

# The Calibrated Error-Driven Ranking Algorithm as a Solution to Oscillation in Antagonistic Constraints: A Necessary Bias for Algorithmic Learning of Kihnu Estonian

Kaili Vesik

*University of British Columbia*

## 1 Introduction

Kihnu Estonian (KE) is a variety of Estonian (Balto-Finnic) that has progressive, front-back vowel harmony. Using Kiparsky & Pajusalu’s (2003; hereafter K&P) proposed analysis of harmony in Balto-Finnic languages as a starting point, I present a set of constraints to account for the vowel harmony (VH) patterns attested in KE. Furthermore, I use this constraint set as the structure within which to apply the Gradual Learning Algorithm (GLA; Boersma & Hayes, 2001), with the aim to model the acquisition of KE using a theoretical framework that considers the broader vowel typology of Balto-Finnic languages. However, GLA learning simulations show that antagonism of constraints in the set I adopt poses an obstacle in learning.

Antagonistic constraints in an Optimality-Theoretic (OT) gradual learning framework have the potential to oscillate with respect to each other in terms of their constraint values or ordinal rankings, due to the fact that promoting one necessarily involves demoting the other. Such oscillation can be either beneficial or detrimental to the goal of learning the grammar in question. Magri & Storme (2020) argue that calibration of constraint promotion (Magri, 2012) is not useful in learning variation in stochastic OT. However, I show here that such calibration, though originally proposed to address the Credit Problem (Dresher, 1999), elegantly solves the learning problems caused by oscillating constraints in this context.

## 2 Learning Kihnu Estonian

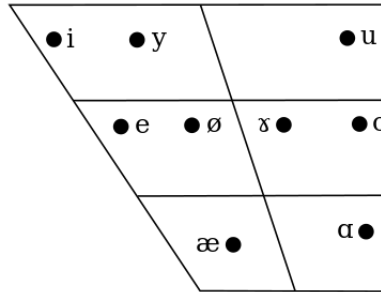
This paper is grounded in the modeling of acquisition of Kihnu Estonian VH. Following K&P, I investigate the Optimality-Theoretic characteristics that are required to describe the vowel patterns of KE within the broader context of other Balto-Finnic languages. I propose a constraint set sufficient to construct a grammar correctly describing KE. Following this, I analyze the parameters and biases necessary for a gradual learning algorithm to acquire a restrictive grammar for KE using the proposed constraint set. I show that even for simple patterns such as straightforward VH, a bias such as the update rule used in the Calibrated Error-Driven Ranking Algorithm (CEDRA; Magri, 2012) needs to be built in to the learning process for an algorithmic learner to successfully acquire the dialect in question.

**2.1 *Kihnu Estonian vowel harmony*** Estonian is in the Balto-Finnic branch of the Uralic language family, and the Kihnu variety is spoken primarily on the island of Kihnu. KE has the same nine-vowel inventory as Standard Estonian: /i, e, a, o, u, ɤ, æ, ø, y/ (Asu & Teras, 2009; Asu et al., 2012; Léonard, 1993). Figure 1 illustrates the Estonian vowels.

KE is one of just a few Estonian varieties with VH. VH in KE is front/back and progressive, with alternating pairs /y/-/u/, /ø/-/o/, /æ/-/ɑ/, and /e/-/ɤ/, and neutral transparent /i/ (Léonard,

---

\* My thanks to Anne-Michelle Tessier, Gunnar Ólafur Hansson, participants at AMP 2022 (particularly Bruce Hayes), and the UBC Child Phonology Lab. This research is funded in part by a SSHRC CGS-D.



**Figure 1:** Estonian vowels (Asu & Teras, 2009).

1993; Sang, 2009); note that this paper focuses on the distribution of monophthongs only.<sup>1</sup>

The examples in (1) show that individual words are either back-harmonic (containing /ɑ, o, u, ɤ/) or front-harmonic (containing /æ, ø, y, e/), with transparent /i/ appearing in both types of words.

- (1) KE VH examples (Lindström, 2013). Note /e/-/ɤ/ alternation in the 1PL suffix [me ~ mɤ].

Front harmonic	Back harmonic	Disharmonic
yhes (one.SG.INE)	puhast (clean.SG.PART)	*yhɤs, *puhæst
pærny (Pärnu.SG.NOM)	svtku (knead.SG.IMP)	*pærnu, *svtkɤ
ekæ (neither)	kopsu (lung.SG.GEN)	*ekɑ, *kopsy
tekijæ (doer.SG.NOM)	tulinɤ (hot.SG.NOM)	*tekijɑ, *tuline
elæsime (live.1PL.PST)	olimɤ (be.1PL.PST)	*elæsimɤ, *olime

The /e/-/ɤ/ pair has somewhat variable harmony (Léonard, 1993), with /e/ participating inconsistently as both trigger and target. However, these mid unrounded vowels demonstrate expected harmony behaviour rather more often than not, and for the purposes of this paper I will abstract away from their variability and assume 100% categorical harmony participation along with the other vowel pairs. Though this idealization is not completely accurate, my approach for the learning portion of the project (Section 2.3) will be able to accommodate variation and other frequency-based information as those are reincorporated in subsequent work.

The front marked vowels pattern together in KE, as do the back unmarked vowels.<sup>2</sup> Thus throughout the paper I will refer to the following sets:

- (2) **F<sub>m</sub>**: the set of front marked vowels {æ, ø, y}  
 (3) **B<sub>u</sub>**: the set of back unmarked vowels {ɑ, o, u}

Section 2.2 tackles the question of what kind of constraint-based framework is required in order to account for VH in KE. The Balto-Finnic languages and their various dialects, Estonian included, demonstrate varying types and degrees of front/back VH. K&P present a set of constraints with an aim to characterize the diversity of harmony behaviour exhibited by all of these related languages. However, their constraints fail to account for a particular type of disharmony (specifically, transparency) that is characteristic of KE. The constraint set I propose in Section 2.2.1 is adapted from K&P's, but also includes some crucial additions in order to ensure that it is capable of accounting for the VH patterns of KE.

