One-shot vs. competitions phonotactics in modeling constraint cumulativity

Marisabel Cabrera
University of California, Los Angeles

1 Introduction

Speakers of any language have systematic intuitions about which sounds or sound sequences are probable and improbable in their language. However, only relatively recently have we started to understand the effects of co-occurring unlikely structures in the same word, referred to as *cumulativity* or *additivity* in phonological literature. Specifically, how does the introduction of a second or third unlikely structure in a word affect its frequency in the language's lexicon, and also speakers' intuitions about such word? Do the widely assumed models of phonological grammar capture these patterns and intuitions? In this paper, I compare the predicted hypothesis spaces of two probabilistic model structures that differ in the structure of the candidate competition. This work finds that these two models predict widely different cumulative patterns, specifically differing in their predictions about the extent to which probability falls as constraint violations accumulate.

Typically, in the modeling of variable phonotactic patterns, probabilistic phonological models such as Maximum Entropy grammars (Goldwater and Johnson, 2003; Hayes and Wilson, 2008) assign a single probability distribution over a set of forms. For example, in modeling the distribution of English words with zero, one, and two [θ]'s, an uncommon sound in the language, all words with varying quantities of [θ] are in a single competition. All candidates in the grammar are normalized by the constant Z, the sum of all exponentiated harmonies of all candidates, hence all candidates form part of the same (and only) probability distribution. I'll call these kinds of models "one-shot" models of phonotactics in this work.

(1)

	*θ w_m	Harmony	e^{Harmony}	predicted probability
hamper	0	0	1	1 / **Z**
english	0	0	1	1 / **Z**
...				
theta	1	$-w_m$	e^{-w_m}	e^{-w_m} / **Z**
thought	1	$-w_m$	e^{-w_m}	e^{-w_m} / **Z**
...				
ba**th**oli**th**	2	$-2w_m$	e^{-2w_m}	e^{-2w_m} / **Z**
or**th**opa**th**y	2	$-2w_m$	e^{-2w_m}	e^{-2w_m} / **Z**

An alternative grammar structure used in the modeling of phonotactics involves setting each structural candidate in competition against a null candidate $\odot$ (Breiss and Albright, 2022). The null candidate represents a null output in the grammar, hence violating MPARSE, a constraint requiring valid outputs, only once (McCarthy and Wolf, 2005). Now, structural candidates are in separate competitions as their exponentiated

* Huge thank you to the audiences of AMP 2024 and the Phonology Seminar at UCLA. I am especially grateful for Claire Moore-Cantwell, Tim Hunter, Canaan Breiss, Bruce Hayes, Kie Zuraw, and Megha Sundara for their guidance.

Proceedings of *AMP 2023/2024*

harmonies are normalized by different constants. I'll call these models "multiple competitions" models in this work, since the MaxEnt model now assigns multiple probability distributions across structural candidates.

(2)

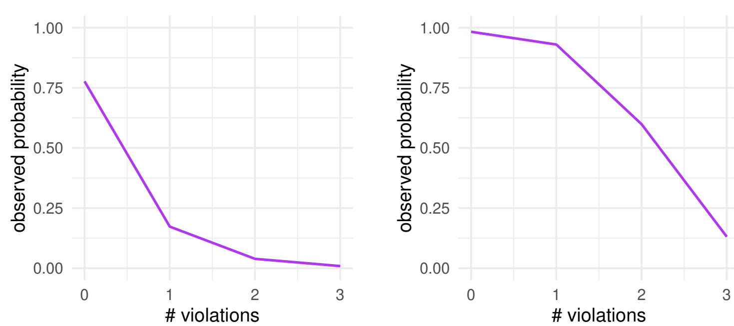
		MPARSE w_{mp}	*θ w_m	e^{Harmony}	predicted probability
hamper	hamper		0	1	1 / $1 + e^{-w_{mp}}$
	⊙	1		$e^{-w_{mp}}$	
theta	**th**eta		1	e^{-w_m}	e^{-w_m} / $(e^{-w_m} + e^{-w_{mp}})$
	⊙	1		$e^{-w_{mp}}$	
ba**th**oli**th**	ba**th**oli**th**		2	e^{-2w_m}	e^{-2w_m} / $(e^{-2w_m} + e^{-w_{mp}})$
	⊙	1		$e^{-w_{mp}}$	

Multiple competitions model structures are required in deriving "wug-shaped" curves, deemed to be MaxEnt's "quantitative signature" (Hayes, 2022; Kawahara, 2021).[1] The wug-shaped curve is a sigmoidal frequency pattern widely attested in various studies of quantitative and probabilistic patterns, including gradient phonotactic patterns. Ultimately, multiple competitions models also bring the modeling of phonotactics closer to the modeling of alternations, since in alternation patterns the structural candidates exhibiting cumulative phonotactic effects may in separate competitions and competing against different candidates.

