One-shot vs. competitions phonotactics in modeling constraint cumulativity

Marisabel (Isa) Cabrera
UCLA

AMP 2024
1-3 November
Rutgers University

Introduction

- Model variable phonological patterns with log-linear (MaxEnt) models: output of grammar is a probability distribution over candidates (Goldwater & Johnson 2003; Hayes & Wilson 2008).

- Both phonotactics and alternations show variable patterns.

English phonotactics
hæ**mp**ɹ̩
ɛ**nt**ɹ̩
ɪ**ŋg**lɪʃ
…
ɪ**np**ʌt
kæ**md**ən

Tagalog alternations (Zuraw 2010)	
/ma**ŋ-b**igáj/	[ma-**m**igáj]
/ma**ŋ-s**úlat/	[mà-**n**ulát]
/ma-pa**ŋ-k**amkám/	[ma-pa-**ŋ**amkám]
…	…
/pa**ŋ-p**oʔók/	[pa**m-p**oʔók]
/pa**ŋ-s**úlat/	[pa**n-s**úlat]

Introduction

- In phonotactics, model assigns a ***single*** probability distribution over a (big) list of forms.

candidates	Agree[place] *w*	H	e^H	probability
hæ**mp**ɹ	0	0	1	1 / **Z**
ɪŋglɪʃ	0	0	1	1 / **Z**
…	…	…	…	…
ɪnpʌt	1	-w	e^{-w}	e^{-w} / **Z**
kæ**md**ən	1	-w	e^{-w}	e^{-w} / **Z**

$$Z = \sum_{c} e^{H(c)}$$

$$\sum_{c} P(c) = 1$$

Hayes & Wilson 2008

Introduction

- In phonotactics, model assigns a ***single*** probability distribution over a (big) list of forms.

"one-shot" phonotactics

all candidates in a single competition

Hayes & Wilson 2008

Introduction

- We can also model phonotactics as a ***binary choice*** between a structural candidate (observed form) and null candidate ⊙.

inputs	**cands.**	**freq.**	**Agree[pl]** w_1	**MParse** w_2	**probability**	
hæ**mp**ɹ	hæ**mp**ɹ	1	0		$1 / Z_1$	Z_1
	⊙	0		1	e^{-w2} / Z_1	
ɪ**ŋg**lɪʃ	ɪ**ŋg**lɪʃ	1	0		$1 / Z_2$	Z_2
	⊙	0		1	e^{-w2} / Z_2	
kæ**md**ən	kæ**md**ən	0	1		e^{-w1} / Z_3	Z_3
	⊙	1		1	e^{-w2} / Z_3	

McCarthy & Wolf 2005; Kawahara 2021; Hayes 2022; Breiss & Albright 2022

Introduction

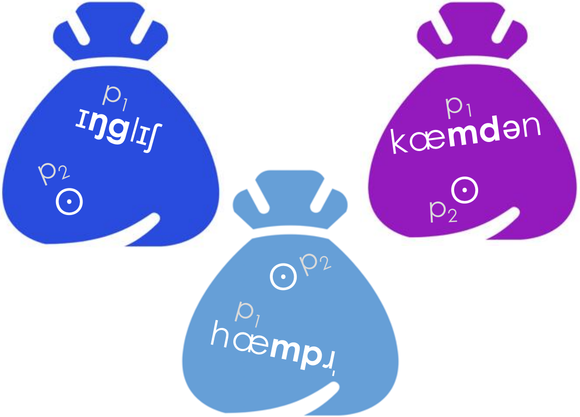
- We can also model phonotactics as a ***binary choice*** between a structural candidate (observed form) and null candidate ⊙.

inputs	cands.	freq.	Agree[pl] w_1	MParse w_2	probability	
hæmpɹ	hæmpɹ	1	0		$1 / Z_1$	$\sum_c = 1$
	⊙	0		1	e^{-w2} / Z_1	
ɪŋglɪʃ	ɪŋglɪʃ	1	0		$1 / Z_2$	$\sum_c = 1$
	⊙	0		1	e^{-w2} / Z_2	
kæmdən	kæmdən	0	1		e^{-w1} / Z_3	$\sum_c = 1$
	⊙	1		1	e^{-w2} / Z_3	

McCarthy & Wolf 2005; Kawahara 2021; Hayes 2022; Breiss & Albright 2022

Introduction

- We can also model phonotactics as a ***binary choice*** between a structural candidate (observed form) and null candidate ⊙.