<sup>1</sup> For a more in-depth description of KE VH, including behaviour of diphthongs, see Vesik (under review).

<sup>2</sup> The idea of particular vowels being marked or not is based on their distribution in other Balto-Finnic varieties; see Kiparsky & Pajusalu (2003) for further detail.

Following this, Section 2.3 addresses the issue of what assumptions and biases are necessary in order for a constraint-based learning algorithm to successfully learn a KE grammar. I focus on one bias in particular that was found to be crucial for success; that is, the update rule used in the CEDRA. Section 2.3 also presents results of learning simulations, after which Section 3 presents an in-depth discussion of the conditions under which CEDRA’s update rule is a benefit or a hindrance to learning in a stochastic OT environment. Finally, Section 4 concludes.

## 2.2 Describing the grammar

**2.2.1 Constraints** I draw the basis for my proposed constraints from K&P. They present a constraint set to account the typology of vowel distributions for a broad set of Balto-Finnic languages (see K&P, in particular examples (2), (3), (4), (12), (14)). (4) through (9) below summarize K&P’s constraints that are relevant to the grammar of KE.

Segmental markedness constraints:

- (4)  $*\mathbf{F}_m$  (\*æ, \*ø, \*y): if a vowel is front, it must be nonlow and unrounded.
- (5)  $*\mathbf{v}$ : if a vowel is nonlow and unrounded, it must be front.

KE VH is rooted in agreement of the feature [back]. I assume access to a vowel tier, and build all of the more specific harmony constraints that follow on this foundation:

- (6) AGREE(Back): adjacent segments on the vowel tier must have the same value of the feature [back]. Abbreviated as AGR(Bk).

However, AGR(Bk) on its own is not capable of describing the co-occurrences that surface in KE. For example, /i..a/ is a valid vowel sequence even though it is disharmonic, because /i/ is transparent. A harmony constraint such as the one in (7) permits sequences such as /i..a/ (disharmonic) and /i..æ/ (marked) but disallows those such as /æ..a/ (both marked and disharmonic). It is constructed using constraint conjunction:

- (7) Marked Vowel Harmony for / $\mathbf{F}_m$ /: AGR(Bk) &  $*\mathbf{F}_m$ .  
A word may not contain both a pair of disharmonic syllable-adjacent vowels and a vowel in  $\mathbf{F}_m$ . Abbreviated as VH( $\mathbf{F}_m$ ).

Faithfulness constraints ensure that harmony is progressive (via the privileged status of vowels in initial syllables) while also regulating the influence of markedness constraints on all vowels in general. I include both a position-specific as well as a general identity constraint for the feature [back].

- (8) IDENT- $\sigma_1$ (Back): an [ $\alpha$  back] input segment in an initial syllable must not have a [- $\alpha$  back] output correspondent. Abbreviated as ID- $\sigma_1$ (Bk).
- (9) IDENT(Back): an [ $\alpha$  back] input segment must not have a [- $\alpha$  back] output correspondent. Abbreviated as ID(Bk).

K&P’s proposed constraints and rankings successfully describe (dis)harmony in various Estonian dialects as well as other Balto-Finnic languages. However, their analysis is not able to account for the transparent behaviour of /i/ as in KE without also inadvertently characterizing /e/ as transparent as well. See (10), in which harmony is enforced when involving vowels other than /e/, vs (11), in which /e/’s behaviour is transparent rather than harmonic. This is due to the fact that transparency under their account is a property of unmarked front vowels, a description that matches both /i/ and /e/. In order to address this issue, I propose several additional constraints below.<sup>3</sup>

- (10) Conjoined VH( $\mathbf{F}_m$ ) constraint bans disharmony involving vowel(s) in  $\mathbf{F}_m$ .

/æ..o/	VH( $\mathbf{F}_m$ )	ID- $\sigma_1$ (Bk)	ID(Bk)	$*\mathbf{F}_m$
a. æ..o	*!			*
b. æ..ø			*	**
c. a..o		*!	*	

<sup>3</sup> Interested readers can find a much more detailed argument in Vesik (2022).

- (11) K&P’s constraints cannot account for KE’s alternating /e/ but transparent i.

/u..i..e/	VH( $\mathbf{F}_m$ )	ID- $\sigma_1$ (Bk)	ID(Bk)	* $\gamma$
☹ a. u..i..e				
☹ b. u..i.. $\gamma$			*!	*
c. y..i..e		*!	*	
d. y..i.. $\gamma$	*!	*	**	*

KE requires the ability to penalize /e/ in particular harmony-related contexts. Therefore to facilitate the construction of harmony constraint (13) below, I adopt this third segmental markedness constraint:

- (12) \*e

In order to ensure that disharmony with /e/ is penalized, we need a VH constraint involving /e/. If such a constraint is not included in the set, then /e/ (as an unmarked front vowel) falls into the same patterns as transparent /i/, even though it does have a back counterpart and participates in harmony. I propose this additional marked harmony constraint for /e/:

- (13) Marked Vowel Harmony for /e/: AGR(Bk) & \*e. A word may not contain both a pair of disharmonic syllable-adjacent vowels and the vowel /e/. Abbreviated as VH(e).