This work finds that one-shot and multiple competitions MaxEnt model structures make different predictions about the extent to which accumulating violations affect probability. Specifically, one-shot models can only predict that additional violations take increasingly smaller hits on probability relative to earlier violations, therefore predicting probabilities that look "concave-up" across an increasing number of violations of the same constraint (Figure 1a). On the other hand, multiple competitions models can predict that later violations may take larger hits on probability relative to previous violations, producing probabilities that look "concave-down" across increasing violations (Figure 1b).

Figure 1: Examples of "concave-up" vs. "concave-down" patterns across accumulating violations of the same constraint.

(a) "concave-up" cumulativity **(b)** "concave-down" cumulativity



These findings are important in various ways. First, phonotactic patterns have been modeled using both one-shot and multiple competitions model structures in previous literature. However, it is unclear why a certain model structure is chosen over another and if such choice has crucial consequences for the analysis and modeling of the given empirical facts. Second, assuming that the one-shot model structure is the right model of phonotactics and that multiple competitions grammar structures are reserved for alternations, we expect phonotactics and alternations to show empirical differences in constraint cumulativity - as will be described in this paper, differences in the structure of the candidate competition affects how constraints interact in the grammar. If we assume instead that the multiple competitions model is a suitable model of phonotactics, it remains unclear how phonotactic learning proceeds when the null candidate ⊙ is unobservable from empirical

[1] Also see Bruce Hayes's Gallery of Wug-Shaped Curves: https://brucehayes.org/GalleryOfWugShapedCurves/index.htm.

data. Ultimately, these findings shed light on the grammar structures necessary for capturing cumulative constraint interactions and on their consequences for phonotactics and phonotactic learning in general.

2 Background

Constraint cumulativity in phonology refers to the general observation that two or more constraint violations together have an additive effect on the phonology of a language. Put simply, a form with $n + 1$ violations is somehow worse than a form with n violations. These "worsening" effects are attested in a wide range of empirical phenomena, such as in lexical frequencies, acceptability judgments, and rates of repair of marked structures.

An example of a cumulative phonotactic effect in lexical frequencies is found in the lexicon of Lakhota (Albright, 2008). In Lakhota, words with fricatives, ejectives, aspirates, and consonant clusters are relatively uncommon in the language, but are way more uncommon when there are two of these marked structures in the same word. For example, Albright (2008) reports that 32% of Lakhota CVCV words have a fricative as their first consonant and 18% have a fricative as a second consonant. However, only 1% of words have fricatives as both consonants. The expected proportion of words with two fricatives is approximately 6%, given the joint proportion of words with fricatives at first and second position (0.32 x 0.18 = 0.06). This shows that the presence of two fricatives in the same word has an effect larger than can be predicted from the simple combination of the two marked structures. Similar findings hold for words with various combinations of fricatives, ejectives, aspirates, and consonant clusters in Lakhota.

Constraint cumulativity is also attested in experimental settings. Pizzo (2015) finds that native English speakers judge noncewords with an illegal onset and an illegal coda (***tl**a**vb**) as worse than noncewords with only one of these marked structures (***tl**ag, *pla**vb**). And, in a series artificial grammar learning experiments, Breiss (2020) finds that participants rate forms with two violations of the learned backness harmony and nasal consonant harmony phonotactics as worse than forms with only one violation. These results were found for both counting and ganging cumulativity: counting cumulativity involves multiple violations of the same constraint (***po**n**u**m**i**te, *p**o**t**i**n**u**m**e** in the case of Breiss (2020)), while ganging cumulativity involves single violations of different constraints (***poni**) (Jäger and Rosenbach, 2006). In a follow-up study, Breiss and Albright (2022) find that the degree to which acceptability decreases for doubly-violating forms is a function of the strength of the learned phonotactic. They trained participants on a series of artificial grammars with differing levels of exceptions to backness harmony and nasal consonant harmony phonotactics and found that participants trained on grammars with more exceptions rate doubly-violating forms proportionally worse than singly-violating forms compared to participants that learned grammars with less exceptions and more robust phonotactics.

Finally, constraint cumulativity is also attested in the rates at which singly-marked vs. doubly-marked structures are repaired. For example, Smith and Pater (2020) find that French speakers are more likely to realize a schwa ([œ]) in forms where such schwa breaks up a three-consonant cluster (as in mã**ʒ l**[œ] **g**ato 'eat the cake') and in forms where it breaks up a stress clash (as in ló s[œ] vá 'the water was sold'). But, the rate of schwa realization is significantly higher when a consonant cluster and a stress clash are present in the same form and the realization of schwa simultaneously repairs both marked structures (as in yn vέ**st**[œ] **ʁ**úʒ 'a red jacket'). Another example is attested in Dioula d'Odienné (Mande, Ivory Coast) tonal melody alternations. Dioula bisyllabic indefinite nouns with two underlying L tones alternate to either L.H (Type I) or H.H (Type II) tonal melodies in the definite form, and Shih (2017) identifies three phonotactic factors that determine the type of the noun: (i) nouns with more sonorous final consonants are more likely to be Type II nouns, (ii) nouns in which the final vowel and intervocalic consonant agree in nasality are more likely to be Type II, and (iii) nouns whose last two vowels are identical are more likely to be Type II. Shih (2017) crucially finds that forms with these three phonotactic properties are overwhelmingly Type II compared to forms with two or one of these properties.

