] \quad 100\%
\]

<table>
<thead>
<tr>
<th>/CV/</th>
<th>/V/</th>
<th>( w = 10 )</th>
<th>( w = 0 )</th>
<th>( w = 0 )</th>
<th>( w = 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{DEP} )</td>
<td>( \text{ONSET} )</td>
<td>( \text{MAX} )</td>
<td>( \text{HARM} )</td>
<td>( \text{PROB} )</td>
<td></td>
</tr>
<tr>
<td>a. CV</td>
<td>( -1 )</td>
<td>( -1 )</td>
<td>0</td>
<td>.5</td>
<td></td>
</tr>
<tr>
<td>b. V</td>
<td>( -1 )</td>
<td>( -1 )</td>
<td>0</td>
<td>.5</td>
<td></td>
</tr>
<tr>
<td>a. CV</td>
<td>( -1 )</td>
<td>( -10 )</td>
<td>( \sim 0 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. V</td>
<td>( -1 )</td>
<td>( -1 )</td>
<td>0</td>
<td>( \sim 1 )</td>
<td></td>
</tr>
</tbody>
</table>
Problem

MaxEnt can model grammars with probabilistic markedness reversals.

- It is impossible to restrict typology in MaxEnt using just the grammar.
- When all weights equal zero, all candidates for each input receive the same probability.
- Any input output mapping can receive probability.

Not ALL grammars are equally likely.

- Patterns that give substantial probability to harmonically bounded candidates are much harder to learn than patterns that do not.
- After learning is applied, MaxEnt does not severely overgenerate.
Patterns Under Investigation

There are two types of variation predicted by MaxEnt:

- **Normal Variation** - Most of the probability is split between candidates that could surface categorically.

  \[
  \text{Variable Onset Epenthesis}
  \]
  \[
  /\text{pa}/ \rightarrow \text{[pa]} \quad 100\% \quad /\text{u}/ \rightarrow \text{[?u]} \quad 50\% \quad /\text{u}/ \rightarrow \text{[u]} \quad 50\%
  \]

- **Harmonically-Bounded Variation** - Most of the probability is split between candidates some of which are harmonically bounded.

  \[
  \text{Variable Onset Deletion}
  \]
  \[
  /\text{pa}/ \rightarrow \text{[pa]} \quad 50\% \quad /\text{a}/ \quad 50\%\quad /\text{u}/ \rightarrow \text{[u]} \quad 100\%
  \]
Solution

I present learning simulations that show:

- Harmonically-bounded variation takes more data to learn than normal variation.
- Harmonically-bounded variation is less stably transmitted across generations.
- If both normal and harmonically bounded variation are found in a pattern, the harmonically-bounded variation will be lost first.

When filtered by learnability, MaxEnt is unlikely to give (much) probability to harmonically bounded candidates.
Relevant weighting condition

In Categorical HG:
- Onsets epenthesize when Onset outweighs Dep.
- Onsetless syllables remain faithful when Dep outweighs Onset.

In MaxEnt:
- Onsets epenthesize more when Onset outweighs Dep more.
- Onsetless syllables remain faithful more when Dep outweighs Onset more.
Categorical Harmonic Grammar

Onset Epenthesis

Onsetless Syllables allowed

<table>
<thead>
<tr>
<th>V</th>
<th>ONSET</th>
<th>DEP</th>
<th>HARMONY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weights</td>
<td>$w = 2$</td>
<td>$w = 1$</td>
<td></td>
</tr>
<tr>
<td>a. V</td>
<td>-1</td>
<td></td>
<td>-2</td>
</tr>
<tr>
<td>☛ b. CV</td>
<td>-1</td>
<td>-1</td>
<td></td>
</tr>
</tbody>
</table>
MaxEnt

Onset Epenthesis

Onsetless Syllables allowed

<table>
<thead>
<tr>
<th>/V/</th>
<th>ONSET</th>
<th>DEP</th>
<th>HARM</th>
<th>PROB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weights</td>
<td>$w = 2$</td>
<td>$w = 1$</td>
<td>HARM</td>
<td>PROB</td>
</tr>
<tr>
<td>a. V</td>
<td>-1</td>
<td></td>
<td>-2</td>
<td>.27</td>
</tr>
<tr>
<td>b. CV</td>
<td>-1</td>
<td>-1</td>
<td>.73</td>
<td></td>
</tr>
</tbody>
</table>
In Categorical HG:
- Onsets delete when $\text{Max} + \text{Onset}$ is lower than zero.
- Onsets are preserved when the sum of $\text{Max} + \text{Onset}$ is above zero.