The next constraint is not strictly necessary for building a grammar of KE; however, simulations show that successful gradual learning of Standard Estonian requires a counterbalance for \* $\mathbf{F}_m$  (Vesik, 2022); thus I adopt a fourth segmental markedness constraint:

- (14) \*
- $\mathbf{B}_u$
- (\*a, \*o, \*u)

I propose the addition of a third and a fourth conjoined harmony constraint, which penalize disharmony in the domain of / $\gamma$ / and  $\mathbf{B}_u$  respectively:

- (15) Marked Vowel Harmony for / $\gamma$ /: AGR(Bk) & \* $\gamma$ . A word may not contain both a pair of disharmonic syllable-adjacent vowels and the vowel / $\gamma$ /. Abbreviated as VH( $\gamma$ ).
- (16) Marked Vowel Harmony for / $\mathbf{B}_u$ /: AGR(Bk) & \* $\mathbf{B}_u$ . A word may not contain both a pair of disharmonic syllable-adjacent vowels and a vowel in  $\mathbf{B}_u$ . Abbreviated as VH( $\mathbf{B}_u$ ).

**2.2.2 Ranking** To ensure that the constraints proposed are in fact capable of characterizing KE, I use the OTSoft (Hayes et al., 2013) implementation of the Low-Faithfulness Constraint Demotion algorithm (LFCD; Hayes, 2004) to demonstrate that a restrictive grammar for KE can be constructed from these constraints, given appropriate learning data. In order to mimic a human learner, I use only positive evidence (that is, underlying forms are assumed to be identical to surface forms) as learning data for the simulations. Once the learner has completed its batch learning process, I test it on a larger set of ungrammatical forms to assess what it has learned about the grammar of KE. See Table 1 for sample data.

Learning from faithful mappings	Testing on ungrammatical inputs
/ $\mathbf{F}_m$ ..e/ → [ $\mathbf{F}_m$ ..e]	/ $\mathbf{F}_m$ .. $\gamma$ / → [ $\mathbf{F}_m$ ..e]
/i.. $\gamma$ / → [i.. $\gamma$ ]	/ $\gamma$ ..e/ → [ $\gamma$ .. $\gamma$ ]
/ $\mathbf{B}_u$ ..i.. $\mathbf{B}_u$ / → [ $\mathbf{B}_u$ ..i.. $\mathbf{B}_u$ ]	/ $\mathbf{B}_u$ ..i.. $\mathbf{F}_m$ / → [ $\mathbf{B}_u$ ..i.. $\mathbf{B}_u$ ]

**Table 1:** Sample data for learning and testing a KE grammar.

The LFCD does in fact succeed in producing a correct, restrictive grammar for KE. It installs markedness constraints in higher strata than faithfulness constraints, wherever possible. This inherent bias, which Hayes (2004) refers to as *Favour Activeness*, implements a phonotactic learner in that rather than interpreting the learning data as being entirely faithful, the algorithm prioritizes markedness constraints that are satisfied by the learning data. The LFCD first installs markedness constraints that are never violated; that is, the harmony constraints for  $\mathbf{F}_m$  and /e/. Following this,

the faithfulness constraints that determine how to resolve harmony errors are installed, with specific faithfulness prioritized over general faithfulness (Hayes's (2004) *Favour Specificity*). At this point, all winning candidates have been identified and the remaining markedness constraints are installed in the lowest stratum.


(17) LFCD-generated ranking for KE:

$$\text{VH}(\mathbf{F}_m); \text{VH}(e) \gg \text{ID-}\sigma_1(\text{Bk}) \gg \text{ID}(\text{Bk}) \gg *F_m; *B_u; *\gamma; *e; \text{AGR}(\text{Bk}); \text{VH}(\mathbf{B}_u); \text{VH}(\gamma)$$


The tableaux in (18) and (19) show how this ranking enforces harmony between participating vowels, while preserving the value of [back] for the vowel in the first syllable as well as treating /i/ transparently.<sup>4</sup>

In (18) we see that  $\text{ID}(\text{Bk})$  must be ranked below all of the top three constraints. (18a) shows that  $\text{VH}(e)$  must outrank  $\text{ID}(\text{Bk})$  in order to prevent the fully faithful candidate [e.. $\gamma$ ] (which fails to harmonize for /e/) from being preferred to one that violates faithfulness but obeys harmony for /e/; a parallel argument can be made for  $\text{VH}(\mathbf{F}_m)$  outranking  $\text{ID}(\text{Bk})$ . (18b) shows that  $\text{ID-}\sigma_1(\text{Bk})$  must outrank  $\text{ID}(\text{Bk})$  in order to prevent the candidate [ $\mathbf{B}_u$ .. $\mathbf{B}_u$ .. $\mathbf{B}_u$ ], which achieves harmony by changing only the first vowel's value of [back], from being selected as the optimal candidate over a candidate such as [ $\mathbf{F}_m$ .. $\mathbf{F}_m$ .. $\mathbf{F}_m$ ], which achieves harmony by violating  $\text{ID}(\text{Bk})$  twice in non-initial vowels. In other words,  $\text{ID-}\sigma_1(\text{Bk}) \gg \text{ID}(\text{Bk})$  is necessary in order to avoid Majority Rule harmony (Lombardi, 1999; Baković, 2000).