It is important to note that not all constraint-based frameworks in phonological theory model cumulative constraint interactions. Classic Optimality Theory (Prince and Smolensky, 1993) evaluates constraint violations according to their strict-ranking in the grammar, where the choice between candidates is solely determined by the highest-ranking constraint that distinguishes among them. In (3) below, only Constraint A determines which candidate is the winner even when Candidate 2 violates two constraints and Candidate 1 violates only one.[2]

(3)

	Constraint A	Constraint B	Constraint C
Candidate 1	*!		
☞ Candidate 2		*	*

Alternatively, Harmonic Grammar (HG; Legendre et al. (1990)) considers all constraint violations in its evaluation of the outcome. The Harmonic Grammar tableau below shows that a gang effect can now be predicted, given that the two lower-weighted violations together (but not separately) overtake the higher-weighted violation of Constraint A. Therefore, Candidate 1 is now predicted as the winner.

(4)

	Constraint A $w_A = 3$	Constraint B $w_B = 2$	Constraint C $w_C = 2$	H
☞ Candidate 1	*			-3
Candidate 2		*	*	-4

Harmony in HG is strictly linear by definition, so Harmonic Grammar does not predict concave-up or concave-down patterns of cumulativity. By definition, the harmony of a candidate is the negative sumproduct of the weight of the constraints and the number of constraint violations. The non-linearity in cumulativity that MaxEnt grammars install occurs in the exponentiation (and subsequent normalization) of Harmony.

3 Cumulativity in one-shot phonotactics

In evaluating the predictions of one-shot MaxEnt models regarding the effect of accumulating violations on probability, I assume the highly simplified grammar with counting cumulativity in (5) below. Recall that, in what I call "one-shot" models of phonotactics, all observed forms are in a single competition and all candidates' exponentiated harmonies are normalized by the same constant Z, which is the sum of all exponentiated harmonies of all candidates in the grammar (Hayes and Wilson, 2008).

(5)

inputs	MARKEDNESS w_m	Harmony	predicted probability
c_0	0	0	1 / **Z**
c_1	1	$-w_m$	e^{-w_m} / **Z**
c_2	2	$-2w_m$	e^{-2w_m} / **Z**
c_3	3	$-3w_m$	e^{-3w_m} / **Z**

In essence, one-shot MaxEnt grammars with counting cumulatiivity define the geometric series in (6) below, where each candidate is a term in the geometric series. The series sums to one because all candidates' probabilities sum to one in the grammar in (5).[3]

(6) $$\frac{1}{Z} + \frac{1}{Z}e^{-w} + \frac{1}{Z}(e^{-w})^2 + \frac{1}{Z}(e^{-w})^3 = 1$$

Note that the common ratio among the terms (i.e. candidates) in the series is e^{-w}, which means that probability in one-shot MaxEnt falls by a factor of e^{-w} at each additional violation. In other words, the predicted probability of a candidate with $n+1$ violations is $\frac{1}{e^{-w}}$ of the probability of the candidate with n violations. We can also deduce this from the proportional probability between a candidate with n violations and a candidate with $n+1$ violations, as in (7) below.[4]

[2] Constraint conjunction has been proposed as a mechanism for predicting gang effects in strict-ranking OT grammars (Smolensky, 2006; Itô and Mester, 1998, 2003), and is even proposed for grammars with weighted constraints (Shih, 2017). In (3), when a conjoined constraint composed of Constraint B and Constraint C (Constraint [B&C]) outranks Constraint A, Candidate 2 now violates the conjoined constraint and Candidate 1 is predicted as the winner.

[3] By definition, OT's candidates are an infinite set, therefore the geometric series in (6) should be a infinite sum. For simplicity and tractability, I focus on the predictions of one-shot models with candidates that show zero to three violations.