"multiple competitions" phonotactics

structural candidates are in separate competitions

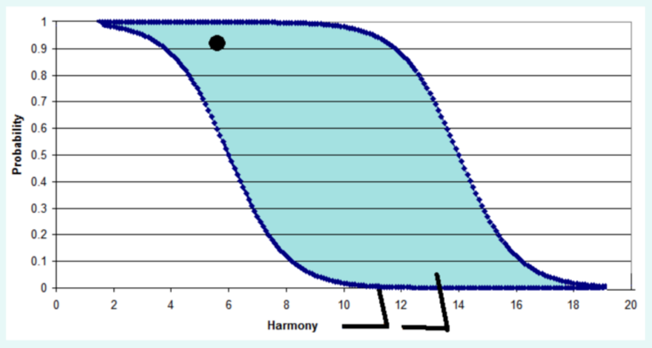
Introduction

- Multiple competitions models bring phonotactics closer to ***alternations*** - assign multiple probability distributions.

/pa**ŋ**-**s**úlat/ /pa**ŋ**-**p**oʔók/ /ma**ŋ**-**b**igáj/

Introduction

- Binary competition is required to derive wug-shaped curves – MaxEnt's "quantitative signature" (Kawahara 2021, Hayes 2022).
  - Frequency pattern widely found in quantitative studies of variable patterns.



Introduction

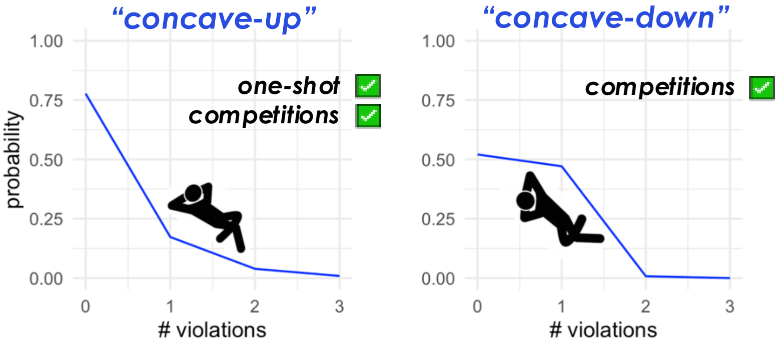
One-shot vs. multiple competitions phonotactics raises some important questions…

- Can we model ***frequencies*** with the multiple competitions model?
- How would phonotactic learning proceed when "say nothing" ⊙ is ***unobservable***?
- How do the models differ in predictions?
- Are their predictions empirically attested?

TODAY: models make different empirical predictions regarding ***cumulative phonotactic effects***…

Introduction

- ***One-shot*** models: additional violations take a ***decreasing hit*** on probability relative to previous violations.
- ***Multiple competitions*** models: additional violations may take a ***greater hit*** *(under certain weighing conditions)*.