(18) a.  $\text{VH}(e) \gg \text{ID}(\text{Bk})$


/e.. $\gamma$ /	$\text{VH}(\mathbf{F}_m)$	$\text{VH}(e)$	$\text{ID-}\sigma_1(\text{Bk})$	$\text{ID}(\text{Bk})$
a. e.. $\gamma$		*!		
 b. e..e				*
c. $\gamma$ .. $\gamma$			*!	*
d. $\gamma$ ..e		*!	*	**

b.  $\text{ID-}\sigma_1(\text{Bk}) \gg \text{ID}(\text{Bk})$

/ $\mathbf{F}_m$ .. $\mathbf{B}_u$ .. $\mathbf{B}_u$ /	$\text{VH}(\mathbf{F}_m)$	$\text{VH}(e)$	$\text{ID-}\sigma_1(\text{Bk})$	$\text{ID}(\text{Bk})$
a. $\mathbf{F}_m$ .. $\mathbf{B}_u$ .. $\mathbf{B}_u$	*!			
b. $\mathbf{F}_m$ .. $\mathbf{F}_m$ .. $\mathbf{B}_u$	*!			*
 c. $\mathbf{F}_m$ .. $\mathbf{F}_m$ .. $\mathbf{F}_m$				**
d. $\mathbf{B}_u$ .. $\mathbf{B}_u$ .. $\mathbf{B}_u$			*!	*

In (19) we see that these same rankings hold for inputs with a medial (or later) /i/ as well, because the VH constraints are defined over the entire word rather than being restricted to immediate neighbours. /i/ does not participate in harmony but is transparent. Subsequent non-/i/ vowels harmonize to the first, even if they are [+back] and therefore disagree with /i/.

(19)

/ $\gamma$ ..i..e/	$\text{VH}(\mathbf{F}_m)$	$\text{VH}(e)$	$\text{ID-}\sigma_1(\text{Bk})$	$\text{ID}(\text{Bk})$
a. $\gamma$ ..i..e		*!		
 b. $\gamma$ ..i.. $\gamma$				*
c. e..i..e			*!	*
d. e..i.. $\gamma$		*!	*	**

Given the subset of constraints included in the tableaux of (18) and (19), it appears that the last candidate in each of (18a) and (19) might be harmonically bounded by the penultimate. However, the extended view of tableau (18a) shown in (20) demonstrates that this is not the case. Broadening our view to consider all constraints at once confirms that the segmental markedness constraints in the lowest stratum (in particular  $*\gamma$  and  $*e$ ) have a role to play in avoiding harmonic bounding.

<sup>4</sup> However, I set aside inputs such as /i..e..u/, whose harmony is driven by a non-initial vowel.

(20)

$/e..ɣ/$	VH( $\mathbf{F}_m$ )	VH(e)	ID- $\sigma_1$ (Bk)	ID(Bk)	* $\mathbf{F}_m$	* $\mathbf{B}_u$	* $\gamma$	*e	AGR(Bk)	VH( $\mathbf{B}_u$ )	VH( $\gamma$ )
a. e..ɣ		*!					*	*	*		*
b. e..e				*				**			
c. ɣ..ɣ			*!	*			**				
d. ɣ..e		*!	*	**			*	*	*		*

As mentioned in Section 2.1, the  $/e/-/\gamma/$  pair exhibits variable participation in KE harmony. Although this project assumes categorical behaviour on the part of this pair, it is crucial to note that this variability does exist, as this is what informs the work in Section 2.3. A batch learner such as LFCD is capable of producing only discrete, ordered rankings and therefore of capturing only categorical patterns. Since it is my eventual goal to investigate learning simulations that better parallel reality in that they are both gradual as well as able to incorporate and reflect the variation in KE, this approach will not be suitable for the broader problem. Instead, I will take a step past the success of LFCD and consider an algorithm that both learns from and is able to replicate noisy data; that is, the GLA (Boersma & Hayes, 2001). This is the algorithm that will be used to model acquisition of variable patterns in future work.

**2.3 Learning simulations** This research was carried out via learning simulations focusing on acquisition of KE grammar. In order to simulate a human learner acquiring KE, I chose to use a phonotactic learning model, which is presented with only positive learning evidence (that is, it assumes that the underlying form is identical to the surface form). The LFCD is a batch learner and functionally assumes that absence of evidence is evidence of absence. This produces ranking (17) presented in Section 2.2.2.

As mentioned above, the GLA has the ability to produce a variable grammar based on input frequencies. It is an online learner, processing one piece of data at a time. The GLA assesses an input based on current constraint ordering, and adjusts values based on errors. Since the GLA works gradually, without seeing the entire dataset at once, it is not able to make the same conclusion that LFCD does and assume that lack of evidence corresponds to negative evidence. This results in the inability of unviolated constraints to rise, which I discuss further in Section 3.

**2.3.1 Data** The learning data for these simulations are based on the KE subset of the Estonian Dialect Corpus (Lindström, 2013). Appendix A and especially Vesik (under review) contain more detailed information about the raw data, but here I will focus on how the data were adapted and presented as learning inputs to the GLA. Since I am idealizing the vowel patterns in both dialects to be categorical, the small amounts of variation attested in the corpus will be ignored by means of omitting the disharmonic KE forms when determining relative input frequencies of data for the learner.

Initial bigrams and trigrams of vowels were extracted from the KE subset and the frequency of each type calculated. I then determined frequencies (relative to 1000 inputs) of each grammatical n-gram using the pair of equations in (21), for KE bigrams and trigrams.

(21)

$$x'_i = 1000 \left( \frac{x_i}{n} \right)$$

$$X_i = \begin{cases} 1 & x'_i < 1 \\ \lfloor x'_i \rfloor & x'_i \geq 1 \end{cases}$$

where  $x_i$  = number of instances of the  $i^{\text{th}}$  ngram,  
 $n$  = total number of ngrams,  
 $x'_i$  = relative frequency of ngram  $i$  proportional to a total of 1000, and  
 $X_i$  = number of times to use form  $i$  as an input to the GLA.

This ensured that even grammatical forms with very low representation in the corpus could inform the learning process (i.e., by ensuring that small numbers were not rounded down to 0 out of 1000).