In MaxEnt:
- Onsets delete more when $\text{Max} + \text{Onset}$ is lower.
- Onsets are preserved more when the sum of $\text{Max} + \text{Onset}$ is higher.
Harmonically-Bounded Variation

Harmonic Bounding - Categorical HG

<table>
<thead>
<tr>
<th>/CV/</th>
<th>Onset</th>
<th>Dep</th>
<th>Harmony</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weights</td>
<td>$w = .4$</td>
<td>$w = 1$</td>
<td>HARMONY</td>
</tr>
<tr>
<td>a. CV</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>b. V</td>
<td>-1</td>
<td>-.4</td>
<td></td>
</tr>
</tbody>
</table>
## Harmonic Bounding - MaxEnt

![Graph showing onsets and weights](image)

<table>
<thead>
<tr>
<th>/CV/</th>
<th>ONSET</th>
<th>DEP</th>
<th>HARM</th>
<th>PROB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weights</td>
<td>$w = .4$</td>
<td>$w = 1$</td>
<td>HARM</td>
<td>PROB</td>
</tr>
<tr>
<td>a. CV</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.60</td>
</tr>
<tr>
<td>b. V</td>
<td>-1</td>
<td>-.4</td>
<td>.40</td>
<td></td>
</tr>
</tbody>
</table>
Learning bias is a diachronic pressure

A child may not end up learning the same grammar as their parent
  - If the parent’s target grammar is harder to learn, the learner has higher probability of mislearning

Even a small probability of mislearning can have a large effect on typology over many generations.

By modeling generational transmission (Kirby & Huford, 2002; Staubs, 2014; Kirby, 2017; Hughto, 2018; O’Hara, in prep), we can observe effect of learning bias.
Simulation Methodology: Within-Generation

I use the truncated perceptron algorithm Magri (2015); Rosenblatt (1958); Boersma & Pater (2016).

- An input is randomly selected (here, a syllable structure).
- The teacher and learner select outputs for that input based on their current grammar.
- If they differ, the learner updates their grammar to make the teacher’s form more likely in the future.

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Learner</th>
</tr>
</thead>
<tbody>
<tr>
<td>/CV/</td>
<td>[CV] 100%</td>
</tr>
<tr>
<td>[V] 0%</td>
<td>[V] 0%</td>
</tr>
<tr>
<td>/V/</td>
<td>[CV] 50%</td>
</tr>
<tr>
<td>[V] 50%</td>
<td>[V] 50%</td>
</tr>
<tr>
<td>/V/</td>
<td>[CV] 75%</td>
</tr>
<tr>
<td>[V] 25%</td>
<td>[V] 25%</td>
</tr>
</tbody>
</table>
Simulation Methodology: Within-Generation

I use the truncated perceptron algorithm Magri (2015); Rosenblatt (1958); Boersma & Pater (2016).

- An input is randomly selected (here, a syllable structure).
- The teacher and learner select outputs for that input based on their current grammar.
- If they differ, the learner updates their grammar to make the teacher’s form more likely in the future.

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Learner</th>
</tr>
</thead>
<tbody>
<tr>
<td>/CV/</td>
<td>/V/</td>
</tr>
<tr>
<td>[CV] 100%</td>
<td>[CV] 100%</td>
</tr>
<tr>
<td>[V] 0%</td>
<td>[V] 0%</td>
</tr>
<tr>
<td>/V/</td>
<td>/V/</td>
</tr>
<tr>
<td>[CV] 50%</td>
<td>[CV] 75%</td>
</tr>
<tr>
<td>[V] 50%</td>
<td>[V] 25%</td>
</tr>
</tbody>
</table>
Simulation Methodology: Within-Generation

I use the truncated perceptron algorithm Magri (2015); Rosenblatt (1958); Boersma & Pater (2016).