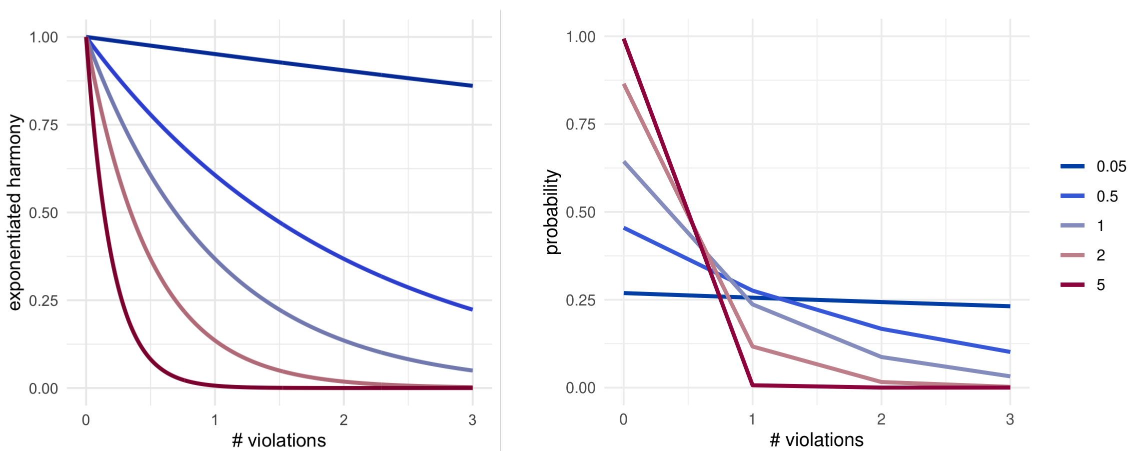
[4] The demominator Z in $P(c_n)$ and $P(c_{n+1})$ can be ignored since both terms have the same denominator.

(7) $$\frac{P(c_n)}{P(c_{n+1})} = \frac{e^{-wn}}{e^{-w(n+1)}} = \frac{e^{-wn}}{e^{-wn+(-w)}} = \frac{e^{-wn}}{e^{-wn} \times e^{-w}} = \frac{1}{e^{-w}}$$

Therefore, one-shot MaxEnt's predictions are all *concave-up*: a later violation always causes a smaller decrease in probability relative to the decrease in probability associated with the preceding violation, since the predicted probability at each additional violation is $\frac{1}{e^{-w}}$ of the probability of the immediately preceding violation.

To further concretize this, Figure 2 below plots one-shot MaxEnt's predicted probabilities across a wide range of weights of markedness. The plot on the left shows exponentiated harmony, and the plot on the right shows predicted probability (normalized exponentiated harmonies). Crucially, all curves in both exponentiated harmony and predicted probability are *concave-up*: the decrease in probability at a later violation is always less than the decrease in probability at the immediately preceding violation. Figure 2 also shows that one-shot MaxEnt's predictions are *increasingly concave-up*: the decrease in probability between, for example, the 0$^{\text{th}}$ and the 1$^{\text{st}}$ violation increases across increasing values of w_m, but the decrease in probability between the 1$^{\text{st}}$ and the $^{\text{2nd}}$ violation doesn't increase proportionally.

Figure 2: Predicted exponentated harmony and probability of one-shot MaxEnt model structures at increasing weights of markedness, across zero to three violations of the same constraint.



4 Cumulativity in multiple competitions phonotactics

To investigate the predictions of multiple competitions models for counting cumulativity, I take the same hypothetical candidates in (5) above but structure these into separate competitions, each competing against the null candidate ⊙ (Breiss and Albright, 2022). The null candidate ⊙ violates the MPARSE constraint only once, while structural candidates variably violate MARKEDNESS.

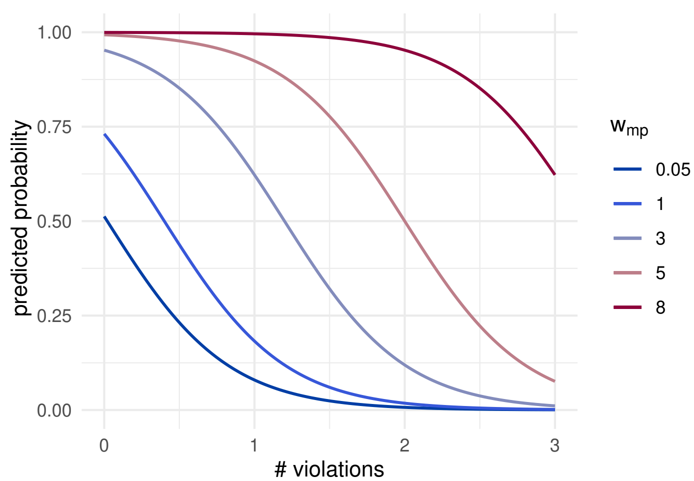
(8)

		MPARSE w_{mp}	MARKEDNESS w_m	Harmony	predicted probability
i_0	a. c_0		0	0	$1 / (1 + e^{-w_{pm}})$
	b. ⊙	1		$-w_{mp}$	
i_1	a. c_1		1	$-w_m$	$e^{-w_m} / (e^{-w_{mp}} + e^{-w_m})$
	b. ⊙	1		$-w_{mp}$	
i_2	a. c_2		2	$-2w_m$	$e^{-2w_m} / (e^{-w_{mp}} + e^{-2w_m})$
	b. ⊙	1		$-w_{mp}$	
i_3	a. c_3		3	$-3w_m$	$e^{-3w_m} / (e^{-w_{mp}} + e^{-3w_m})$
	b. ⊙	1		$-w_{mp}$	