Roadmap

1. Background on cumulativity in phonology
2. Formal properties of one-shot vs. multiple competitions phonotactics
3. Learning concave-up and concave-down patterns
   - sanity check
   - can the models predict concave-down patterns in absence of such pattern in training?
4. Discussion

Roadmap

1. **Background on cumulativity in phonology**
2. Formal properties of one-shot vs. multiple competitions phonotactics
3. Learning concave-up and concave-down patterns
   - sanity check
   - can the models predict concave-down patterns in absence of such pattern in training?
4. Discussion

Cumulativity in phonology

- When 2+ constraint violations together have an additive effect on the phonology of a language.
- A form with *n+1* violations is somehow worse than a form with *n* violations (when n >= 1).
- Additive "worsening" effect evidenced in:

1. lexical frequencies (Albright 2008; Shih 2017; Yang et al. 2018)
2. acceptability judgments (Pizzo 2015; Breiss 2020; Breiss & Albright 2022)
3. repairs (alternations) (Farris-Trimble 2008; Green & Davis 2014; Shih 2017; Smith & Pater 2020; Kim 2022)

Cumulativity in phonology

1. lexical frequencies: Albright (2008)

- Lakhota fricatives, ejectives, aspirates, and consonant clusters are marked structures - they're quite uncommon.
- But, words with two of these are ***way more*** uncommon…
  - Bisyllabic words: 32% have fricative as C_1 and 18% have fricative as C_2.
  - Only **1%** have two fricatives.
  - But we expect **6%** from joint probability (0.32 x 0.18)

Cumulativity in phonology

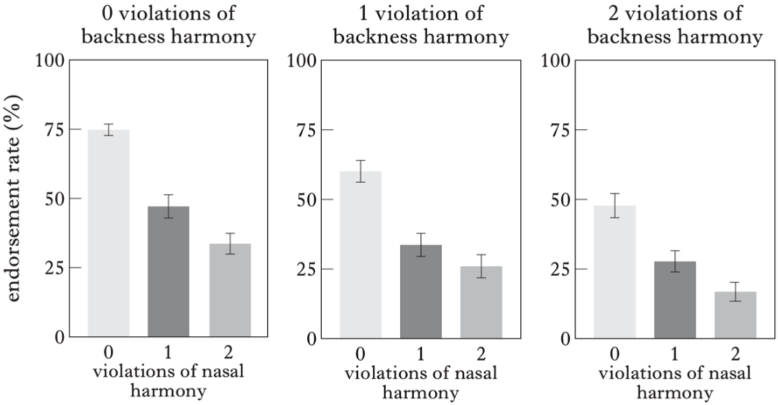
2 acceptability judgments: Breiss (2020)

- Familiarized participants with exceptionless backness and nasal harmony (*potu*, *nime*)
- Asked to rate zero-, singly-, and doubly-violating words.
- ***Result:*** speakers assume cumulativity even when there's no evidence for it in the input.

Cumulativity in phonology

2 acceptability judgments: Breiss (2020)

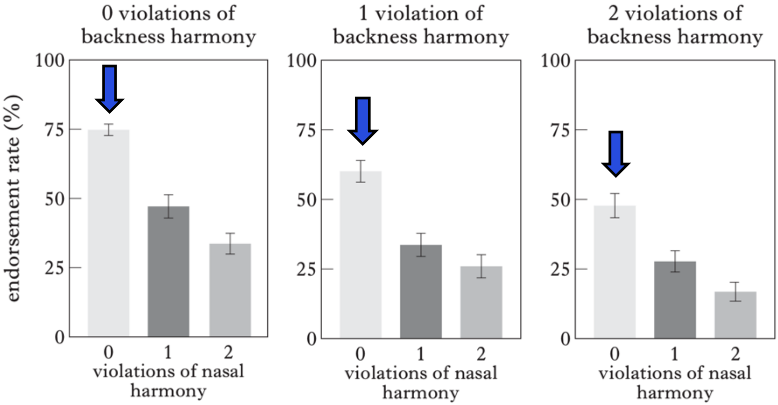
- Results for binary decision tasks.



Cumulativity in phonology

2 acceptability judgments: Breiss (2020)

- Results for binary decision tasks.