- An input is randomly selected.
- The teacher and learner select outputs for that input based on their current grammar.
- If they differ, the learner updates their grammar to make the teacher’s form more likely in the future.

<table>
<thead>
<tr>
<th>Teacher</th>
<th>/CV/</th>
<th>/V/</th>
</tr>
</thead>
<tbody>
<tr>
<td>[CV]</td>
<td>100%</td>
<td>[V]</td>
</tr>
<tr>
<td>[V]</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>[CV]</td>
<td>50%</td>
<td>[V]</td>
</tr>
<tr>
<td>[V]</td>
<td>50%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Learner</th>
<th>/CV/</th>
<th>/V/</th>
</tr>
</thead>
<tbody>
<tr>
<td>[CV]</td>
<td>100%</td>
<td>[V]</td>
</tr>
<tr>
<td>[V]</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>[CV]</td>
<td>75%</td>
<td>[V]</td>
</tr>
<tr>
<td>[V]</td>
<td>25%</td>
<td></td>
</tr>
</tbody>
</table>
Simulation Methodology: Within-Generation

I use the truncated perceptron algorithm Magri (2015); Rosenblatt (1958); Boersma & Pater (2016).

- An input is randomly selected.
- The teacher and learner select outputs for that input based on their current grammar.
- If they differ, the learner updates their grammar to make the teacher’s form more likely in the future.

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Learner</th>
</tr>
</thead>
<tbody>
<tr>
<td>/CV/</td>
<td>/CV/</td>
</tr>
<tr>
<td>[CV] 100%</td>
<td>[CV] 100%</td>
</tr>
<tr>
<td>[V] 0%</td>
<td>[V] 0%</td>
</tr>
<tr>
<td>/V/</td>
<td>/V/</td>
</tr>
<tr>
<td>[CV] 50%</td>
<td>[CV] 75%</td>
</tr>
<tr>
<td>[V] 50%</td>
<td>[V] 25%</td>
</tr>
<tr>
<td>[V] x [CV]</td>
<td>[V] x [CV]</td>
</tr>
</tbody>
</table>
Simulation Methodology: Within-Generation

I use the truncated perceptron algorithm Magri (2015); Rosenblatt (1958); Boersma & Pater (2016).

- An input is randomly selected.
- The teacher and learner select outputs for that input based on their current grammar.
- If they differ, the learner updates their grammar to make the teacher’s form more likely in the future.

<table>
<thead>
<tr>
<th></th>
<th>Teacher</th>
<th>Learner</th>
</tr>
</thead>
<tbody>
<tr>
<td>/CV/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[CV]</td>
<td>100%</td>
<td>[CV] 100%</td>
</tr>
<tr>
<td>[V]</td>
<td>0%</td>
<td>[V] 0%</td>
</tr>
<tr>
<td>/V/</td>
<td>[CV] 50%</td>
<td>[CV] 70%</td>
</tr>
<tr>
<td>[V]</td>
<td>50%</td>
<td>[V] 30%</td>
</tr>
</tbody>
</table>
Two Phases of Error Driven Learning

There are two major phases of error-driven learning of stochastic grammars.

- **Learning Phase**: Most updates move the learner towards the target grammar and away from the starting grammar.
- **Oscillation Phase**: Updates cause the learner to oscillate around the target pattern.
Oscillatory Phase

When the teacher’s grammar is variable, errors continue even when the learner has the same grammar.

<table>
<thead>
<tr>
<th>Input</th>
<th>Teacher</th>
<th>Learner</th>
<th>Error?</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>/V/</td>
<td>[CV]</td>
<td>[V]</td>
<td>YES</td>
<td>p([CV])↑ 25%</td>
</tr>
<tr>
<td></td>
<td>[V]</td>
<td>[CV]</td>
<td>YES</td>
<td>p([V])↑ 25%</td>
</tr>
<tr>
<td></td>
<td>[V]</td>
<td>[V]</td>
<td>no</td>
<td></td>
</tr>
</tbody>
</table>

Errors occur 50% of the time, but they are balanced in both directions, so the average across many runs will remain at the target pattern.
Simulation Methodology: Across-Generation

Generational Learning Model⁴

- Simulated learners using MaxEnt grammars
- Learners are initialized with Markedness constraints high, faith low⁵
- Train a learning agent off of some limited number of forms⁶ from a teacher.
- Then train a new learner on that agent’s final grammar.
- Patterns that remain stable across generations are likely to be better attested.