The predicted probability of a given structural candidate c_n in multiple competitions MaxEnt models is $\frac{e^{-nw_m}}{e^{-nw_m}+e^{-w_{mp}}}$, where n is the number of violations of MARKENDESS for that candidate.[5] Therefore, the different structural candidates whose probabilities are used to evaluate the concavity of cumulativity are now in separate competitions and hence in separate probability distributions. Another core property of multiple competitions models is that the constraints are in an *asymmetric trade-off* (Pater, 2009). Namely, variable violations of MARKEDNESS are avoided by violating MPARSE only once. This is in opposition to a *symmetric trade-off*, where variable violations of MARKEDNESS are avoided by violating the opposing constraint by the same number of violations. Grammar models with symmetric trade-offs make very unnatural predictions for constraint cumulativity (see Appendix A).

Crucially, multiple competitions MaxEnt models predict a wide range of concavity patterns compared to one-shot models: they predict both *concave-up* and *concave-down* counting cumulativity. Figure 3 below shows predicted probabilities at various weights of MPARSE with the weight of markedness held constant at 2.5. Note that the type of concavity of the curve (concave-up or concave-down) is a function of the weight of MPARSE: at lower weights of MPARSE the grammar predicts concave-up curves, but concave-down patterns are predicted once the weight of MPARSE surpasses the weight of MARKEDNESS. As the weight of MPARSE increases, the markedness constraint loses its effect at the first few violations: the null candidate is highly penalized by higher weights of MPARSE and its competing structural candidate is assigned high probability.

Figure 3: Predicted probabilities of multiple competitions MaxEnt at varying weights of MPARSE, with the weight of MARKEDNESS held constant at 2.5.



Additionally, in multiple competitions models, the steepness of the predicted curve is a function of the strength of the phonotactic. When the weight of MPARSE is held constant and the weight of MARKEDNESS increases, the existing concave-up or concave-down pattern becomes more strongly concave-up or concave-down. This is visualized in the leftmost plot in Figure 4 below. The solid lines plot predicted probabilities at different weights of MPARSE when the weight of markedness is constant at 2.5. When w_m increases to 3 (dashed lines), the predicted curve becomes steeper, resulting in a stronger concave-up pattern for $w_{mp} = 1$ and a stronger concave-down pattern for $w_{mp} = 5$. Conversely, when the weight of markedness is decreased to 2 (dotted lines), the predicted cumulativity becomes less concave-up or concave-down.[6] Recall from Figure 3 that the type of concavity is determined by the weight of MPARSE, where higher weights of MPARSE predict concave-down patterns and lower weights of MPARSE predict concave-up patterns.

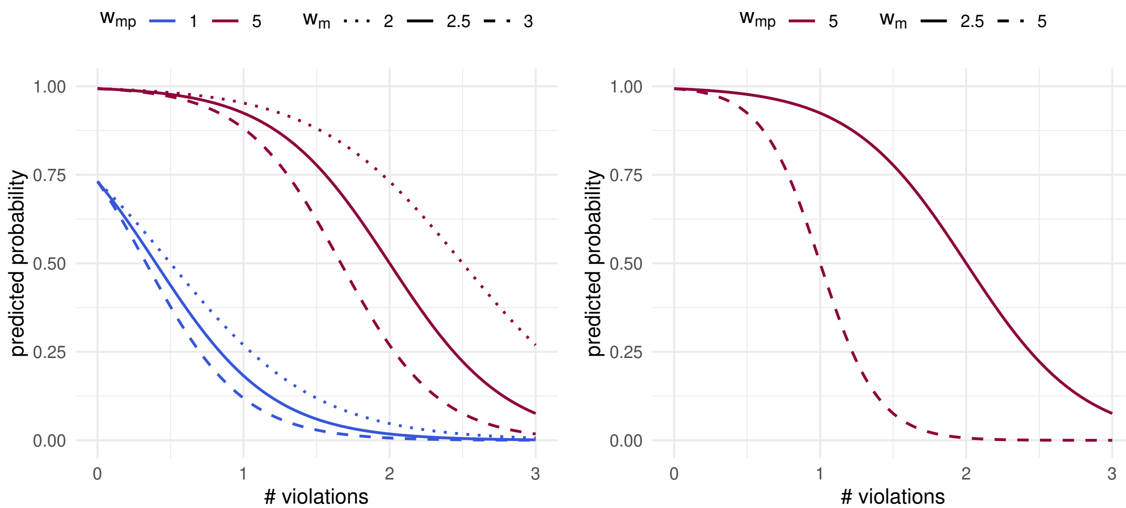
The last important property of multiple competitions models is that the location of the inflection point of the predicted probability curve depends on the relative weights of MPARSE and MARKEDNESS. As observed

[5] The predicted probability of a given nulll candidate $\odot_n$ is $\frac{e^{-w_{mp}}}{e^{-w_{mp}}+e^{-nw_m}}$, where n is the number of violations of MARKEDNESS of the structural candidate in competition with $\odot$.