Cumulativity in phonology

- Cumulativity not predicted by all theories of phonology.
  - Strict-ranking OT

	Constraint A	Constraint B	Constraint C
Candidate A	*!		
☞ **Candidate B**		*	*

  - Harmonic Grammar (Legendre et al. 1990)

	Constraint A *w = 3*	Constraint B *w = 2*	Constraint C *w = 2*	*H*
☞ **Candidate A**	*			**3**
Candidate B		*	*	4

The one-shot model

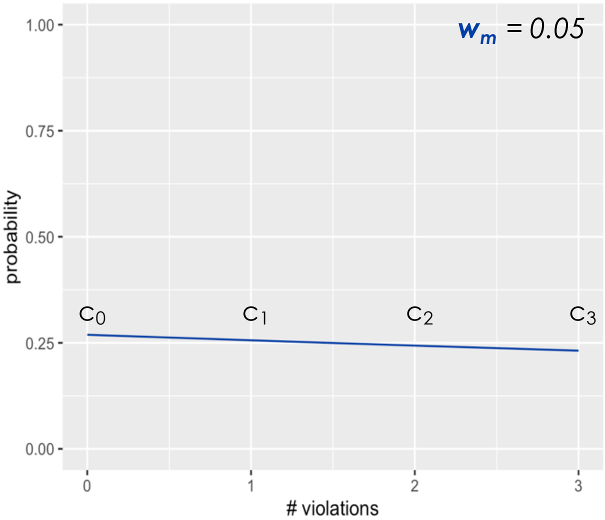
- Assigns a ***single*** probability distribution over all inputs.
- ***Counting*** cumulativity: multiple violations of the same constraint (vs. *ganging* cumulativity)
- ***How do subsequent violations affect predicted probability?***

inputs	mark w_m	H	probability
c_0	0	0	1 / **Z**
c_1	1	-w	e^{-w} / **Z**
c_2	2	-2w	e^{-2w} / **Z**
c_3	3	-3w	e^{-3w} / **Z**

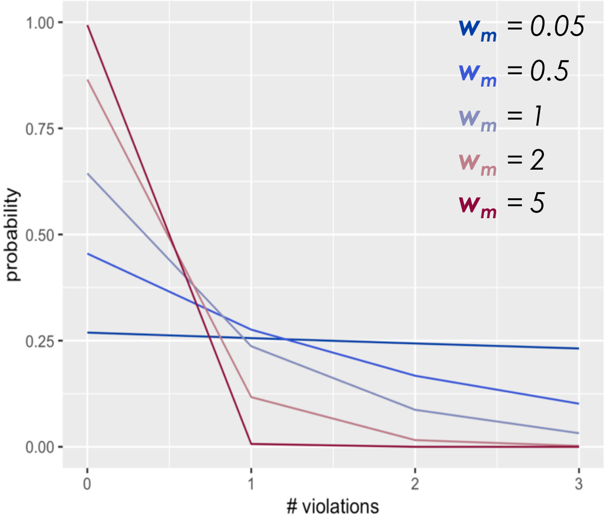
$$\frac{P(c1)}{P(c2)} = \frac{e^{-w}}{e^{-2w}} = \frac{\mathbf{1}}{\boldsymbol{e^{-w}}}$$

Each additional violation decreases probability by $\mathbf{e^{w}}$.

The one-shot model



The one-shot model



- All curves are ***concave-up***
- Later violations cause ***smaller*** dips in probability.
- Increasingly concave- up as w_m increases.

The competitions model

- Assigns ***multiple*** probability distributions, one for each input.

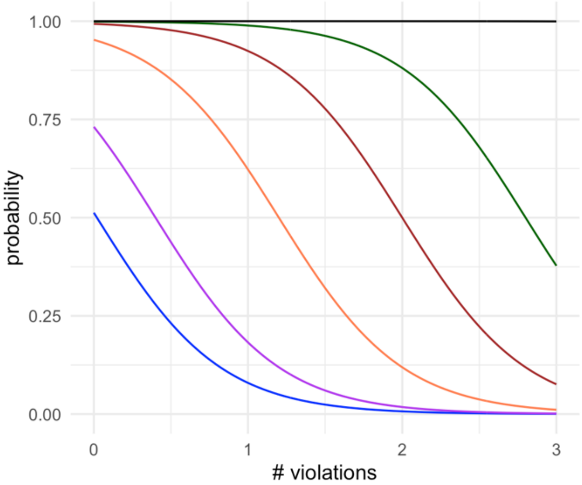
inputs	cands	mark w_m	MParse w_{mp}	prob
C_0	C_0	0		?
	⊙		1	
C_1	C_1	1		?
	⊙		1	
C_2	C_2	2		?
	⊙		1	
C_3	C_3	3		?
	⊙		1	

