# Laying a foundation for bidialectalism: necessary biases for algorithmic learning of two dialects of Estonian

Kaili Vesik
University of British Columbia
kaili.vesik@ubc.ca

February 24, 2022

**Abstract**

This paper investigates how the vowel patterns of two closely-related dialects of Estonian can be described using as much shared representation as possible, as well as what parameters and biases are necessary for the Gradual Learning Algorithm (Boersma & Hayes, 2001) to be able to learn restrictive grammars for both dialects. Standard Estonian and the minority Kihnu dialect share the same vowel inventory but differ in their distribution of those vowels, with Standard Estonian being subject to positional restrictions and Kihnu Estonian demonstrating front-back vowel harmony. I extend the constraint set that Kiparsky and Pajusalu (2003) propose to account for the vowel harmony typology in Balto-Finnic languages, and show via Low-Faithfulness Constraint Demotion (Hayes, 2004) that there exist ranking of these constraints that account for the patterns in both dialects. I also process the Estonian Dialect Corpus (Lindström, 2013) and use the contents as learning data for runs of the Gradual Learning Algorithm implemented by both OTSoft (Hayes et al., 2013) and the author. Convergence on grammars for both dialects is contingent on three biases being employed during the learning process: low initial faithfulness, specific over general faithfulness, and the Magri (2012) update rule. This suggests that even for relatively simple phonological patterns, the learning environment must be delicately balanced in order to account for the behaviour of two different grammars grounded in the same shared framework.

***Keywords***— phonology, learning algorithms, acquisition, Estonian, vowel harmony

# 1 Introduction

Common sounds, structures and patterns across languages occupy shared representational networks and cognitive processes in bilingual speakers (Simonet, 2016). Linguists, too, use shared abstractions such as phonetic symbols, constraints, and processes, across the world's languages. I endeavour to shed light on how we can formally describe closely-related dialects using as much shared representation as possible. And, given that shared representation, I investigate what we can discover about modeling the paths to and pitfalls of learning these dialects.

This paper focuses on two such dialects of Estonian. One of these, Kihnu Estonian (spoken by about 1300 people, primarily on the island of Kihnu) is virtually unstudied and its vowel harmony phenomena challenge current theories of Balto-Finnic phonology. Following Kiparsky and Pajusalu (2003), I investigate the Optimality-Theoretic characteristics that are required to describe the vowel patterns of Standard Estonian and Kihnu Estonian. I propose a shared constraint set and demonstrate that it is sufficient to construct grammars correctly describing both dialects. Following this, I analyze the parameters and biases necessary for a gradual learning algorithm to acquire restrictive grammars - those that produce all and only correct forms - for both dialects using the common constraint set. I show that even for simple patterns such as positional restrictions or straightforward vowel harmony, several biases need to be built in to the learning process for an algorithmic learner to successfully acquire either of these dialects.

# 2 Estonian

## 2.1 Data

The Estonian Dialect Corpus (EDC; Lindström, 2013) was used as the primary source of data for many portions of this project. Over 20,000 tokens for each of Standard Estonian and Kihnu Estonian were extracted from the corpus (whose entries are tagged with dialect group and parish). The relevant entries were analyzed for word type counts, identification of monophthongs, and adherence to the patterns described below (Section 2.2) for each dialect. This nontrivial work is discussed in detail in Appendix A and the corpus is referred to in examples throughout the paper.

## 2.2 Dialects and vowel distributions

Estonian is in the Balto-Finnic branch of the Uralic language family. This paper focuses on the description and acquisition of two dialects of Estonian: Standard Estonian (SE) and Kihnu Estonian (KE). SE and KE share the same vowel inventory /i, e, ɑ, o, u, ɤ, æ, ø, y/ but differ in their distribution of these vowels (Asu et al., 2012; Asu & Teras, 2009; Léonard, 1993). Figure 1 illustrates the Estonian vowels, with primary vowels in boldface. The highlighted vowels are discussed below.

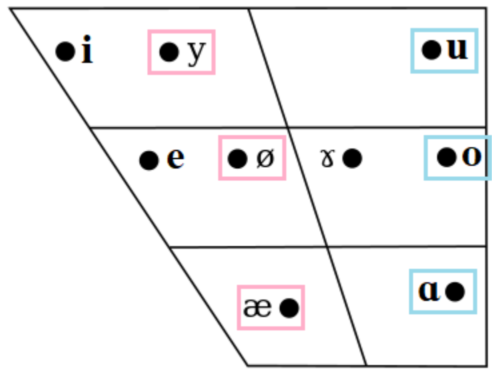


**Figure 1:** Estonian vowels (Asu & Teras, 2009). Primary vowels are in boldface. Set F is highlighted in pink; set B is highlighted in blue.

SE restricts vowels in non-initial syllables of non-compound native words to the primary vowels /i, e, ɑ, o, u/, with /o/ available only in proper names (Asu & Teras, 2009). KE, spoken primarily on the island of Kihnu, is not subject to these distributional restrictions but has progressive front/back vowel harmony involving /y/-/u/, /ø/-/o/, and /æ/-/ɑ/ pairs, with /i/ being neutral and transparent (Léonard, 1993; Sang, 2009). The /e/-/ɤ/ pair has somewhat variable harmony (Léonard, 1993), with /e/ participating inconsistently as both trigger (propagates harmony in 85% of potential cases) and target (undergoes harmony in 78% of potential cases). These mid unrounded vowels demonstrate expected harmony behaviour more often than not, and it is my eventual goal to model learning of this variable data. However, for the purposes of establishing a starting point for this analysis, 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 4) will be able to accommodate variation and other frequency-based information as those are reincorporated in subsequent work. I will also focus only on the distribution of monophthongs. Figure 2 shows examples of how the vowel distributions for the two dialects intersect and Table 1 enumerates their vowel co-occurrences in more detail, again highlighting where the two dialects intersect.

