One-shot vs. competitions phonotactics in modeling constraint cumulativity

Marisabel Cabrera (UCLA)

Background. Constraint-based probabilistic grammars like MaxEnt (Goldwater & Johnson 2003) differ with respect to how the input data is organized and evaluated. On the one hand, in what I call "one-shot" models of phonotactics, all words are in the same competition and the grammar assigns a single probability distribution over all candidates (Hayes & Wilson 2008). On the other hand, in what I call "competitions" models of phonotactics, inputs are in separate competitions and are compared against a single candidate that violates an opposing binary constraint (Hayes 2021; Kawahara 2021; Breiss & Albright 2022). This latter model derives MaxEnt's "quantitative signature" – sigmoid curves.

Proposal. This work finds that the "one-shot" and "competitions" models (Figure 1) make distinct and testable predictions regarding *cumulative constraint interactions* – the phonological phenomenon wherein forms with multiple constraint violations show decreased lexical frequency and acceptability in phonotactics (Albright 2008; Breiss 2020) and increased rates of repair (Smith & Pater 2020; Kim 2022) compared to forms with less violations. The predictions of these models follow directly from their structure. The one-shot model only predicts that subsequent violations take a *decreasing* hit on probability relative to previous violations, while the competitions model can predict that subsequent violations take a *greater* hit on probability (under certain weighting conditions). Since patterns wherein greater hits on well-formedness as violations accumulate are attested, I argue that the "competitions" approach to modeling phonotactics is necessary, as the one-shot model fails to straightforwardly analyze the full range of cumulative phonotactic effects. I present results for counting cumulativity (multiple violations of the *same* constraint), but analogous results are found for ganging cumulativity (violations of *different* constraints).

Model probability space. I first compare the predicted probability spaces of the one-shot and competitions models in counting cumulativity. In the one-shot model, predicted probability across

***(1)** the "one-shot" and "competitions" models in counting cumulativity.*

(a) the "one-shot" model

cands	Mark w_m	H	p
0 viols	0	0	?
1 viol	1	-w	?
2 viols	2	-2w	?
3 viols	3	-3w	?

$\sum_c P(c) = 1$

(b) the "competitions" model

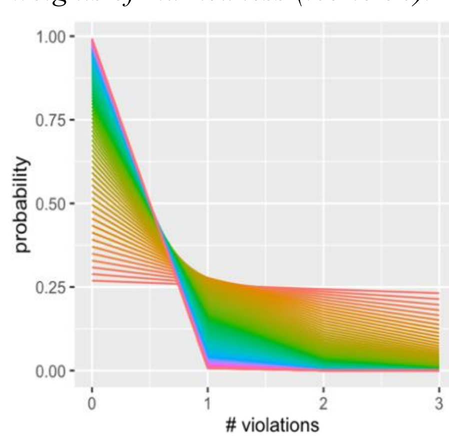
input	cands	Mark w_m	MParse w_{mp}	H	p
0 viols	0 viols	0		0	?
	⊙		1	$-w_{mp}$	
1 viol	1 viol	1		$-w_m$	?
	⊙		1	$-w_{mp}$	
2 viols	2 viols	2		$-2w_m$	?
	⊙		1	$-w_{mp}$	
3 viols	3 viols	3		$-3w_m$	?
	⊙		1	$-w_{mp}$	

$\sum_c P(c) = 1$ (for each input)

***(2)** Proportional probability across n violations.*

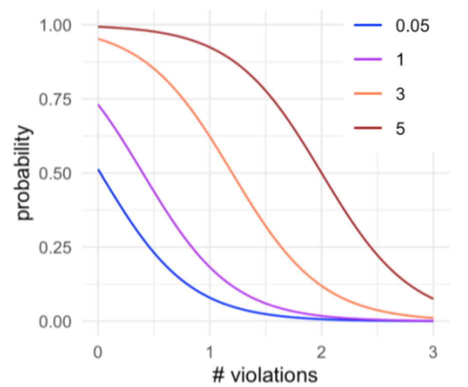
$$\frac{P(c_n)}{P(c_{n+1})} = \frac{e^{-nw}}{e^{-(n+1)w}} = \frac{1}{e^{-w}}$$

***(3)** One-shot model probability space. Lines plot predicted prob. at varying weights of markedness (.05 to 50).*



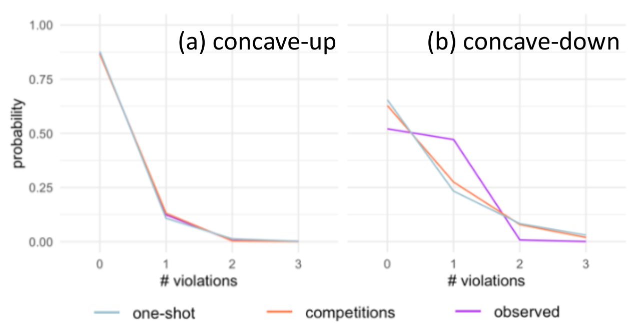
subsequent violations changes proportional to $\frac{1}{e^{-w}}$, as in (2). The plot in (3) shows that the one-shot model only predicts "concave-up curves": the effect of subsequent violations on probability diminishes as they accumulate. On the other hand, the competitions model predicts a wider range of cumulative interactions. As shown in (4), the model predicts both concave-up and concave-down curves: subsequent violations have a greater hit in probability when the weight of MPARSE is high compared to the weight of markedness (red and orange lines). The competitions model can predict concave-down curves given its ability to install a "threshold of markedness" (Hayes 2021; Breiss & Albright 2022).

(4) Competitions predictions for w_m = 2.5 at different weights of MParse.

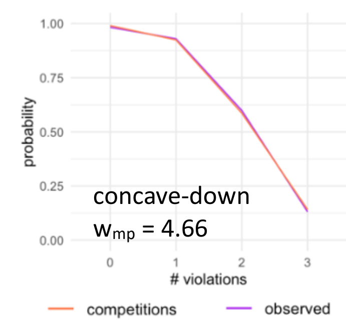


Learning. We next fit the one-shot and competitions models on empirically-motivated simulated data with different kinds of cumulative behavior. Consistent with the previous result, both models can fit concave-up curves (example in 5a), but neither model can fit concave-down curves whose probabilities sum to one (example in 5b). Instead, both models fit the best concave-down curve. However, the competitions model fits concave-down curves when the MPARSE constraint gets substantial weight, as shown in (6). Therefore, we can only predict that subsequent violations have a greater hit in probability when structural candidates are in separate distributions, as in the competitions model.

(5)



(6)



Implications. These results provide novel insights as to how our particular assumptions of the structure of probabilistic grammars, frequently used in phonotactic modeling, influence their predictions. However, phonotactic learning is only possible under the competitions model assuming negative evidence: in this model, the null candidate is assigned observed frequency when competing against an unobserved input provided by GEN. The null candidate ⊙ is otherwise unobservable when learning phonotactics only from positive evidence, as in Hayes & Wilson (2008). The competitions model also obviates the need for additional mechanisms such as constraint conjunction (Shih 2017) and the exponentiation of violations (Kim 2022). These are proposed as ways to directly lower the probability of 2+ violations, which MaxEnt otherwise fails to predict. Additionally, the competitions model brings phonotactics closer to the modeling of alternations since both models have comparable structures when in an asymmetric tradeoff (Pater 2009). The current difference in the modeling of phonotactics and alternations is problematic for cumulativity since both domains show "concave-down" patterns (Breiss & Albright 2022; Kim 2022), which, as it stands, do not receive a straightforward analysis under one-shot models.