**2.3.2 Parameters and biases** GLA learning simulations were performed using implementations provided by OTSoft (Hayes et al., 2013)<sup>5</sup> and the author (see Appendix B), and were run with the parameter settings and widely-used biases summarized in Tables 2 and 3. These implementations used a ranked, rather than a weighted, version of the GLA. Constraint values move gradually, but at evaluation time are strictly ranked based on their current values (which include evaluation noise).

Parameter	Value
# Batches	4
# Learning trials per batch	$\in \{1000, 5000, 50000, 500000\}$
Evaluation noise for M constraints	2
Evaluation noise for F constraints	2
Plasticity (learning rate) for M constraints	2, 0.2, 0.02, 0.002 over the 4 batches
Plasticity (learning rate) for F constraints	2, 0.2, 0.02, 0.002 over the 4 batches

**Table 2:** GLA learning simulation parameters

Bias	Result if omitted
Low initial faithfulness (Gnanadesikan, 1995)	Fully-faithful grammar
Specific over general faithfulness (Hayes, 2004)	VH present but initial-syllable vowel not preserved

**Table 3:** GLA learning simulation biases

In addition to the biases mentioned in Table 3, it was necessary to introduce a third bias into the GLA for it to be able to learn a correct grammar for KE: the calibrated re-ranking rule as per the CEDRA, introduced by Magri (2012).

In order to acquire a grammar equivalent to (17), the parameters of the learning algorithm must be such that markedness constraints other than  $VH(\mathbf{F}_m)$  and  $VH(e)$  are demoted far enough, and quickly enough, that the faithfulness constraints can surpass those (without also rising above  $VH(\mathbf{F}_m)$  and  $VH(e)$ ) with enough room that noisy evaluation doesn't cause any unintended variation. Without this third bias, the necessary movement cannot occur; it must be ensured that constraints should be demoted more aggressively than they are promoted. Such a bias can be implemented by employing the update rule from Magri's (2012) CEDRA, in which demotions are always applied in full force, but the effect of each individual promotion depends on both the number of demotions and the number of promotions being applied:

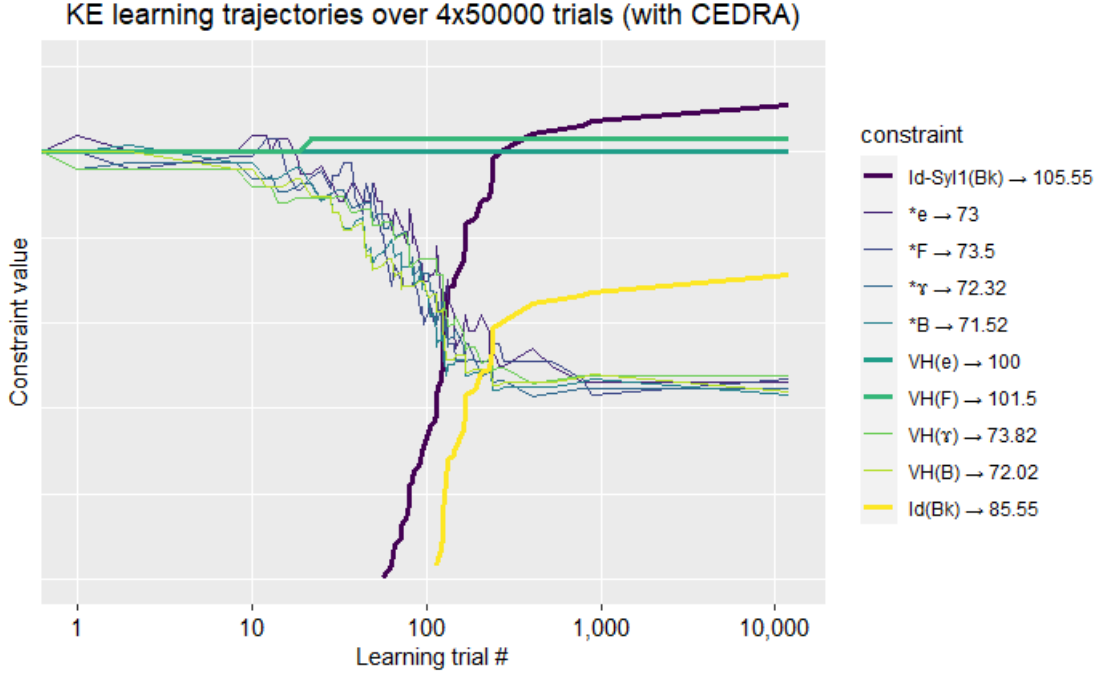
<sup>5</sup> Note that the CEDRA update rule is referred to as "Magri update" in OTSoft.

(22)

$$\text{promotion amount} = \frac{\text{number of constraints demoted}}{1 + \text{number of constraints promoted}} \times \text{plasticity}$$

Since the majority of updates in the KE learning process involve at least as many promotions as demotions, the CEDRA update rule serves to effectively reduce the amount by which constraints are promoted in most learning trials.

**2.3.3 Results** With parameters set as specified in Section 2.3.2, simulations with 1000 or more trials per batch converged to a grammar that correctly describes the VH of KE; see constraint trajectories and final values in Figure 2.



**Figure 2:** Changes in constraint values over 200,000 trials of KE learning data. The horizontal axis is truncated as there was only a single learning error occurring after trial 10,000.

This result aligns with ranking (17) produced by LFCD. The ordering of constraints differing by 10 or greater is almost never<sup>6</sup> affected by evaluation noise; therefore the resulting grammar can be summarized as in ranking (23).