In both dialects, the front secondary (marked) vowels pattern together, as do the back primary (unmarked) vowels. Thus throughout the paper I will refer to the following sets:

(1) **F**: the set of front marked vowels {æ, ø, y}

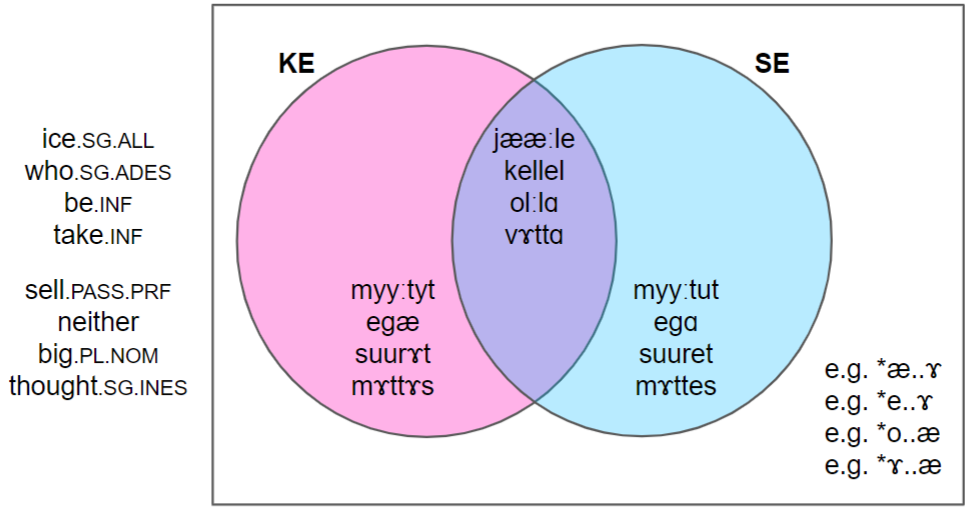


**Figure 2:** Grammatical and ungrammatical vowel sequences in KE and SE. Examples from EDC and the author, glossed at left.

(2) **B**: the set of back unmarked vowels {ɑ, o, u}

Under these definitions, the privileged (unmarked) vowels in SE are those in the set {i, e} ∪ **B** whereas the marked vowels - those that only surface in initial syllables - are those in the set {ɤ} ∪ **F**. From a harmony perspective as for KE, each front-back harmony pair has one member in set {e} ∪ **F** and the other in set {ɤ} ∪ **B**.

Section 3 tackles the question of what kind of OT framework is required in order to account for the vowel distributions in KE and SE. The Balto-Finnic languages and their various dialects, Estonian included, demonstrate varying types and degrees of front/back vowel harmony. Kiparsky and Pajusalu (2003, hereafter 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 3.1 is adapted from K&P's, but also includes some crucial additions in order to ensure that it is capable of accounting for KE's vowel harmony patterns (see Section 6 for in-depth discussion about adapting K&P's constraint set for KE).

Following this, Section 4 addresses the issue of what parameters, biases, or assumptions are necessary in order for a constraint-based learning algorithm to successfully learn both of these grammars. I focus on three biases in particular that were found to be crucial for success. Section 4 also presents results of learning simulations, after which Section 5

| | | V2 | | | | |
|---|---|---|---|---|---|---|
| | | F | e | i | B | ɤ |
| V1 | F | SE: *<br>KE: [kæry]<br>cart.SG.NOM | [yle]<br>over | [æri]<br>shop.SG.NOM | SE: [kæru]<br>cart.SG.NOM<br>KE: * | *{æ,ø,y}..ɤ |
| | e | SE: *<br>KE: [elæ]<br>live.SG.IMP | [hele]<br>light.SG.NOM | [vesi]<br>water.SG.NOM | SE: [elɑ]<br>live.SG.IMP<br>KE: * | *e..ɤ |
| | i | SE: *<br>KE: [kivigæ][2]<br>rock.SG.COM | [sile]<br>smooth.SG.NOM | [igi]<br>sweat.SG.NOM | [ilu]<br>beauty.SG.NOM | SE: *<br>KE: [pisiksɤ][2]<br>rock.SG.GEN |
| | B | *{ɑ,o,u}..{æ,ø,y} | SE: [pɑne]<br>put.SG.IMP<br>KE: * | [hɑni]<br>goose.SG.NOM | [tɑlu]<br>farm.SG.NOM | SE: *<br>KE: [pɑnɤ]<br>put.SG.IMP |
| | ɤ | *ɤ..{æ,ø,y} | SE: [hɤpe]<br>silver.SG.NOM<br>KE: * | [ɤli]<br>oil.SG.NOM | [ɤlɑ]<br>shoulder.SG.GEN | SE: *<br>KE: [hɤpɤ]<br>silver.SG.NOM |

**Table 1:** SE and KE vowel co-occurrences. Examples selected for brevity from EDC and the author. Pink highlighted cells are grammatical in KE only; blue in SE only; purple in both.

presents an in-depth discussion of the decisions regarding which biases and constraints to use in learning the grammars described in Sections 3 and 4.

