

Globally Normalizing MaxEnt

Andrew Hennessey
University of York

Key Words: Stratal OT, MaxEnt, global normalization

Concerns over Stratal OT: Among the criticisms of Stratal OT (e.g. Kiparsky 2015) are 1) the learnability (or lack thereof) of a multi-level cophonological system, 2) the purported need for intra-level opaque effects when relying on the standard three-level model, and 3) the fact that expanding the number of levels in the model to mitigate the second issue exacerbates the first issue and might undermine the theoretical underpinnings of the model (e.g. McCarthy 2007, Obiri-Yeboah & Rasin 2025). An expanded set of viable constraint typologies is possible but is at the expense of one of the appeals of the model: that the multi-level structure bypasses the need for controversial and exotic constraint types.

In addition, standard Stratal OT is a deterministic model and lacks the mechanisms in stochastic OT models like Maximum Entropy Grammar (MaxEnt; Goldwater & Johnson 2003) to tackle gradience and variability. Opaque phenomena can also be variable and gradient (e.g. Vance 1987's descriptions of variable outcomes in Canadian raising), and it would appear that probabilistic models would be better suited to analyze them. Indeed, Nazarov & Pater (2017) provided a version of Stratal MaxEnt in which each level chooses a winning candidate probabilistically, and in any given running of the grammar, that winner then feeds into the next level as the input. However, this model too relies on cophonologies (as in (1)), and as such the primary concerns listed above still stand:

(1) Example of cophonologies in Canadian raising analysis from Nazarov & Pater (2017 p. 311)

Word level				Phrase level			
ID[son]	ID[low]	*'VTV	*aI[-vce]	ID[son]	ID[low]	*'VTV	*aI[-vce]
10.44	5.02	0	11.13	0	6.81	6.12	0

Nazarov & Pater found that, no matter the combinations of weights used, if the same set of weights is used in each level, the system makes the wrong predictions. The question becomes, is there a stratal model that can simultaneously account for both variable and opaque outcomes and dispense with cophonologies?

Novel Approach – Globally Normalizing MaxEnt: This presentation describes Globally Normalizing MaxEnt (GNoME; AUTHOR, submitted). Like Nazarov & Pater's model, GNoME uses multiple strata to account for opacity: candidates at one level generate their own candidate sets at the next, and surface probabilities are calculated over derivational paths. GNoME differs in when normalization – converting raw harmony scores into probabilities – occurs. Nazarov & Pater normalize at every stratum and multiply the resulting probabilities along a path, the final probability times the probabilities of each of its progenitors at previous strata; GNoME sums raw harmony scores along each path and normalizes once, at the final stratum. Weights can therefore be optimized against every violation vector in the derivation, giving the best single-weight-vector fit.

A well-known example from Canadian Raising is shown in (2). Opaque *writer* ['ɾɹɪ.rə] and transparent *rider* ['ɹɪ.rə] are derived with one weight vector (Table 1). Four strata are employed to accommodate different affix levels, although in this example, only one affix is necessary. Below, the best possible paths to the surface for each variant are presented. Both the opaque and transparent forms are almost guaranteed to win.

(2) GNoME paths for *writer* and *rider* (probabilities over all complete paths)

	writer /rait-ə/ <i>opaque: raised before a surface flap</i>		rider /raid-ə/ <i>transparent: unraised before a voiced flap</i>	
	unraised → ['ɹɪ.rə]	raised → ['ɾɹɪ.rə]	unraised → ['ɹɪ.rə]	raised → ['ɾɹɪ.rə]
S1 stem	16.1 *Cai/ [-vce]ft	24.9 IDENT-low + IDENT-F(root)	0	24.9 IDENT-low + IDENT-F(root)
S2 -ə	38.2 IDENT-vce + *(CaiCV)	31.9 IDENT-vce	6.2 *(CaiCV)	0
S3	6.2 *(CaiCV)	0	6.2 *(CaiCV)	0
S4	6.2 *(CaiCV)	0	6.2 *(CaiCV)	0
Sum	66.7	56.8	18.7	24.9
P	0.0001	0.9999	0.998	0.002

Harmony is summed along each path and normalized once via the following formula:

$$(3) P(d) = e^{-H(d)} / \sum_d e^{-H(d)}$$

The key mechanism is the behavior of faithfulness constraints. Every candidate at one level becomes an input at the next level and spawns its own set of candidates; that progenitor form, regardless of its violation vector profile at the prior level is now an input, and a candidate identical to it will have no faithfulness violations. However, if it has not fixed a markedness issue, it will continue to be penalized for that markedness issue. This penalty is imposed at every stratum and the violations add up. For faithfulness constraints to keep up, they often have to have very high weights, because their impact is diffused over the remainder of the derivation.

Table 1: Constraint weights	
*Cai/ [-vce]ft	16.08
*(CaiCV)	6.22
FLAP[+son]	3.76
*'VTV	23.54
IDENT-son	0
IDENT-vce	31.95
IDENT-low	8.93
ID-low('V)	18.28
IDENT-F(root)	15.96
IDENT-Ft	24.89

In *writer*, unraised ('ɹɪ)t leads after stratum 1 but keeps incurring *(CaiCV), a constraint penalizing a disyllabic foot with an unclipped diphthong, and as a result the raised path wins ($P \approx 0.9999$). In *rider*, the same weights favour the unraised path ($P \approx 0.998$).

Advantages: GNoME produces transparent, opaque, and variable outputs (not shown but analyzed are variable forms *cider*, *tiger* and *Cyclops*), with the same weights at every level. Without cophonologies, any number of levels can be used, which potentially simplifies the treatment of intra-stratum opacity and accommodates any morphosyntactic typology. While this presentation assumes that the number of levels is not known (and potentially might vary from language to language), like other stratal models, it also assumes that the number is constrained by the language's morphosyntax.