[6] This is consistent with the experimental findings in Breiss and Albright (2022), where, at lower weights of markedness (hence grammars with more exceptions) participants rated doubly-violating forms proportionally worse than for grammars with stronger phonotactic restrictions.

in the rightmost plot in Figure 4, when the weight of MPARSE is exactly double the weight of MARKEDNESS (solid curve), the inflection point occurs at the second violation. And, when the weight of MPARSE is exactly the weight of MARKEDNESS, the inflection point of the curve occurs at the first violation (dashed line). Therefore, we can think of the MPARSE constraint as installing a "threshold of markedness", where grammars with lower weights of MPARSE tolerate less violations of markedness above the 50% line and grammars with higher weights of MPARSE allow for more violations of markedness above such line (Breiss and Albright, 2022).

Figure 4: Predicted probability of multiple competitions models across various weights of MPARSE (w_{mp}) and MARKEDNESS (w_m).

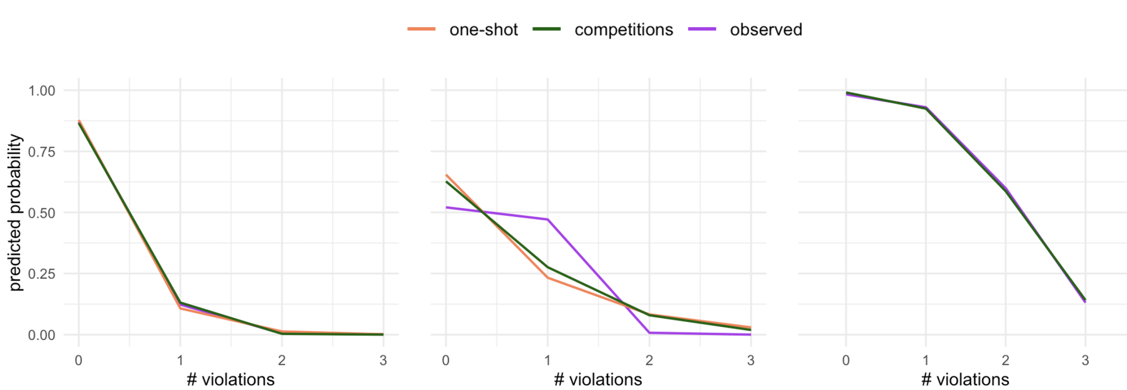


5 Learning concave-up and concave-down patterns

As a brief sanity check, we can further put these models to the test by having them learn different concave-up and concave-down patterns. Naturally, if the model cannot predict a certain pattern under any parameter setting, then it won't be able to learn such pattern either.

Figure 5 below shows the learning results for the one-shot and multiple competitions MaxEnt models under cumulative patterns with different concavities of cumulativity. Both the one-shot and multiple competitions MaxEnt models find a good fit for the given concave-down pattern (leftmost plot). However, neither the one-shot not multiple competitions models find a good fit for the concave-down pattern in the middle plot - both models fit a concave-up pattern that maximally (yet poorly) predicts the observed data. Finally, the rightmost plot shows that the multiple competitions model fits concave-down pattern whose probabilities are in different distributions (hence sum greater than one). Therefore, multiple competitions MaxEnt predicts only certain kinds of concave-down patterns under certain weighing conditions.

Figure 5: Learning results over simulated data with different concavities of cumulativity.



6 Discussion and closing

This paper finds that the structure of the candidate competition in probabilistic (MaxEnt) phonotactic models affects the way constraint violations interact. Specifically, one-shot MaxEnt models only predict that additional violations take a decreasing hit on probability relative to previous violations (concave-up patterns), while multiple competitions MaxEnt models can predict that later violations may take a greater hit on probability than earlier violations (concave-down patterns) under certain weighing conditions.

As briefly mentioned in the Introduction, a crucial advantage of the Hayes and Wilson (2008) one-shot model of phonotactics over the multiple competitions model is that constraint weights may be learned solely from observable data. In the multiple competitions model, the null candidate $\odot$ is unobservable from the data available to the learner, which raises the question of how phonotactic learning proceeds, if at all, under such model structure. If all null candidates are assigned an observed frequency of zero, the model will fail to learn a non-zero weight of MPARSE and will therefore only predict concave-up patterns of cumulativity even in the presence of a concave-down distribution (see Figure 3).