Finally, in Section 7 I propose several directions in which these questions can be carried further in order to investigate challenges underlying the analysis of progressive harmony in KE, variation in vowel harmony behaviour, and second-dialect learning.

# 3 Describing the grammars

## 3.1 Constraints

The two dialects described above are similar enough as to be mutually intelligible. It should be conceivable for a learner to acquire both grammars with as much shared rep-

[2]These types of entries are grammatical in KE, but extremely rare. Such forms are ungrammatical in SE and though their corresponding front/back pairs do appear after [i] in SE, the very fact that they follow [i] means that there is no reason for their value of [back] to change in order to satisfy KE's harmony patterns.

resentation as possible, so I draw the basis for my proposed constraint set from K&P. They present constraints for a broad set of related Balto-Finnic languages in an attempt to characterize the typology of vowel distributions in this set; I will focus on the constraints, adapted from K&P, that are necessary to capture both KE and SE.

SE, due to its positional restrictions, requires segmental markedness constraints in order to prevent /ɤ, æ, ø, y/ from surfacing in non-initial syllables. The first vowel is [+back] while the latter three are [-back] and therefore pattern differently in KE. Thus we begin with the constraints presented in K&P example (2):

(3) ***F** (*æ, *ø, *y): if a vowel is front, it must be nonlow and unrounded.

(4) *ɤ: if a vowel is nonlow and unrounded, it must be front.

Although it is not necessary for SE, KE also requires the ability to penalize /e/ in particular harmony-related contexts. Therefore to facilitate the construction of harmony constraint (11) below, I adopt this third segmental markedness constraint:

(5) *e

The next constraint is not necessary for building a grammar of either KE or SE; however, simulations (see Section 4.3.2) show that successful gradual learning of SE requires a counterbalance for ***F**; thus I adopt a fourth segmental markedness constraint:

(6) ***B** (*ɑ, *o, *u)

KE vowel harmony 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 the foundation of K&P's (example (3)):

(7) 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..ɑ/ is a valid vowel sequence even though it is disharmonic, because /i/ is transparent. Here I introduce the Marked Harmony constraints, which permit sequences such as /i..ɑ/ (disharmonic) and /i..æ/ (marked) but disallow those such as /æ..ɑ/ (both marked and disharmonic). They are constructed using constraint conjunction.

K&P define, with a combination of their examples (12) and (14)a, the first of these:

(8) Generalized Marked Harmony for /**F**/: AGR(Bk) & ***F**.
A word may not contain both a vowel marked for /**F**/ and a pair of syllable-adjacent vowels disharmonic for /**F**/. Abbreviated as GMH(**F**).

I propose the addition of a second and a third conjoined harmony constraint, which penalize disharmony in the domain of /ɤ/ and **B** respectively:

(9) Generalized Marked Harmony for /ɤ/: AGR(Bk) & *ɤ. A word may not contain both a vowel marked for /ɤ/ and a pair of syllable-adjacent vowels disharmonic for /ɤ/. Abbreviated as GMH(ɤ).

(10) Generalized Marked Harmony for **/B/**: AGR(Bk) & ***B**. A word may not contain both a vowel marked for **/B/** and a pair of syllable-adjacent vowels disharmonic for **/B/**. Abbreviated as GMH(**B**).

In order to ensure that disharmony with /e/ is also penalized, we need a similar conjoined 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 final marked harmony constraint for /e/, which also justifies the inclusion of the earlier segmental markedness constraint *e.

(11) Generalized Marked Harmony for /e/: AGR(Bk) & *e. A word may not contain both a vowel marked for /e/ and a pair of syllable-adjacent vowels disharmonic for /e/. Abbreviated as GMH(e).

Finally, in order to ensure the privileged status of vowels in initial syllables, and to regulate the influence of the markedness constraints on all vowels in general, I will also include both a specific and a general positional faithfulness constraint for the feature [back]. The first of these is from K&P example (4)a; the second is a more general version of K&P's example (4)c.

(12) 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).

(13) IDENT(Back): an [$\alpha$ back] input segment must not have a [-$\alpha$ back] output correspondent. Abbreviated as ID(Bk).

## 3.2 Rankings

To ensure that the constraints proposed are in fact capable of characterizing the two dialects of interest, I use the OTSoft (Hayes et al., 2013) implementation of the Low-Faithfulness Constraint Demotion algorithm (LFCD; Hayes, 2004) to demonstrate that restrictive grammars for both SE and 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, the surface forms in each dialect) 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 the dialect in question. See Table 2 for sample KE data and Table 3 for sample SE data.

The LFCD does in fact succeed in producing correct, restrictive grammars for both dialects. It installs markedness constraints in higher strata than faithfulness constraints,

| Learning from faithful mappings | Testing on ungrammatical inputs |
|---|---|
| /**F**..e/ → [**F**..e] | /**F**..ɤ/ → [**F**..e] |
| /i..ɤ/ → [i..ɤ] | /ɤ..e/ → [ɤ..ɤ] |
| /**B**..i..**B**/ → [**B**..i..**B**] | /**B**..i..**F**/ → [**B**..i..**B**] |

**Table 2:** Sample data for KE

| Learning from faithful mappings | Testing on ungrammatical inputs |
|---|---|
| /**F**..e/ → [**F**..e] | /**F**..ɤ/ → [**F**..e] |
| /i..ɤ/ → [i..ɤ] | /ɤ..e/ → [ɤ..ɤ] |
| /**B**..i..**B**/ → [**B**..i..**B**] | /**B**..i..**F**/ → [**B**..i..**B**] |

**Table 3:** Sample data for SE

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.