(23) GLA-learned ranking for KE, using CEDRA update rule:  
 $\text{ID-}\sigma_1(\text{Bk}); \text{VH}(\mathbf{F}_m); \text{VH}(e) \gg \text{ID}(\text{Bk}) \gg \text{VH}(r); *F_m; *e; *r; \text{VH}(\mathbf{B}_u); *B_u; \text{AGR}(\text{Bk})$

Simulations with KE inputs fail when CEDRA is not applied. Learning trajectories and results presented in Figure 3 show that the top-ranked markedness constraints are the correct ones but they are nevertheless below  $\text{ID-}\sigma_1(\text{Bk})$  and  $\text{ID}(\text{Bk})$ , resulting in a fully faithful grammar as in (24).

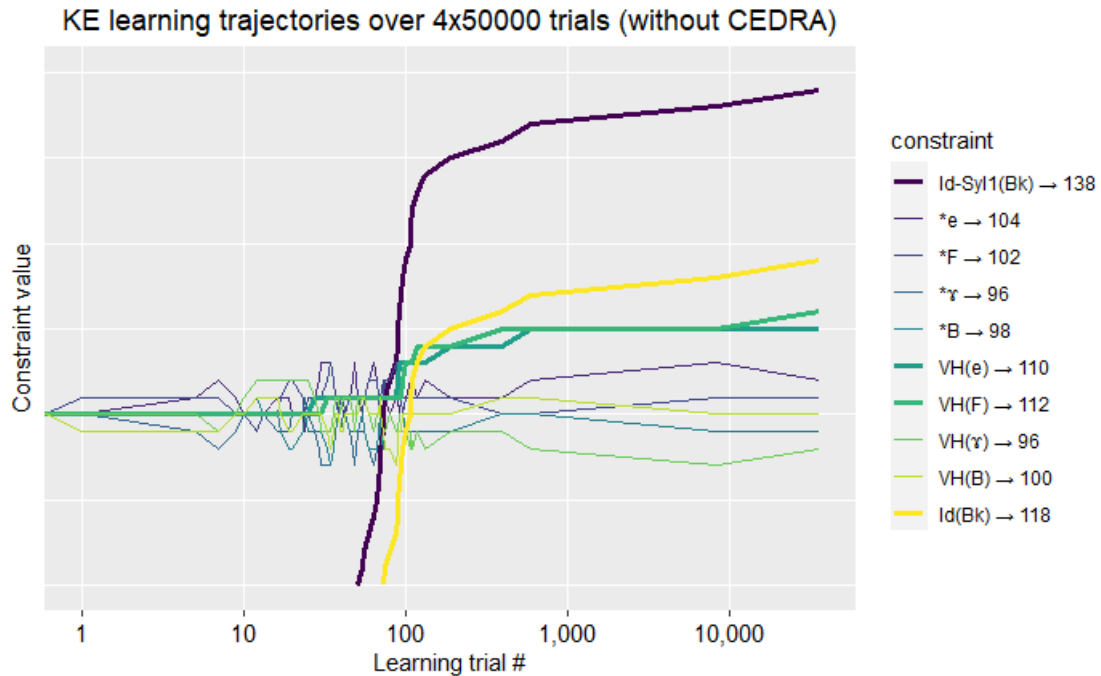
(24) GLA-learned ranking for KE, without CEDRA update rule:  
 $\text{ID-}\sigma_1(\text{Bk}) \gg \text{ID}(\text{Bk}) \gg \text{VH}(\mathbf{F}_m); \text{VH}(e) \gg *e; *F_m; \text{VH}(\mathbf{B}_u); *B_u; \text{VH}(r); *r \gg \text{AGR}(\text{Bk})$

### 3 Discussion

**3.1 Constraint set** K&P frame their work in terms of the typology of VH patterns in many Balto-Finnic languages; however, their typology does not include all related languages. In particular,

<sup>6</sup> Well under 0.05% of the time.





**Figure 3:** Changes in constraint values over 200,000 trials of KE learning data. The horizontal axis is truncated as there was only a single learning error occurring after trial 10,000.

their constraint set is not able to accommodate KE. In light of this, it is clear that some additions and/or adaptations had to be made to their constraint set. In the rankings presented in this paper (LFCD-learned and GLA-learned), the constraints  $VH(\gamma)$  and  $VH(\mathbf{B}_u)$  are installed in the lowest strata, rendered inactive by the higher-ranked constraints. Nor are they referred to in some indirect way like, for example,  $AGR(Bk)$  is via its inclusion in the definition of the conjoined constraints. The justification for inclusion of these constraints does not refer to strategy but rather to symmetry. The need for  $VH(\mathbf{F}_m)$  is clear through the work it does to identify winning candidates.  $*e$  is motivated by the need for  $VH(e)$ , in whose definition it is included. While  $*γ$  and  $*B_u$ , both necessary for different reasons, do not require the existence of their corresponding VH constraints, I have included the harmony constraints for the sake of continuity and symmetry: if  $*F_m$  with  $VH(\mathbf{F}_m)$  and  $*e$  with  $VH(e)$  are members of the set (that is, if CON is permitted to conjoin  $AGR(Bk)$  with some segmental markedness constraints), then  $*B_u$  along with  $VH(\mathbf{B}_u)$  and  $*γ$  along with  $VH(\gamma)$  should be also. With that said, I have run simulations without these two VH constraints, and they succeed (as predicted, given their inactivity in the rankings presented). This is worth noting, as the additional constraints may well add a much greater degree of complexity to this typology than is necessary or justified. However, notwithstanding the potential problems associated with the inclusion of these constraints, it is the case that the constraint set presented provides the opportunity to highlight the fact that oscillation which is detrimental in some learning situations can be beneficial in others.