Asymmetric trade-off (Pater 2009)

3 important properties…

The competitions model

1. Type of curve ("concavity") is a function of weight of ***MParse***.



w_m *constant at* ***2.5***

w_{mp}

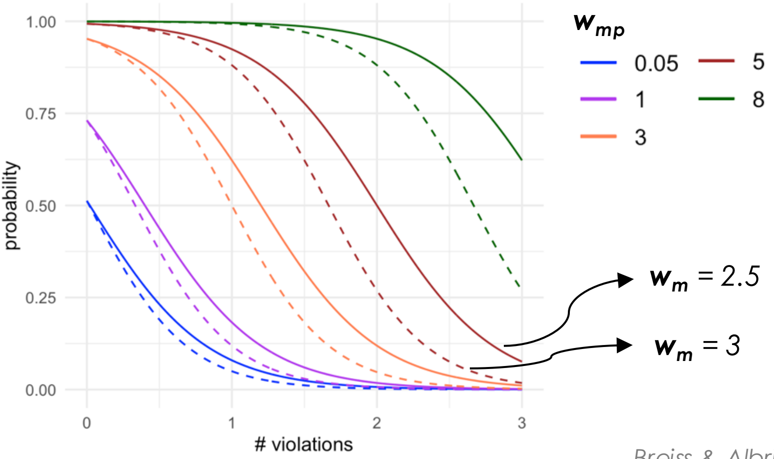
- 0.05
- 1
- 3
- 5
- 8
- 15

Curves are concave-up and concave-down.

Shifted sigmoids. (Kawahara 2021; Hayes 2022)

The competitions model

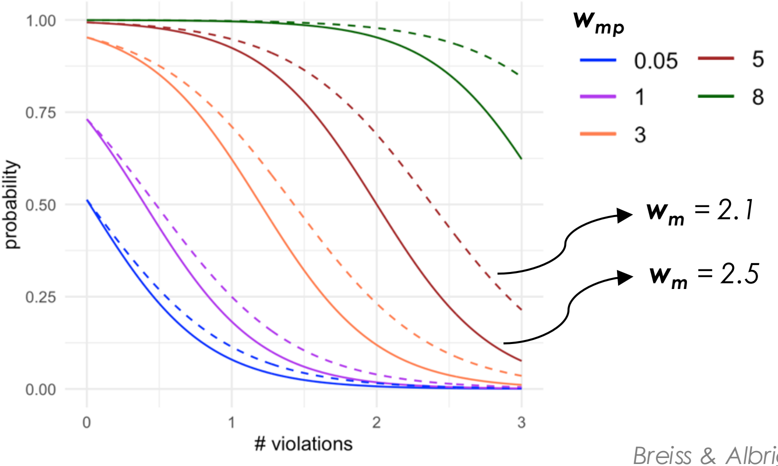
2 Steepness of curve (strength of concavity) is a function of weight of ***markedness***.