### 3.2.1 KE ranking

For the KE learner, the LFCD first installs markedness constraints that are never violated; that is, the harmony constraints for **F** 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.

(14) LFCD-generated ranking for KE:
GMH(**F**); GMH(e) ≫ ID-$\sigma_1$(Bk) ≫ ID(Bk) ≫ **F**; ***B**; *ɤ; *e; AGR(Bk); GMH(**B**); GMH(ɤ)

The tableaux in (15) and (16) 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.[3]

In (15) we see that ID(Bk) must be ranked below all of the top three constraints. (15a) shows that GMH(**F**) must outrank ID(Bk) in order to prevent the fully faithful candidate [ɤ..**F**] (which fails to harmonize for **F**) from being preferred to one that violates

[3]However, I set aside inputs such as /i..e..u/, whose harmony is driven by a non-initial vowel.

faithfulness but obeys harmony for **F**; (15b) shows a parallel argument for GMH(e) outranking ID(Bk); (15c) shows that ID-$\sigma_1$(Bk) must outrank ID(Bk) in order to prevent the candidate [**B**..**B**..**B**], which achieves harmony by changing only the first vowel's value of [back], from being more harmonic than a candidate such as [**F**..**F**..**F**], which achieves harmony by violating /idbk twice in non-initial vowels. In other words, ID-$\sigma_1$(Bk) $\gg$ ID(Bk) is necessary in order to avoid Majority Rule harmony (Baković, 2000; Lombardi, 1999).

In (16) we see that these same rankings hold for inputs with a medial (or later) /i/ as well, because the GMH 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/.

(15) a. GMH(**F**) $\gg$ ID(Bk)

| /ɤ..**F**/ | GMH(**F**) | GMH(e) | ID-$\sigma_1$(Bk) | ID(Bk) |
|---|---|---|---|---|
| a. ɤ..**F** | ∗! | | | |
| ☞ b. ɤ..**B** | | | | ∗ |
| c. e..**F** | | | ∗! | ∗ |
| d. e..**B** | | ∗! | ∗ | ∗∗ |

b. GMH(e) $\gg$ ID(Bk)

| /e..ɤ/ | GMH(**F**) | GMH(e) | ID-$\sigma_1$(Bk) | ID(Bk) |
|---|---|---|---|---|
| a. e..ɤ | | ∗! | | |
| ☞ b. e..e | | | | ∗ |
| c. ɤ..ɤ | | | ∗! | ∗ |
| d. ɤ..e | | ∗! | ∗ | ∗∗ |

c. ID-$\sigma_1$(Bk) $\gg$ ID(Bk)

| /**F**..**B**..**B**/ | GMH(**F**) | GMH(e) | ID-$\sigma_1$(Bk) | ID(Bk) |
|---|---|---|---|---|
| a. **F**..**B**..**B** | ∗! | | | |
| b. **F**..**F**..**B** | ∗! | | | ∗ |
| ☞ c. **F**..**F**..**F** | | | | ∗∗ |
| d. **B**..**B**..**B** | | | ∗! | ∗ |

(16)

| /ɤ..i..e/ | GMH(**F**) | GMH(e) | ID-$\sigma_1$(Bk) | ID(Bk) |
|---|---|---|---|---|
| a. ɤ..i..e | | ∗! | | |
| ☞ b. ɤ..i..ɤ | | | | ∗ |
| c. e..i..e | | | ∗! | ∗ |
| d. e..i..ɤ | | ∗! | ∗ | ∗∗ |

Given the subset of constraints included in the tableaux of (15) and (16), it appears that the last candidate in each of (15a), (15b), and (16) might be harmonically bounded by the penultimate. However, the extended view of tableau (15b) shown in (17) demonstrates that this is not the case. Broadening our view to consider all constraints at once confirms that the segmental markedness constraints (*ɤ; ***F**; *e) in the lowest stratum have a role to play in avoiding harmonic bounding.

(17)

| /e..ɤ/ | GMH(**F**) | GMH(e) | ID-$\sigma_1$(Bk) | ID(Bk) | ***F** | ***B** | *ɤ | *e | AGR(Bk) | GMH(**B**) | GMH(ɤ) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| a. e..ɤ | | *! | | | | | * | * | * | | * |
| ☞ b. e..e | | | | * | | | | ** | | | |
| c. ɤ..ɤ | | | *! | * | | | ** | | | | |
| d. ɤ..e | | *! | * | ** | | | * | * | * | | * |

### 3.2.2 SE ranking

On the other hand, the SE LFCD learner does not find any unviolated markedness constraints on its first pass. Therefore the top stratum for SE contains the only unviolated faithfulness constraint, which privileges the vowel in the initial syllable. Following this, any errors in non-initial syllables are resolved with the installation of the two segmental markedness constraints *ɤ and ***F**. Once these top two strata are constructed, all winning candidates have been identified and the remaining constraints are installed in the bottom strata.

(18) LFCD-generated ranking for SE:
ID-$\sigma_1$(Bk) $\gg$ ***F**; *ɤ $\gg$ ***B**; *e; AGR(Bk); GMH(**F**); GMH(**B**); GMH(ɤ); GMH(e) $\gg$[4] ID(Bk)

The tableaux in (19a) and (19b) show how this ranking places positional restrictions on the marked vowels (/ɤ/ and those in set **F**), permitting them only in initial syllables and neutralizing to the corresponding unmarked vowel in subsequent syllables. /i/ has no special status here, other than the fact that it happens to not have a [+back] correspondent in the inventory.