---

⁴ Following Staubs (2014); Hughto (2018)
⁵ Gnanadesikan (2004); Tesar & Smolensky (2000); Jesney & Tessier (2011)
⁶ Kirby & Huford (2002)
NORMAL VARIATION SIMULATIONS

Variable Onset Epenthesis

/pa/ → [pa]  100%  /u/ → [u]  50%  /ʔu/ → [ʔu]  50%
Normal variation simulations clearly show the oscillation phase, but the average run converges towards the target grammar.
Normal variation is learned here in around 2100 iterations.

![Graph showing the learning of normal variation over iterations. The graph includes a line indicating the average rates of change. The x-axis represents iterations ranging from 2,000 to 5,000, and the y-axis represents probability ranging from 0.2 to 0.8. There are two lines: one for CV→CV rate and another for V→V rate.](image-url)
There is variation across runs in terms of generational change. Typological implications are respected in all runs throughout.

**Average Rates**
- CV→CV rate
- V→V rate
Generational Change of Normal Variation

40% remained within a .25 probability window, but 12 runs lost variation: six categorically epenthesize onsets, and six never epenthesize onsets.\textsuperscript{7}

\textsuperscript{7}bias for categorical patterns replicating Hughto (2018)
Generational Change of Normal Variation

40% remained within a .25 probability window, but 12 runs lost variation: six categorically epenthesize onsets, and six never epenthesize onsets.  

---

8 bias for categorical patterns replicating Hughto (2018)
40% remained within a .25 probability window, but 12 runs lost variation: **six categorically epenthesize onsets**, and six never epenthesize onsets. \(^9\)

---

\(^9\) bias for categorical patterns replicating Hughto (2018)
Generational Change of Normal Variation

40% remained within a .25 probability window, but 12 runs lost variation: six categorically epenthesized onsets, and six never epenthesized onsets.  

---

bias for categorical patterns replicating Hughto (2018)
Weights for Normal Variation

First generation weighting dynamics are consistent, Onset and Dep meet.

Average Rates
- NoCoda
- Onset
- Max
- Dep

Graph showing weight changes over iterations with lines for different categories.
And then they oscillate around each other.
Weights at later generations

At the seventeenth (last) generation, there are 3 types of weighting dynamics observed.

![Graph showing weight variation over iterations]

**Average Rates**
- **NoCoda**
- **Onset**
- **Max**
- **Dep**
At the seventeenth (last) generation, there are 3 types of weighting dynamics observed. **Variation**

**Average Rates**
- **NoCoda**
- **Onset**
- **Max**
- **Dep**

![Graph showing weight dynamics over iterations](image)
At the seventeenth (last) generation, there are 3 types of weighting dynamics observed. **Variation, Categorical Faithfulness**
Weights at later generations

At the seventeenth (last) generation, there are 3 types of weighting dynamics observed. Variation, Categorical Faithfulness, and **Categorical Epenthesis**.
HARMONICALLY-BOUNDED VARIATION SIMULATIONS

Variable Onset Deletion

\[
/\text{pa}/ \rightarrow [\text{pa}]\ 50\% \quad [\text{a}]\ 50\% \\
/\text{u}/ \rightarrow [\text{u}]\ 100\%
\]
Learning of Harmonically-Bounded Variation

Given enough data, learners can learn to delete onsets.

Average Rates
- CV→CV rate
- V→V Rate
Learning of Harmonically-Bounded Variation

Given enough data, learners can learn to delete onsets. Converges at around 2700 iterations.
Learning of Harmonically-Bounded Variation

But notably, it doesn’t converge quite to the target pattern (gray dashed line).

![Graph showing probability distribution over iterations]

**Average Rates**
- Blue line: CV → CV rate
- Red line: V → V Rate
Generational Change of Harmonically-Bounded Variation

Harmonically-Bounded Variation is far less stable—lost in all but one run.