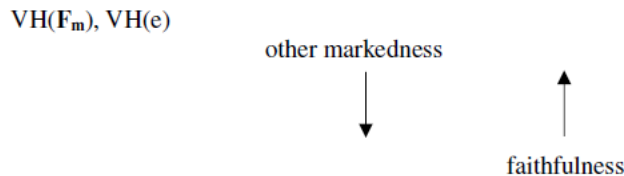
**3.2 Utility of the CEDRA update rule** Due to the nearsightedness of the GLA and the use of positive evidence only, markedness constraints that are never violated by the learning data (eg,  $VH(\mathbf{F}_m)$  and  $VH(e)$  in KE) are highly unlikely to ever be violated by a generated output; the only way this would happen is due to evaluation noise. Therefore, they have negligible opportunity to be promoted as a result of such an error. However, the symmetrical properties of /e/ and /γ/ (due to their segmental markedness as well as their harmony constraints) result in  $*e$  and  $*γ$  staying

relatively stable relative to each other, and also fairly close to their initial value, as errors that promote one demote the other and vice versa; see (25). We run the risk of producing a strictly faithful grammar (which accounts for all of the learning data but no potential unfaithful test data) if the general faithfulness constraint is permitted to rise above the markedness constraints as their values oscillate.

(25) Violation profile for a sample GLA learning error.

$/\gamma..B_u/$	$*\gamma$	$*e$	$*F_m$	VH(e)	VH( $F_m$ )	ID- $\sigma_1$ (Bk)	ID(Bk)
$\checkmark$ a. $\gamma..B_u$	L**	→					
$*\text{?}$ b. $e..F_m$		← W*	← W*			← W*	← W**

Magri’s (2012) CEDRA is motivated by a combination of two goals. The first is to enable a learning algorithm to incorporate promotion as well as demotion, in order to permit faithfulness to start low, for example, and also to allow for adjustments to rankings as new learning inputs are encountered. The second is to avoid full-fledged promotion of constraints in the case of an Elementary Ranking Condition (ERC; Prince, 2002) that contains two or more constraints that prefer the intended winner, in order to avoid overpromoting when it is not clear which of those constraints should be credited with preference of the winner (Credit Problem; Dresher, 1999). The promotion amount is thus shared among the winner-preferring constraints. This effectively creates a learner whose demotions tend to be greater than its promotions, successfully “creating space” for faithfulness between the unviolated markedness constraints, VH( $F_m$ ) and VH(e), and the rest (Figure 4).



**Figure 4:** Creating space for faithfulness constraints between VH( $F_m$ ) and VH(e) vs other markedness constraints.

Magri & Storme (2020) revisit CEDRA, arguing that in fact it does not live up to Magri’s (2012) original claim of solving the GLA’s convergence problem. In the Ilokano metathesis test case that they investigate, two constraints LINEARITY and  $*\text{?}$  must crucially have ranking values that are (a) equal to each other and (b) greater than that of MAXIO(?). However, due to the variation inherent to this grammar, LINEARITY and  $*\text{?}$  end up oscillating during the learning simulation. Ideally, the overall trajectories of these oscillating constraints should be relatively stable over time; other algorithms such as the GLA and the Minimal GLA (Boersma, 1997, 1998) are able to produce such behaviour through valuing promotions and demotions equally. However, because of the CEDRA’s calibrated promotion calculation, this pair of constraints oscillate continually downward instead of around an equilibrium value, and the learner is not able to converge on a grammar.

This type of downward oscillation is precisely what is so useful about the CEDRA update rule in the learning of KE VH, whereas it is the downfall of the CEDRA’s attempt to learn Ilokano metathesis. The two cases are underlyingly different, however, in that the oscillation in KE springs from the violation profiles (that is, it is inherent to the constraints themselves) as in (25), whereas the oscillation in Ilokano is caused by conflicting ERCs (due to variation in inputs) as in Figure 5. In the first case the downward oscillation helps to move antagonistic constraints “out of the way” of active ones, whereas in the second we wish that that learning algorithm would recognize the oscillation as being indicative of variation and stop updating the conflicting pair.

## 4 Conclusion

The constraint set proposed by K&P to account for the VH patterns of Balto-Finnic languages is insufficient to describe all such languages, in particular KE. Adaptations to this set were necessary

$(/x/, [y], \{z\})$	ONSET	MAXIO(V)	*LOWGLIDE	IDENTIO[low]	*[ʔC]	MAXOO(?)	DEPIO(?)	IDENTIO[syll]	LINEARITY	*[ʔ]	MAXIO(?)
ERC 1 = $(/paʔlak/, [pa.lak], \{paʔak\})$									W		L
ERC 2 = $(/paʔlak/, [pa.lak], \{paʔ.lak\})$										W	L
ERC 3 = $(/taʔo-en/, [taw.ʔen], \{taʔ.wen\})$									W		L
ERC 4 = $(/taʔo-en/, [taʔ.wen], \{taw.ʔen\})$									L		W

**Figure 5:** Magri & Storme’s (2020) Table 7 demonstrates conflicting ERCs (3 and 4) after some simplification.

in order to be able to produce a correct grammar for KE. Once these adaptations were made, the constraint set presented in this paper was shown via LFCD to successfully produce a ranking that corresponds to a correct and restrictive grammar for the VH of KE. An existing corpus (the Estonian Dialect Corpus) was processed and analyzed to inform learning data and distributions for a gradual learner. In using the GLA (implemented by Hayes et al. (2013) as well as the author) to learn the vowel phenomena for either dialects, the CEDRA update rule must be used even though Magri & Storme (2020) argue against it in later work. This finding shows that the oscillation in constraint values caused by contradictory learning updates springing from (a) ERCs (due to variation in learning noise) vs (b) violation profiles (due to antagonistic constraints) can be either detrimental or beneficial to converging on a correct grammar, respectively.