Considering input /ɤ..ɤ/ in (19a), we see that ID-$\sigma_1$(Bk) is crucially ranked above *ɤ. This ranking ensures that candidate [e..e], which completely avoids the marked segment /ɤ/ but at the expense of unfaithfulness to the initial vowel, is not preferred over the

[4] It is not in fact necessary for the success of this grammar that ID(Bk) be ranked below the set of constraints above it. This is strictly due to the LFCD's bias of installing markedness constraints before faithfulness.

actual winner [ɤ..e], which preserves a marked first segment in order to satisfy faithfulness to the first vowel.

(19b)'s input /***F**..i/ supports a parallel argument for ID-$\sigma_1$(Bk) $\gg$ ***F**, while also underscoring the fact that /i/ simply patterns with the other unmarked vowels in that it does not face any positional restrictions in SE.

(19) a.

| /ɤ..ɤ/ | ID-$\sigma_1$(Bk) | *ɤ |
|---|---|---|
| a. ɤ..ɤ | | **! |
| ☞ b. ɤ..e | | * |
| c. e..ɤ | *! | * |
| d. e..e | *! | |

b.

| /**F**..i/ | ID-$\sigma_1$(Bk) | ***F** |
|---|---|---|
| ☞ a. **F**..i | | * |
| b. **B**..i | *! | |

As mentioned in Section 2.2, the /e/-/ɤ/ 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 4. 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 Gradual Learning Algorithm (GLA; Boersma & Hayes, 2001). This is the algorithm that will be used to model acquisition of variable patterns in future work.

# 4 Learning

This research was carried out via learning simulations focusing on acquisition of the KE and SE grammars. In order to simulate a human learner acquiring either KE or SE, 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 rankings (14) and (18) presented in Section 3.2, which correctly describe each dialect.

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. My intent is to model the same learner acquiring both dialects; therefore if we are going to use GLA for one (KE) then it should also be used for the other (SE). However, it should be noted that 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 5.1.3.

## 4.1 Data

The learning data for these simulations are based on the KE and SE subsets of the EDC. Appendix A contains a detailed analysis of 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 and SE subsets 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 (20), for each of KE (bigrams and trigrams) and SE (bigrams and trigrams).

(20)

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

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

where, for the relevant dialect,
$x_i$ = 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,
$X_i$ = 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).

## 4.2 Parameters and biases

GLA learning simulations were performed using implementations provided by OTSoft (Hayes et al., 2013) and the author (see Appendix B), and were run with the parameter settings summarized in Table 4. 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 4:** GLA learning simulation parameters

It was necessary to introduce three biases into the GLA for it to be able to learn correct grammars for both KE and SE: (1) specific over general faithfulness, (2) high markedness / low faithfulness, and (3) the Magri update rule.

In order to parallel LFCD's *Favour Specificity* principle, an *a priori* ranking condition was imposed on the constraint set. The ranking of the specific positional faithfulness constraint ID-$\sigma_1$(Bk) over the general ID(Bk) was enforced by requiring that the value of ID-$\sigma_1$(Bk) always be at least 20 greater than the value of ID(Bk). This means that if there is an error in a non-initial syllable that would cause ID(Bk) to be promoted, then ID-$\sigma_1$(Bk) must also be promoted. 20 was selected because it is the default for such *a priori* rankings in OTSoft; it is "very close probabilistically to being an obligatory ranking" (Hayes et al., 2013).

A low-faithfulness bias (*Favour Activeness* in LFCD) in the GLA is implemented via initially high rankings of markedness constraints compared to faithfulness constraints. I assign all markedness constraints an initial value of 100 and faithfulness constraints an initial value of 0. This allows for markedness constraints such as GMH(e) and GMH(**F**) for KE, or *ɤ and ***F** for SE, the opportunity to be active in a grammar that (ideally) converges before the markedness constraints are overtaken by rising faithfulness constraints. This risk is particularly relevant to phonotactic learners because there are no learning data with unfaithful winners; the faithfulness constraints are always satisfied.

In order to acquire a grammar equivalent to (14), the parameters of the learning algorithm must be such that markedness constraints other than GMH(**F**) and GMH(e) are demoted far enough, and quickly enough, that the faithfulness constraints can surpass those without also rising above GMH(**F**) and GMH(e) with enough room that noisy evaluation doesn't cause any unintended variation. Without the third and final 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 Magri (2012) update rule, 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:

(21)

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

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

## 4.3 Results

### 4.3.1 KE simulations

With parameters set as specified in Section 4.2, simulations with 1000 or more trials per batch converged to a grammar that correctly describes the vowel harmony of KE; see Table 5 and Figure 3.

| Constraint | Final value |
|---|---|
| ID-$\sigma_1$(Bk) | 105.6 |
| GMH(**F**) | 101.5 |
| GMH(e) | 100.0 |
| ID(Bk) | 85.6 |
| GMH(ɤ) | 73.8 |
| ***F** | 73.5 |
| *e | 73.0 |
| *ɤ | 72.3 |
| GMH(**B**) | 72.0 |
| ***B** | 71.5 |
| AGR(Bk) | 52.6 |