![Graph showing generational change of Harmonically-Bounded Variation](image-url)
Generational Change of Harmonically-Bound Variation

Harmonically-Bounded Variation is far less stable—lost in all but one run.

![Graph showing probability over generations with average rates]

- **Average Rates**
  - CV → CV rate (blue line)
  - V → V rate (red line)
Weights for Harmonically-Bounded Variation

First generation weighting dynamics are consistent, constraints need to go to zero so harmonic bounded candidates get weight.
In later generations, it takes increasingly long for *Onset* to reach zero.
In later generations, it takes increasingly long for Onset to reach zero.
Weights for Harmonically-Bounded Variation

Once a learner doesn’t reach near zero for Onset, they quickly stop lowering it far below Dep
Weights for Harmonically-Bounded Variation

Once a learner doesn’t reach near zero for $\text{Onset}$, they quickly stop lowering it far below $\text{Dep}$.
### BOTH TYPES OF VARIATION SIMULATIONS

**Variable Onset and Coda Deletion**

<table>
<thead>
<tr>
<th>/pa/</th>
<th>[pa] 50%</th>
<th>/u/</th>
<th>[u] 100%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[a] 50%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>/a/</th>
<th>[a] 100%</th>
<th>/uk/</th>
<th>[uk] 50%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[u] 50%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Learning of Harmonically-Bounded Variation

What if we look at both types of variation in one grammar?

![Graph showing probability over iterations for Onset and Coda Retention Rates.](image)

**Average Rates**
- Onset Retention Rate
- Coda Retention Rate
What if we look at both types of variation in one grammar?
Converges at around 5000 iterations.
Generational Change of Combined Variation

Harmonically-Bounded Variation is far less stable—all runs lose harmonically bounded variation.

![Graph showing Generational Change of Combined Variation](image)

**Average Rates**
- Onset Retention Rate
- Coda Retention Rate
Generational Change of Combined Variation

Harmonically-Bounded Variation is far less stable—all runs lose harmonically bounded variation but 30% maintain normal variation.
Harmonically-Bounded Variation is far less stable—all runs lose harmonically bounded variation but 30% maintain normal variation.
Summary

- Harmonically Bounded Variation is harder to learn than Normal Variation.
- If Harmonically Bounded Variation and Normal Variation are in a pattern, loss of Normal Variation implies loss of Harmonically Bounded Variation.

![Categorical Pattern Diagram]

- Normal Variation
- Harmonically-Bounded Variation
- Harmonically-Bounded Variation + Normal Variation

Numbers represent transition probabilities:
- 0.95 from Harmonically-Bounded Variation to Normal Variation
- 0.70 from Normal Variation to Harmonically-Bounded Variation
- 0.60 between Normal Variation and Categorical Pattern
- 0.30 from Categorical Pattern to Normal Variation + Normal + HB Variation
Why is giving probability to harmonically bounded candidates hard?

- Different types of weighting condition needed to give harmonically bounded candidate probability.

  **Normal Variation**  
  \[ \text{Dep} \sim \text{Onset} \]  
  2100 Iterations

  **HB Variation**  
  \[ \text{Max} \sim \text{Onset} \sim 0 \]  
  2700 Iterations

- Oscillation phase works different for harmonically-bounded variation.
  - In a normal variation pattern, the learner is equally likely to oscillate towards either candidate.
  - In a harmonically-bounded variation pattern, the truncated aspect of the algorithm bounds how much probability the learner can give the harmonically bounded candidate.
Generational Differences

Oscillating constraint weights

Oscillating probabilities

- **ONSET**
- **DEP**

- **V→V rate**
- **Target probability**
Generational Differences

Oscillating constraint weights

Oscillating probabilities

- Onset
- Max

CV→CV rate
Target probability
Harmonically Bounded Candidates

With learning, MaxEnt rarely gives harmonically bounded candidates much probability, but it always will give them SOME probability. Giving some probability to harmonically bounded candidates may not be the worst thing in the world.