Breiss & Albright 2022

The competitions model

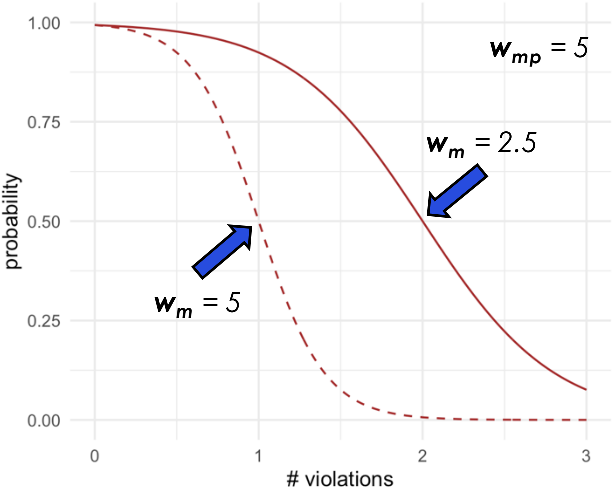
2. Steepness of curve (strength of concavity) is a function of weight of ***markedness***.



Breiss & Albright 2022

The competitions model

3 Installs a "threshold of markedness" (inflection point)



Quickly prefer structural candidate above threshold, and ⊙ below threshold.

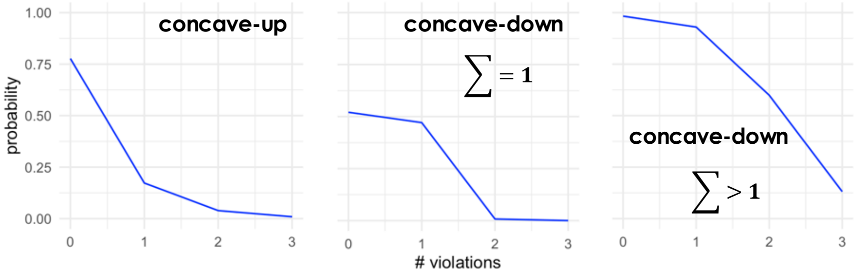
How quick: weight of markedness (property 2)

Breiss & Albright 2022

Learning

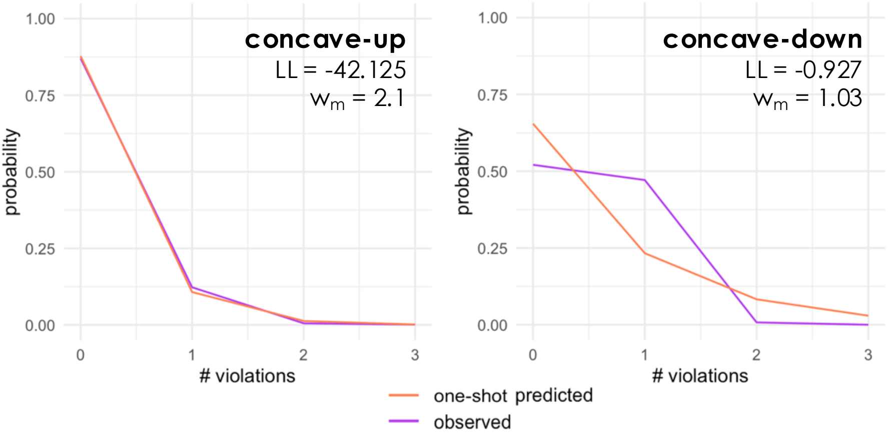
- Simulated different kinds of concave-up and concave-down curves to test model learning.

- I'll focus on these:



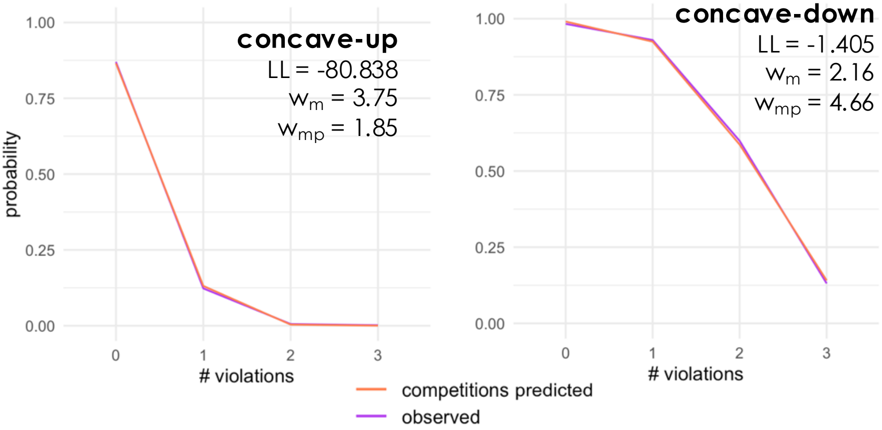
Learning

- As expected, ***one-shot*** only fits ***concave-up*** curves.