**Table 5:** Final constraint values after $4 \times 50000$ trials of KE learning data.

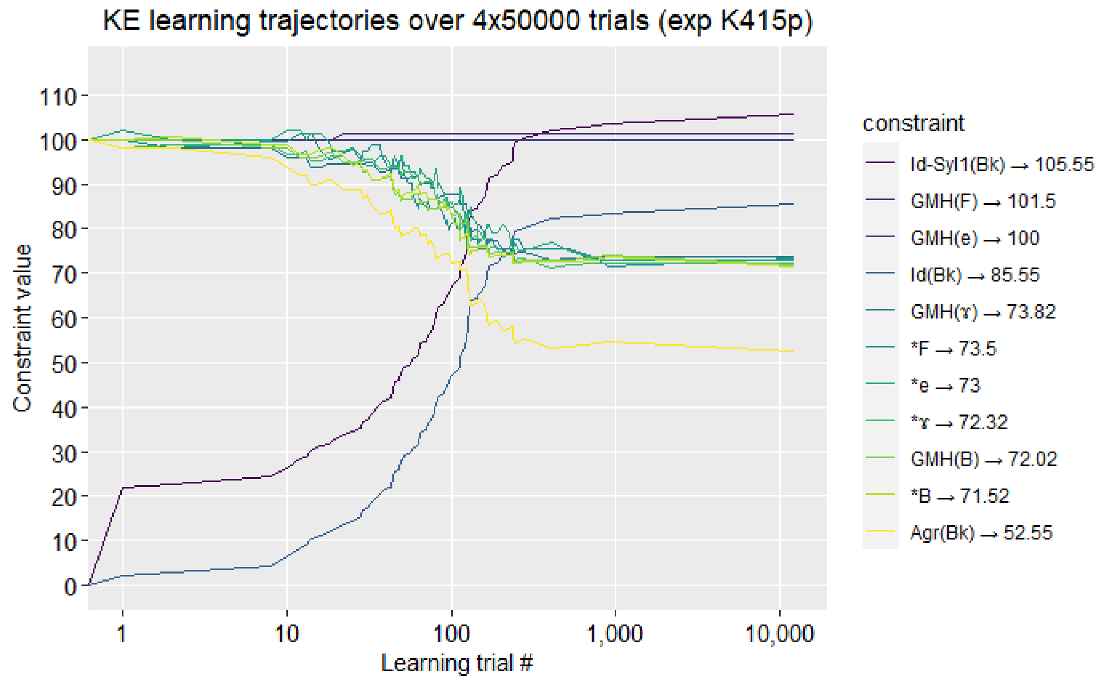


**Figure 3:** Changes in constraint values over 200,000 trials of KE learning data. The horizontal axis is truncated as there were no further learning errors after trial 10,000.

This result aligns with ranking (14) produced by LFCD. The ordering of constraints differing by 10 or greater is almost never[5] affected by evaluation noise; therefore the resulting grammar can be summarized as in ranking (22).

(22) GLA-learned ranking for KE:
ID-$\sigma_1$(Bk); GMH(**F**); GMH(e) $\gg$ ID(Bk) $\gg$ GMH(ɤ); ***F**; *e; *ɤ; GMH(**B**); ***B**; AGR(Bk)

### 4.3.2 SE simulations

With parameters set as specified in Section 4.2, simulations with at least 5000 iterations per batch converged to a grammar that correctly describes the positional restrictions of SE; see Table 6 and Figure 4.

This result does align with ranking (18) produced by LFCD. However, the difference between the values of ***F** and *ɤ vs the value of ID(Bk) is relatively small (only about 6),

[5] Well under 0.05% of the time.

| Constraint | Final value |
|---|---|
| ID-$\sigma_1$(Bk) | 110.6 |
| ***F** | 96.5 |
| *ɤ | 96.5 |
| ID(Bk) | 90.6 |
| *e | 85.7 |
| ***B** | 85.4 |
| GMH(**F**) | 78.9 |
| GMH(ɤ) | 74.1 |
| GMH(**B**) | 64.3 |
| GMH(e) | 47.4 |
| AGR(Bk) | 35.6 |

**Table 6:** Final constraint values after $4 \times 50000$ trials of SE learning data.

which results in evaluation noise effectively promoting ID(Bk) over at least one of ***F** and *ɤ (and therefore producing incorrectly faithful outputs) about 3% of the time. As such, the resulting grammar can be summarized as per (23) below only if the relative ordering of ***F**, *ɤ, and ID(Bk)) is understood to be very slightly variable.

(23) GLA-learned ranking for SE:
ID-$\sigma_1$(Bk) $\gg$ ***F**; *ɤ $\gg$ ID(Bk); *e; ***B**; GMH(**F**); GMH(ɤ); GMH(**B**); GMH(e); AGR(Bk)

# 5 Discussion

## 5.1 Learning biases

The GLA simulations for KE and SE addressed in this paper show that even for quite simple patterns such as the ones described for these two dialects, successfully modelling the acquisition of such phonological patterns (on the basis of as much shared representation as possible) requires the incorporation of several learning biases, of varying levels of complexity. In addition to this, it should also be noted that at least one of the patterns is artificially simple: in reality, KE involves variable rather than categorical harmony of its /e/-/ɤ/ pair and therefore incorporating this noise into the learning process has potential to require an even more nuanced learning environment for successful alignment with the data.