- A harmonically bounded candidate never receives the most probability of the candidates.
- Harmonically bounded candidates can be observed in speech errors\(^\text{11}\)
- Gradient well-formedness of harmonically bounded candidates can be greater than non-bounded candidates \(^\text{12}\)

---

\(^{11}\) Goldrick & Daland (2009)

\(^{12}\) Hayes & Wilson (2008); Hayes & Moore-Cantwell (2011); Hayes (2017); Hayes & Schuh (to appear)
Takeaway

- **Is grammatical overgeneration a problem?**
  - Not necessarily, if the unattested languages can be ruled out independently by learning (or other factors)

- **Does grammatical structure still matter?**
  - Yes! Properties of the grammar (like harmonic bounding) still have some effect.
Takeaway

- Is grammatical overgeneration a problem?
  - Not necessarily, if the unattested languages can be ruled out independently by learning (or other factors)

- Does grammatical structure still matter?
  - Yes! Properties of the grammar (like harmonic bounding) still have some effect.
Takeaway

- Is grammatical overgeneration a problem?
  - Not necessarily, if the unattested languages can be ruled out independently by learning (or other factors)

- Does grammatical structure still matter?
  - Yes! Properties of the grammar (like harmonic bounding) still have some effect.
Takeaway

- Is grammatical overgeneration a problem?
  - Not necessarily, if the unattested languages can be ruled out independently by learning (or other factors)

- Does grammatical structure still matter?
  - Yes! Properties of the grammar (like harmonic bounding) still have some effect.


Works Cited II


Works Cited III


Questions

I focused here on simply harmonically bounded forms. Collectively bounded forms may act different.

<table>
<thead>
<tr>
<th>/bababa/</th>
<th>B</th>
<th>FAITH</th>
</tr>
</thead>
<tbody>
<tr>
<td>weights</td>
<td>w = 5</td>
<td>w = 5</td>
</tr>
<tr>
<td>a. bababa</td>
<td>-3</td>
<td>-15</td>
</tr>
<tr>
<td>b. bapaba</td>
<td>-2</td>
<td>-15</td>
</tr>
<tr>
<td>c. papapa</td>
<td>-3</td>
<td>-15</td>
</tr>
</tbody>
</table>

- Noisy HG performs differently than MaxEnt here (Hayes, 2017).
  - The version discussed in most of this paper gives no probability to the bounded candidate.
  - Other versions can create a u-shaped distribution across these forms.
- Can these types of patterns cause subversion of t-orders?
- Can the distribution of probability across collectively bounded forms (local optionality) differentiate between theories? Maybe (Hayes, 2017)
I focused here on simply harmonically bounded forms. Collectively bounded forms may act different.

<table>
<thead>
<tr>
<th>/bababa/</th>
<th>*B</th>
<th>FAITH</th>
<th>HARM</th>
<th>PROB</th>
</tr>
</thead>
<tbody>
<tr>
<td>weights</td>
<td>$w = 5.1$</td>
<td>$w = 5$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. bababa</td>
<td>-3</td>
<td></td>
<td>-15.3</td>
<td>.33</td>
</tr>
<tr>
<td>b. bapaba</td>
<td>-2</td>
<td>-1</td>
<td>-15.2</td>
<td>.33</td>
</tr>
<tr>
<td>c. papapa</td>
<td>-3</td>
<td>-15</td>
<td></td>
<td>.33</td>
</tr>
</tbody>
</table>

- Noisy HG performs differently than MaxEnt here (Hayes, 2017).
  - The version discussed in most of this paper gives no probability to the bounded candidate.
  - Other versions can create a u-shaped distribution across these forms.

Can these types of patterns cause subversion of t-orders?
Can the distribution of probability across collectively bounded forms (local optionality) differentiate between theories?
Maybe (Hayes, 2017)
Collectively Bounded Forms

Noisy HG

MaxEnt (*b=Faith)
Collectively Bounded Forms

MaxEnt (*b=Faith+.5)

MaxEnt (*b=Faith+1.5)
Do learners see harmonically bounded forms?

In these simulations, learners had full access to the teacher’s underlying form.

- This is unlikely from a learning standpoint.
- If a teacher produces /CVC/-[VC], a learner only hears [VC].
- By Lexicon Optimization, the learner will usually choose /VC/ as the underlying representation, rather than the harmonically bounded /CVC/.
- Harmonically bounded mappings are all either unfaithful, or involve hidden structure.
- Thus, learners would perceive even fewer harmonically bounded mappings than in these simulations.