The way observed frequencies were assigned to null candidates when having these learn different concave-up and concave-down patterns of cumulativity (Section 5) was to assign these a frequency of $1 - p$, where p is the frequency of the structural candidate. For the concave-up pattern, this p frequency of the structural candidate was obtained by assuming that structural candidates are simultaneously in two probability distributions: in their smaller distribution with their null candidate, and in a larger cross-competition distribution with all structural candidates. However, this approach is not possible for concave-down patterns: we can't assume that the structural candidates are in the same cross-competition distribution (and also in probability distributions with their respective null candidates) because the concave-down patterns that multiple competitions models predict necessarily have structural candidates with probabilities that sum to greater than one. So, as we saw in Section 5, structural candidates *must* be in separate frequency distributions in order for multiple competitions MaxEnt models to predict concave-down cumulativity. Then, how might we possibly assign observed distributions for multiple competitions grammars to make these a viable model of phonotactic learning? I propose two alternatives for future work to explore.

The first alternative is to assume that, instead of the structural candidates constituting only of observed words in a language, all structural candidates are provided by GEN, the component of optimality-theoretic grammars that generates all possible candidate forms in the language (Prince and Smolensky, 1993). Structural candidates that are attested in the language are assigned their respective frequencies, and structural candidates that remain unobserved have a frequency of 0 and their opposing null candidate is assigned a frequency of 1. This way, null candidates are now assigned a frequency, and observed frequency distributions can be defined across all competitions regardless of the concavity of the pattern.

The second alternative is to assume that speakers employ different grammar structures for different tasks. For example, speakers might learn constraint weights with the one-shot Hayes and Wilson (2008) model structure, therefore they learn from the distribution of observable forms in their data. But, when performing acceptability judgments, speakers might employ the multiple competitions grammar structure with the learned constraint weights. Notice that the multiple competitions model structure recapitulates the structure of a binary "yes"/"no" judgment task. In modeling such tasks, the frequency of "yes" responses to a given form corresponds to the frequency of the structural candidate, and the frequency of "no" responses corresponds to the frequency of the null candidate. Essentially, learning constraint weights and providing acceptability judgments are different tasks for the speaker, and we can model these differences as the use of different task-specific grammar structures. However, it remains an open question in this approach how the weight of MPARSE is learned, especially when the learning data contains a concave-down pattern of cumulativity. Future empirical studies should test if learners generalize concave-down patterns from concave-down data, and if learners vary in the concavities of cumulativity they generalize.

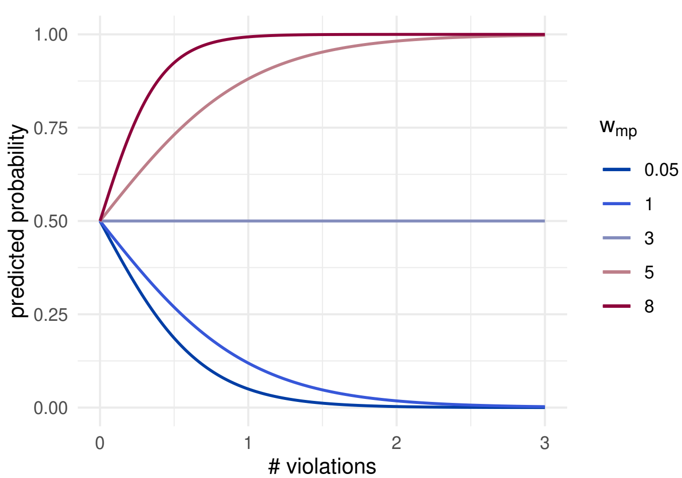
Finally, most previous literature uses the notion of *linearity*, not concavity, to describe and analyze the extent to which later violations decrease probability and acceptability and increase rates of repair. The linearity of cumulativity is defined as the comparison between observed and expected penalties, where the expected penalty is the joint combination of the penalties associated with singly-violating forms. For example, the Lakhota cumulative phonotactic effect of fricatives is a case of *superlinear* cumulativity because the observed penalty of multiply-violating forms is lower than expected (Albright, 2008). The expected proportion of CVCV words with two fricatives is 6%, given that 32% of words have a fricative as the first consonant and 18% of words have a fricative as the second consonant, but the observed proportion of words with two fricatives is only 1%. Conversely, *sublinear* patterns of cumulativity occur when the observed probability is higher than is expected. Although concavity and linearity are both ways of describing and analyzing the extent to which probability falls across an increasing number of violations, it is important to note that they describe different kinds of cumulative patterns. For example, concave-up patterns may be sublinear, approximately linear, or superlinear. But, all concave-down patterns are superlinear: since the penalty introduced by the first violation is relatively low (and probability/acceptability are relatively high) in concave-down patterns, the combination of the single penalties, the expected penalty, is always higher than what is observed at a later violation. In this paper I describe different kinds of cumulative patterns using the notion of concavity, and not linearity, because concavity clearly demarcates the differences between the predicted hypothesis spaces of the models: one-shot models can only predict concave-up cumulativity, but multiple copetitions models may predict both concave-up and concave-down cumulativity.