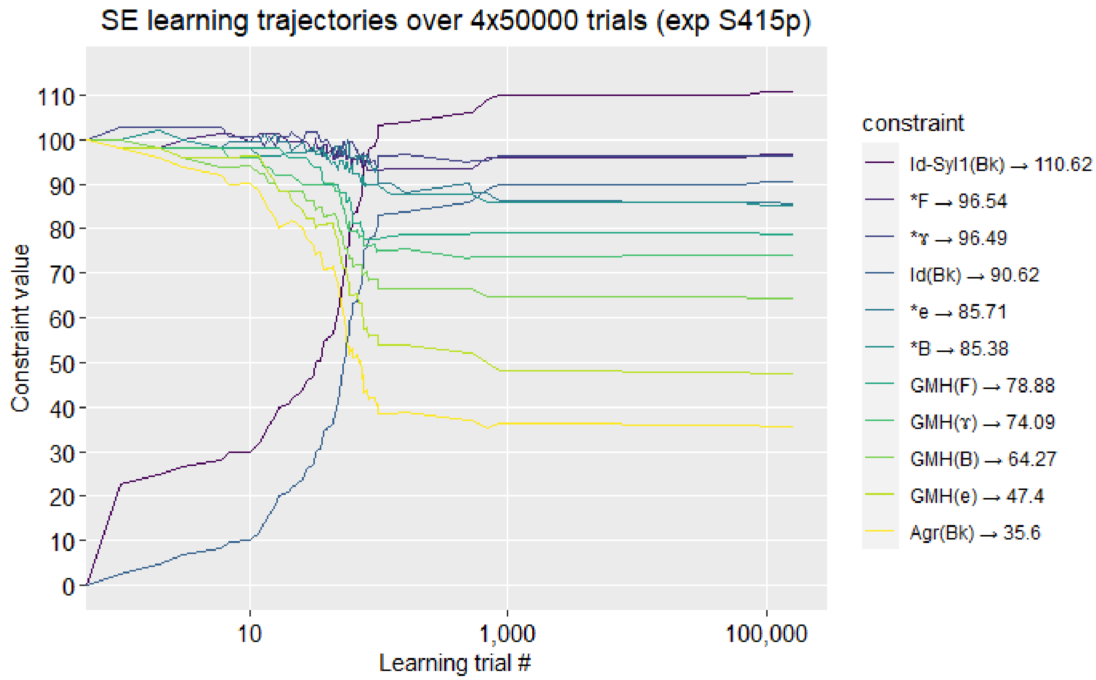


**Figure 4:** Changes in constraint values over 200,000 trials of SE learning data. The horizontal axis is truncated as there were no further learning errors after trial 100,000.

### 5.1.1 Specific over general faithfulness

The rationale for favouring specific over general faithfulness constraints is conservatism (Hayes, 2004). If a particular candidate can be ruled out with a specific constraint rather than with a general one, then doing so provides an opportunity for markedness constraints to outrank the general version of the constraint, avoiding overgeneration. In a phonotactic learning context such as the one assumed in this paper, this is a crucial element in avoiding the acquisition of a fully faithful grammar (which accounts for all of the learning data but cannot discriminate ungrammatical inputs). In terms of this model's correspondence to a human learner, favouring specificity might reflect the likelihood of a language learner to take narrow, conservative steps toward their target grammar rather than making sweeping generalizations where they are not necessary or justified.

GLA simulations with both KE and SE inputs fail when the *a priori* ranking condition of ID-$\sigma_1$(Bk) $\gg$ ID(Bk) is not included. Results in Table 7a show that for KE, ID-$\sigma_1$(Bk) is not able to rise high enough to have a final ranking overID(Bk)). This bias is needed in order to preserve the vowels in initial syllables, even at the potential expense of being unfaithful to several subsequent vowels. Results in Table 7b show that for SE, it is the case both that ID-$\sigma_1$(Bk) has not risen far enough to outrank the segmental markedness constraints and preserve the first vowel even when it is marked, and also that ID(Bk) has risen too far, resulting in a fully faithful grammar.

| Constraint | Final value |
|---|---|
| GMH(e) | 100.3 |
| GMH(**F**) | 100.0 |
| ID(Bk) | 86.3 |
| *e | 75.0 |
| GMH(ɤ) | 73.1 |
| *ɤ | 70.0 |
| GMH(**B**) | 66.8 |
| ***F** | 64.7 |
| ID-$\sigma_1$(Bk) | 64.6 |
| ***B** | 62.2 |
| AGR(Bk) | 50.7 |

**(a)** With KE learning data.

| Constraint | Final value |
|---|---|
| ID(Bk) | 103.3 |
| *ɤ | 90.6 |
| *e | 90.4 |
| ***F** | 83.0 |
| ***B** | 82.9 |
| GMH(ɤ) | 81.7 |
| GMH(**F**) | 72.7 |
| ID-$\sigma_1$(Bk) | 70.3 |
| GMH(e) | 62.4 |
| GMH(**B**) | 38.2 |
| AGR(Bk) | 21.2 |

**(b)** With SE learning data.

**Table 7:** Final values of constraints after $4 \times 50000$ iterations of learning data, omitting bias for specific over general faithfulness.

### 5.1.2 Markedness over faithfulness

Conservatism again provides the rationale for favouring markedness over faithfulness constraints (Hayes, 2004). If a particular candidate can be ruled out with a markedness constraint rather than with a faithfulness constraint, then doing so provides an opportunity for markedness constraints to be active in the acquired grammar. For a phonotactic learner, this is crucial in terms of avoiding the acquisition of a fully faithful grammar. In terms of this model's correspondence to a human learner, favouring markedness might reflect the process of a language learner noting which kinds of forms seem to be permitted and which are marked, better equipping them to repair ungrammatical inputs when they are encountered.