Learning

- Multiple competitions can fit ***both***.



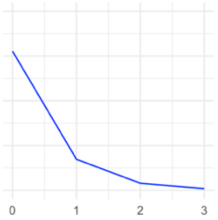
Summary

- "One-shot" vs. "multiple competitions" MaxEnt models differ in the kinds of cumulative phonotactic effects they predict.
    - **One-shot**: later violations take a ***decreasing*** hit on probability.
    - **Multiple competitions**: later violations may take a ***greater*** hit on probability than earlier violations (under some weighing conditions).
- Competitions model only learns concave-down patterns when explicitly trained on them.
- *Are concave-down patterns empirically attested?*
- *How is the weight of MParse learned when the null "say nothing" candidate is unobservable?*

Learning under competitions model

- ***Concave-up*** learning set-up.

inputs	cands	obs. freq.	mark w_m	MParse w_{mp}	pred. prob
C_0	C_0	0.87	0		?
	⊙	**0.13**		1	
C_1	C_1	0.123	1		?
	⊙	**0.877**		1	
C_2	C_2	0.0055	2		?
	⊙	**0.9945**		1	
C_3	C_3	0.0015	3		?
	⊙	**0.9985**		1	



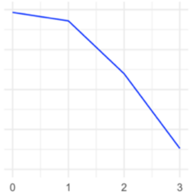
Assumed structural candidates are in the "same distribution"

Assumed ***1-p*** frequency for ⊙

Learning under competitions model

- Not possible with ***concave-down*** patterns that competitions models predict.

inputs	cands	obs. freq.	mark w_m	MParse w_{mp}	pred. prob
c_0	c_0	0.983	0		**?**
	⊙	**0.017**		1	
c_1	c_1	0.93	1		**?**
	⊙	**0.07**		1	
c_2	c_2	0.599	2		**?**
	⊙	**0.401**		1	
c_3	c_3	0.131	3		**?**
	⊙	**0.869**		1	



Can't assume structural candidates are in same distribution.

Alternative 1

- ***Proposal:*** learning with multiple competitions and unrestricted GEN.

inputs	cands	obs. freq.	mark w_m	MParse w_{mp}	…
black	black	1	…		…
	⊙	0		1	
blick	blick	0	…		…
	⊙	1		1	
bnick	bnick	0	…		…
	⊙	1		1	

- ***Testable prediction:*** concave-down patterns are learnable.

Alternative 2

- ***Proposal:*** one-shot and competitions as models for different tasks.

one-shot	→	***competitions***
phonotactic learning (Hayes & Wilson 2008)		model acceptability judgments

Alternative 2

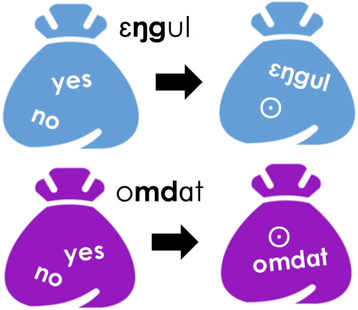
- ***Proposal:*** one-shot and competitions as models for different tasks.

one-shot

phonotactic learning
(Hayes & Wilson 2008)

competitions

model acceptability judgments



Alternative 2

- ***Proposal:*** one-shot and competitions as models for different tasks.

