Equiprobable mappings in weighted constraint grammars

Arto Anttila
Stanford University
anttila@stanford.edu

Scott Borgeson
Stanford University
borgeson@stanford.edu

Giorgio Magri
CNRS
magrigrg@gmail.com

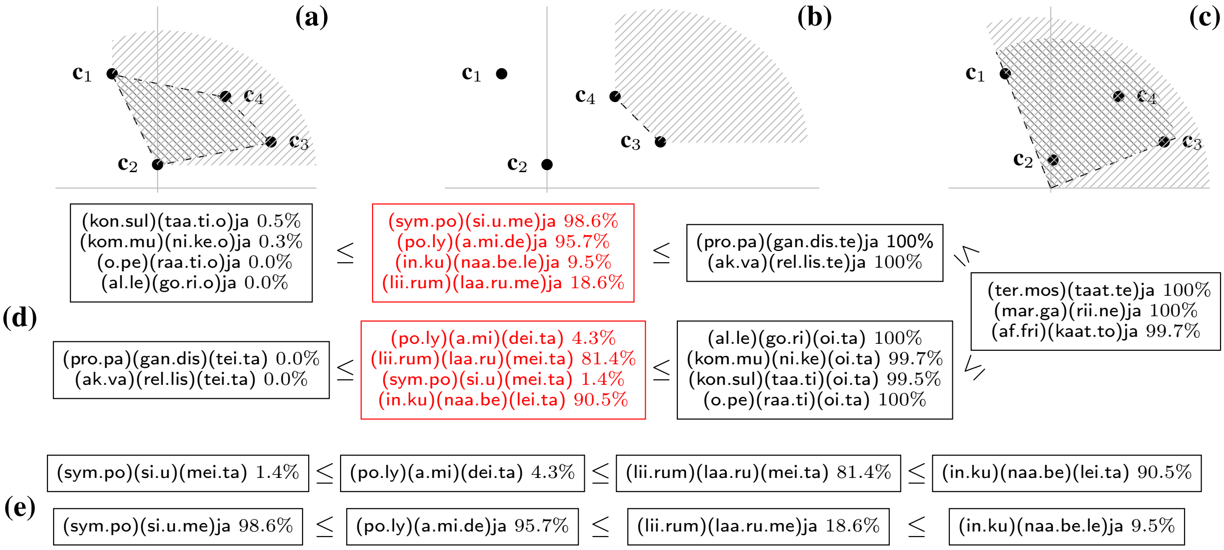
Recent literature (e.g., Smith and Pater 2017) documents cases that admit a better fit in *Maximum Entropy* (ME; Goldwater and Johnson, 2003; Hayes and Wilson, 2008) than in *Stochastic* (or *Noisy*) HG (SHG; Boersma and Pater, 2016). ME is thus richer than SHG. How much richer? This paper addresses this question by comparing ME and SHG in terms of their *equiprobable mappings*.

Equiprobability - A phonological process applies uniformly to all forms that belong to a *natural class* because they share some relevant properties while differing in irrelevant ways. For instance, vowel harmony targets backness but ignores number of syllables. The Finnish mappings (/maa-nä/, [maana]) and (/rakastaja-nä/, [rakastajana]) differ in length, but are *equivalent* for vowel harmony (back). These equivalences are a key property of phonology. How should they be extended to probabilistic phonology? A *probabilistic* grammar assigns to each UR a probability distribution $\mathbb{P}(\text{SR} \mid \text{UR})$ over the set of candidate SRs. Two mappings (UR, SR) and $(\widehat{\text{UR}}, \widehat{\text{SR}})$ are *equiprobable* if every grammar in the typology assigns them the same probability: $\mathbb{P}(\text{SR} \mid \text{UR}) = \mathbb{P}(\widehat{\text{SR}} \mid \widehat{\text{UR}})$. We submit that equiprobability is the proper way of extending phonological equivalence from categorical to probabilistic phonology. E.g., the fact that words that only differ for length are equivalent for harmony means they have the same probability of harmonizing: $\mathbb{P}(\text{[maana]} \mid \text{/maa-nä/}) = \mathbb{P}(\text{[rakastajana]} \mid \text{/rakastaja-nä/})$.

ME - Given a winner and a loser mapping, their *difference vector* consists of the constraint violations of the loser discounted by the violations of the winner. Suppose the mapping (UR, SR) has 4 difference vectors $\mathbf{c}_1, \ldots, \mathbf{c}_4$. The gray region in fig. (a) is their *convex hull*. The lightgray region consists of points larger than a point in this convex hull. Two mappings (UR, SR) and $(\widehat{\text{UR}}, \widehat{\text{SR}})$ are equiprobable in ME iff they define the same lightgray region. The vectors $\mathbf{c}_1$ and $\mathbf{c}_2$ are *extreme points*: they determine the shape of the lightgray region and must therefore be shared by the two mappings in order for them to share the lightgray regions. The vectors $\mathbf{c}_3$ and $\mathbf{c}_4$ are instead *interior points*: they do not contribute to the shape of the region. Yet, since we have established that $\mathbf{c}_1$ and $\mathbf{c}_2$ are shared, we can effectively "peel them off" the two sides of the ME probability identity. In other words, we can ignore $\mathbf{c}_1$ and $\mathbf{c}_2$ and only focus on $\mathbf{c}_3$ and $\mathbf{c}_4$. They are extreme points of the new lightgray region in fig. (b) and must thus be shared. And so on. In conclusion, the two mappings (UR, SR) and $(\widehat{\text{UR}}, \widehat{\text{SR}})$ are equiprobable in ME iff they share exactly the same set of difference vectors. Realistically, this happens only if (UR, SR) and $(\widehat{\text{UR}}, \widehat{\text{SR}})$ are the same mapping. ME thus admits no equiprobable mappings.