## Appendices

### A Corpus

The Estonian Dialect Corpus (Lindström, 2013) was used as the data source for the learning portion of this project. The corpus comprises a total of over 1.2 million words transcribed from spontaneous speech of native Estonian speakers recorded between 1938 and 1996. Speaker dialects are identified via dialect group and parish. The KE subset of the corpus was extracted by restricting entries to those from the Islands dialect group, Kihnu parish. These entries were analyzed for word type counts, identification of monophthongs, and adherence to expected harmony patterns; see Table 4.

Total entries (word tokens)	21,599
Total entries (word types)	5052
Types with $\geq 2$ monophthongs	3375
...of which $\geq 3$	1200
Word-initial monophthong bigrams	1973
Word-initial monophthong trigrams	1173
<i>Harmonic</i> word-initial monophthong bigrams	1935
<i>Harmonic</i> word-initial monophthong trigrams	1052

**Table 4:** KE corpus analysis.

Vowels in sets  $\{\text{æ}, \text{ø}, \text{y}\}$  and  $\{\text{a}, \text{o}, \text{u}\}$  were collapsed and replaced with their representative symbols  $\mathbf{F}_m$  (front marked vowels) and  $\mathbf{B}_u$  (back unmarked vowels). Then, in order to inform

relative frequencies of learning data as described in Section 2.3, the number of instances of each word-initial vowel bigram or trigram was counted.

## B Python implementation of GLA

The implementation of the GLA provided by OTSoft (Hayes et al., 2013) offers the user several ways of modifying the values or processes used by the algorithm. Of particular relevance to this project are the parameters (Table 2) and biases (low faithfulness, specific over general faithfulness, and CEDRA update rule) described in Section 2.3.2. Most of these options can be selected and varied independently of but in concert with the others. However, when both *a priori* rankings (for biasing specific over general faithfulness) and CEDRA update rule are selected, the OTSoft ranking history file shows that only the *a priori* ranking is maintained; the CEDRA update rule is not simultaneously employed. In order to address this gap, I wrote a Python (Van Rossum & Drake, 2009) script<sup>7</sup> to run the GLA, ensuring that the specifics of the implementation would facilitate the simultaneous application of all three necessary biases.

## References

- Asu, Eva Liina & Pire Teras (2009). Estonian. *Journal of the International Phonetic Association* 39:3, p. 367–372.
- Asu, Eva Liina, Pärtel Lippus, Ellen Niit & Helen Türk (2012). The acoustic characteristics of monophthongs and diphthongs in the Kihnu variety of Estonian. *Linguistica Uralica* 48:3, 161–170.
- Baković, Eric (2000). *Harmony, dominance, and control*. Ph.D. thesis, Rutgers University, New Brunswick, URL <http://roa.rutgers.edu>. ROA-360.
- Boersma, Paul (1997). How we learn variation, optionality and probability. van Son, Rob (ed.), *Proceedings of the Institute of Phonetic Sciences (IFA) 21*, University of Amsterdam, Institute of Phonetic Sciences, 43–58.
- Boersma, Paul (1998). *Functional phonology*. Ph.D. thesis, University of Amsterdam.
- Boersma, Paul & Bruce Hayes (2001). Empirical tests of the gradual learning algorithm. *Linguistic inquiry* 32:1, 45–86.
- Dresher, Bezael E. (1999). Charting the learning path: Cues to parameter setting. *Linguistic inquiry* 30:1, 27–67.
- Gnanadesikan, Amalia E. (1995). Markedness and faithfulness constraints in child phonology. URL <http://roa.rutgers.edu>. ROA-67.
- Hayes, Bruce (2004). *Phonological acquisition in Optimality Theory: the early stages*, Constraints in Phonological Acquisition, Cambridge University Press, 158–203.
- Hayes, Bruce, Bruce Tesar & Kie Zuraw (2013). OTSoft [software package]. URL <http://www.linguistics.ucla.edu/people/hayes/otsoft/>. Version 2.6 beta.
- Kiparsky, Paul & Karl Pajusalu (2003). Towards a typology of disharmony. *Linguistic Review* 20, 217–241.
- Lindström, Liina (2013). Eesti murdekorpust [Estonian dialect corpus].
- Lombardi, Linda (1999). Positional faithfulness and voicing assimilation in optimality theory. *Natural Language & Linguistic Theory* 17:2, 267–302.
- Léonard, Jean Léo (1993). Analysis of variation within a typological parameter: Kalevi Wiik’s approach to vowel harmony in southern and eastern Balto-Finnic. *Proceedings of the VIIIth International Conference on Dialectology*.
- Magri, Giorgio (2012). Convergence of error-driven ranking algorithms. *Phonology* 29:2, 213–269.
- Magri, Giorgio & Benjamin Storme (2020). Calibration of constraint promotion does not help with learning variation in stochastic optimality theory. *Linguistic Inquiry* 51:1, 97–123, URL [https://doi.org/10.1162/ling\\_a\\_00328](https://doi.org/10.1162/ling_a_00328).
- Prince, Alan (2002). Entailed ranking arguments. URL <http://roa.rutgers.edu>. ROA-500.
- Sang, Joel (2009). Ühest fonotaktilisest kollisioonist (Kihnu näitel) [On a phonotactic collision (on the example of Kihnu subdialect)]. *Keel Ja Kirjandus* LII:11, 809–817.
- Van Rossum, Guido & Fred L. Drake (2009). Python [programming language]. URL <http://www.python.org/>. Version 3.8.
- Vesik, Kaili (2022). Laying a foundation for bidialectalism: necessary biases for algorithmic learning of two dialects of Estonian. Unpublished manuscript, University of British Columbia.
- Vesik, Kaili (under review). Vowel harmony in the Kihnu variety of Estonian. *Linguistica Uralica* .

<sup>7</sup> Available at <https://github.com/kvesik/learning-Estonian-dialects>