To conclude, although this paper investigates the predictions of one-shot and multiple competitions model structures specifically for cases of counting cumulativity (accumulating violations of the same constraint), the findings described here also generalize to ganging cumulativity (accumulating violations of different constraints). One-shot MaxEnt models can predict that a second violation has a larger effect than the first only when that second violation is of a markedness constraint with a higher weight than the weight of the constraint violated first. In other words, one-shot models fail to predict patterns where the independent violations of both markedness constraints have a small effect on probability but the simultaneous violation of both constraints causes a larger decrease in probability. Such "doubly concave-down" pattern is successfully predicted by multiple competitions MaxEnt grammars at high weights of MPARSE, consistent with the findings for counting cumulativity elaborated in this paper. Additionally, this paper also focuses on the predictions of Maximum Entropy models, but the one-shot and multiple competitions model structures can also be compared assuming other probabilistic phonological models such as Noisy Harmonic Grammar (Boersma and Pater, 2016) and Stochastic OT (Boersma, 1997; Boersma and Hayes, 2001).

Appendix A: the unnatural predictions of symmetric trade-offs

Recall that a crucial property of multiple competitions models is that the constraints are in an asymmetric trade-off, where variable violations of MARKEDNESS always trade-off against a single violation of its opposing constriant MPARSE (Pater, 2009). Conversely, a symmetric trade-off occurs when multiple violations of MARKEDNESS trade-off for the same number of violations of the opposing constraint. Symmetric trade-offs in multiple competitions grammars make very unnatural predictions. These unnatural predictions occur when the weight of MPARSE surpasses the weight of MARKEDNESS. Under these weighing conditions, the model predicts increasing probabilities across an increasing number of violations of MARKEDNESS, which is, naturally, a prediction not expected to hold empirically.

Figure 6: The unnatural predictions of symmetric trade-offs in multiple competitions MaxEnt models at varying weights of MPARSE when weight of MARKEDNESS is 3.



References

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

Boersma, Paul. 1997. How we learn variation, optionality, and probability. In *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*.

Boersma, Paul, and Bruce Hayes. 2001. Empirical tests of 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. Equinox Press.

Breiss, Canaan. 2020. Constraint cumulativity in phonotactics: evidence from artificial grammar learning. *Phonology* 37:551–576.

Breiss, Canaan, and Adam Albright. 2022. Cumulative markedness effects and (non-)linearity in phonotactics. *Glossa: a journal of general linguistics* 7:1–32.

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

Hayes, Bruce. 2022. Deriving the wug-shaped curve: A criterion for assessing formal theories of linguistic variation. *Annual Review of Linguistics* 8:473–494.

Hayes, Bruce, and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistics Inquiry* 39:379–440.

Itô, Junko, and Armin Mester. 1998. Markedness and word structure: OCP effects in Japanese. Ms., University of California, Santa Cruz.

Itô, Junko, and Armin Mester. 2003. *Japanese morphophonemics: markedness and word structure*. MIT Press.

Jäger, Gerhard, and Anette Rosenbach. 2006. The winner takes it all - almost: Cumulativity in grammatical variation. *Linguistics* 44:937–971.

Kawahara, Shigeto. 2021. Testing MaxEnt with sound symbolisms: a stripy wug-shaped curve in Japanese Pokémon names. *Language* 97:e341–e359.

Legendre, Géraldine, Yoshiro Miyata, and Paul Smolensky. 1990. Harmonic Grammar: a formal multi-level connectionist theory of linguistic well-formedness: an application. In *Proceedings of the 12th Annual Conference of the Cognitive Science Society*, 884–891. Erlbaum.

McCarthy, John J., and Matthew Wolf. 2005. Less than zero: Correspondence and the null output. Ms. University of Massachussets Amherst.

Pater, Joe. 2009. Weighted constraints in generative linguistics. *Cognitive Science* 33:999–1035.

Pizzo, Presley. 2015. Investigating properties of phonotactic knowledge through web-based experimentation. Doctoral Dissertation, University of Massachusetts Amherst.

Prince, Alan, and Paul Smolensky. 1993. Optimality Theory: constraint interaction in generative grammar. ROA.

Shih, Stephanie. 2017. Constraint conjunction in weighted probabilistic grammar. *Phonology* 34:243–268.

Smith, Brian, and Joe Pater. 2020. French schwa and gradient cumulativity. *Glossa: a journal of general linguistics* 5:1–33.

Smolensky, Paul. 2006. Optimality in phonology II: harmonic completeness, local constraint conjunction, and feature domain markedness. In *The harmonic mind: from neural computation to optimality theoretic grammar*, ed. Paul Smolensky and Géraldine Legendre, 27–160. MIT Press.