SHG - The gray region in fig. (c) is the *convex cone* of the difference vectors $\mathbf{c}_1, \ldots, \mathbf{c}_4$. The lightgray region consists of points larger than a point in this cone. Indeed, the geometry of SHG is analogous to that of ME, with cones in place of convex hulls. Two mappings (UR, SR) and $(\widehat{\text{UR}}, \widehat{\text{SR}})$ are SHG equiprobable iff they define the same lightgray region. The difference vector $\mathbf{c}_1$ sits on the border but can be shifted (rescaled) without affecting the lightgray region. The equiprobable mapping $(\widehat{\text{UR}}, \widehat{\text{SR}})$ thus needs *not* share this difference vector $\mathbf{c}_1$ but only a rescaling thereof. Furthermore, nothing can be said in this case about the interior vectors $\mathbf{c}_2, \ldots, \mathbf{c}_4$. In conclusion, the two mappings (UR, SR) and $(\widehat{\text{UR}}, \widehat{\text{SR}})$ are equiprobable in SHG iff each *non-interior* difference vector is a *rescaling* of a non-interior difference vector of the other mapping. This SHG condition is weaker than than the ME condition above. First, because ME requires *identity* of difference vectors while SHG only requires *rescaling*. Second, because ME looks at *all* difference vectors while SHG ignores *interior* ones. SHG thus does admit equiprobable mappings.

Test case - We test ME's and SHG's predic-



tions on Finnish secondary stress. In Finnish, (i) primary stress falls on the initial syllable; (ii) secondary stress falls on every other syllable after that (iii) except that a light syllable is skipped if the syllable after that is heavy; (iv) unless that heavy syllable is final (Hanson and Kiparsky 1996). The skipping clause (iii) exhibits probabilistic variation in long words: both (pró.fes.so)(rìl.la) (with skipping) and (pró.fes)(sò.ril)la (without skipping) are attested. The rate of skipping depends on vowel quality and preceding syllable weight (Anttila 2012). Despite secondary stress being hard to hear, Finnish has a segmental alternation that can be used as stress diagnostic: a short underlying /t/ is deleted when *extrametrical*. Thus, skipping correlates with t-retention, as in (pro.fes.so)(rèi.ta); no-skipping correlates with t-deletion, as in (pró.fes)(sò.re)ja.

To model this distribution of Finnish secondary stress, we constructed an input space consisting of 48 noun types systematically varying stem length, syllable weight, and vowel quality. These phonological forms are evaluated by eight constraints capturing the phonological factors mentioned above. We computed SHG/ME uniform probability inequalities for this model using $\mathbb{C}$o$\mathbb{G}$e$\mathbb{T}$o (available online at `[omitted]`), a suite of $\mathbb{T}$ools for studying SHG and ME based on their rich underlying $\mathbb{C}$onvex $\mathbb{G}$eometry, as illustrated by the results above. SGH predicts seven blocks of equiprobable mappings ordered through uniform probability inequalities (denoted $\leq$) as in fig. (d). This confirms the formal result above that SHG does allow for equiprobable mappings.

To evaluate these predicted equiprobable blocks, we computed the observed t-retention/deletion rates for each stem type in a corpus of approximately 9 million nouns (tokens). The five black SHG-equiprobable blocks are consistent with the data (all stems are nearly categorical), but the two red blocks are problematic. Yet, the difference between t-deletion/retention rates for stems of the liirumlaarumi- and inkunaabeli-type is not statistically significant ($\chi^2 = 2.9849$, $df = 1$, $p = 0.08404$). Furthermore, there are only two stems in the symposiumi-type and both could be re-analyzed as 4-syllable stems, consistently with their high t-deletion rate (Anttila and Shapiro 2017). We have no explanation for the high t-deletion rate for stems of the polyamidi-type ($N = 69$). We conclude that the Finnish data are generally consistent with SHG's predictions.

Does ME offer a more principled treatment of the two problematic red blocks? That is not the case. In fact, as expected given the formal result above, ME breaks up these two red equiprobable blocks and orders their stem types through uniform probability inequalities as in fig. (e). On the retention side (top row), ME seems promising: corpus frequencies mirror the predicted probability inequalities. Yet, on the deletion side (bottom row), ME fails to flip the inequalities, yielding the opposite of what we observe. Such counterintuitive probability reversals seem to recur in ME.

Addendum - OT induces even more equiprobable blocks than HG: it predicts "syllable counting" by grouping together odd-parity stems of different lengths, pointing at a linguistically interesting difference between ranked and weighted constraints.