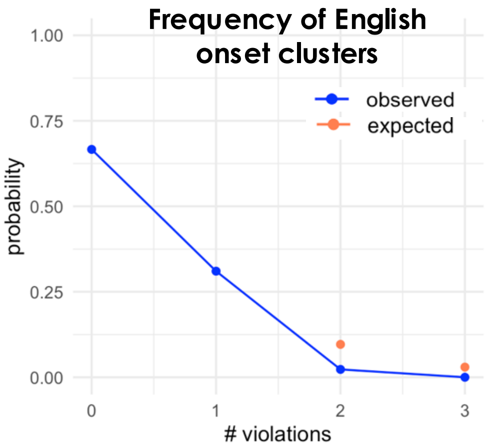
one-shot	→	***competitions***
phonotactic learning (Hayes & Wilson 2008)		model acceptability judgments

- Learning and judgments are different tasks and grammar structure can reflect those differences.

Concavity vs. linearity

- Literature often investigates the ***linearity*** of cumulativity.
- Linearity: **observed** vs. **expected** (expected = joint prob. of candidates with single violations)
- Lakhota fricatives are ***superlinear*** (Albright 2008)
  - Expected prob. of doubly-violating: 32% x 18% = **6%**
  - Observed prob. = **1%**
- ***Concavity and linearity are different…***

Concavity vs. linearity



- English onset clusters are ***concave-up*** but ***superlinear***.
- Superlinear:

  observed < expected

- Concavity aligns with one-shot vs. competitions differences.
- Breiss & Albright (2022) use the competitions model to predict ***superlinearity***.

Closing

- Different candidate competitions structures lead to different and testable empirical predictions.
  - ***One-shot*** models only predict ***concave-up*** patterns.
  - ***Competitions*** can predict ***concave-down*** patterns.

- Consequences for modeling of phonotactics vs. alternations: alternations also involves multiple competitions.
- Extensions to ganging cumulativity: violations of different constraints.
- Extensions to Stochastic OT and Noisy HG (Boersma & Hayes 2001; Boersma & Pater 2016).
- Cumulativity tells us a lot about how grammars should be structured, probabilistic or not.

thank you!

Huge thanks to Claire Moore-Cantwell, Tim Hunter, Bruce Hayes, Canaan Breiss, Megha Sundara, and members of the UCLA Phonology Seminar and UCLA Comp/PsychoLing Seminar.

References

Albright, A. 2008. Cumulative violations and complexity thresholds. Ms., MIT.

Boersma, P. & B. Hayes. 2001. Empirical tests of the gradual learning algorithm. *LI 32(1)*.

Boersma, P. & J. Pater. 2016. Convergence properties of a gradual learning algorithm for Harmonic Grammar.

Breiss, C. 2020. Constraint cumulativity in phonotactics: evidence from artificial grammar learning studies. *Phonology 37*.

Breiss, C & A. Albright. 2022. Cumulative markedness effects and (non-)linearity in phonotactics. *Glossa 7*(1).

Farris-Trimble, A.W. 2008. Cumulative faithfulness effects. Ph.D. Dissertation, University of Indiana.

Goldwater, S. & M. Johnson. 2003. Learning OT constraint rankings using maximum entropy models. *Proceedings of the Workshop on Variation within Optimality Theory.*

Hayes, B. 2022. Deriving the wug-shaped curve: A criterion for assessing formal theories of linguistic variation. *Annual Rev. of Linguistics 8* (1).

Hayes, B. & C. Wilson. 2008. A Maximum Entropy model for phonotactics and phonotactic learning. *LI 39* (3).

Legendre, G., et al. 1990. Harmonic Grammar: a formal multi-level connectionist theory of linguistic well-formedness.

Pater, J. 2009. Weighted constraints in generative linguistics. *Cognitive Science* 33.

Pizzo, J. 2015. Investigating properties of phonotactic knowledge through web-based experimentation. Ph.D. Dissertation, University of Massachusetts Amherst.

Smith, B. & J. Pater. 2020. French schwa and gradient cumulativity. *Glossa 5*(1).

Wolf, M. & J.H. McCarthy. 2008. Less than zero: correspondence and the null output.