GLA simulations with both KE and SE inputs fail when faithfulness and markedness constraints both start at the same value (e.g., 100). Results in Tables 8a and 8b show that for both dialects, although the top-ranked markedness constraints are the correct ones, ID-$\sigma_1$(Bk) and ID(Bk) are ranked above those and all other constraints. This produces fully faithful grammars despite the correct ranking of markedness constraints.

| Constraint | Final value |
|---|---|
| ID-$\sigma_1$(Bk) | 129.0 |
| ID(Bk) | 109.0 |
| GMH(**F**) | 101.1 |
| GMH(e) | 101.0 |
| *e | 97.8 |
| *ɤ | 97.1 |
| GMH(ɤ) | 97.1 |
| ***B** | 97.1 |
| GMH(**B**) | 97.1 |
| AGR(Bk) | 97.0 |
| ***F** | 96.9 |

**(a)** With KE learning data.

| Constraint | Final value |
|---|---|
| ID-$\sigma_1$(Bk) | 130.3 |
| ID(Bk) | 110.3 |
| *ɤ | 106.0 |
| ***F** | 104.3 |
| GMH(**F**) | 99.0 |
| GMH(ɤ) | 98.0 |
| ***B** | 96.0 |
| *e | 96.0 |
| GMH(**B**) | 94.0 |
| GMH(e) | 94.0 |
| AGR(Bk) | 92.0 |

**(b)** With SE learning data.

**Table 8:** Final values of constraints after $4 \times 50000$ iterations of learning data, omitting bias for markedness over faithfulness.

### 5.1.3 Magri update

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, GMH(**F**) and GMH(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 /ɤ/ (due to their segmental markedness as well as their harmony constraints) result in *e and *ɤ 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. 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.

The Magri (2012) update rule 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, GMH(**F**) and GMH(e), and the rest (Figure 5).

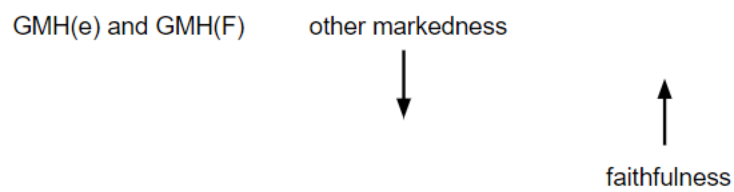


**Figure 5:** Creating space for faithfulness constraints between GMH(**F**) and GMH(e) vs other markedness constraints.

GLA simulations with KE inputs fail when the Magri update rule is not applied. Results in Table 9a show that, similarly to the case where faithfulness starts high (Section 5.1.2), the top-ranked markedness constraints are the correct ones but they are nevertheless below ID-$\sigma_1$(Bk) and ID(Bk), resulting in a fully faithful grammar. For SE, however, the Magri update rule is less relevant. Results in Table 9b show that the final ranking produced by the GLA without this modified update rule is very similar to the result from Section 4.3.2. The only caveat to note is that ID(Bk) continues to rise toward the value of ***F** over additional learning iterations. Simulations with 500,000 or 1,000,000 iterations, for example, produce less accurate grammars than those with 5,000 or 50,000 learning iterations.

## 5.2 Grammars

### 5.2.1 Non-initial harmony trigger

As footnote 3 alluded to, non-initial harmony triggers such as /i..e..u/ are a challenge for the theoretical approach that I have adopted, building on K&P. Access to ID-$\sigma_1$(Bk) along with the various GMH constraints provides a framework in which the vowel in the initial syllable is capable of driving harmony in KE. However, if the first one or more vowels are transparent /i/ there is no way under this system to prioritize a subsequent vowel as a harmony trigger, unless the /i/ is followed by only one subsequent vowel (in which case ID(Bk) will ensure that the second vowel surfaces faithful to its UR). Consider the ERC matrix in (24), in which the two possible harmonic candidates are compared for each input /i..e..**B**/ and /i..**B**..e/. The top-ranked (and up until this point, the only active) constraints are not active in these comparisons. The seven remaining constraints

| Constraint | Final value |
|---|---|
| ID-$\sigma_1$(Bk) | 138.2 |
| ID(Bk) | 118.2 |
| GMH(**F**) | 114.0 |
| GMH(e) | 112.0 |
| ***F** | 106.2 |
| AGR(Bk) | 105.8 |
| *e | 100.0 |
| GMH(ɤ) | 100.0 |
| *ɤ | 100.0 |
| GMH(**B**) | 97.8 |
| ***B** | 93.8 |

**(a)** With KE learning data.

| Constraint | Final value |
|---|---|
| ID-$\sigma_1$(Bk) | 120.0 |
| *ɤ | 108.0 |
| ***F** | 106.3 |
| ID(Bk) | 100.0 |
| GMH(ɤ) | 98.0 |
| ***B** | 94.0 |
| GMH(e) | 94.0 |
| *e | 92.0 |
| GMH(**F**) | 88.0 |
| AGR(Bk) | 86.0 |
| GMH(**B**) | 86.0 |

**(b)** With SE learning data.

**Table 9:** Final values of constraints after $4 \times 50000$ iterations of learning data, omitting Magri update rule.

are unable to consistently select the intended winners, not only because they are currently (as per (14) and (22)) grouped in one stratum, but also because the two lines form an inconsistent matrix, contradicting one another.

(24)

| input | candidates | GMH(**F**) | GMH(e) | ID-$\sigma_1$(Bk) | ID(Bk) | ***F** | ***B** | *ɤ | *e | AGR(Bk) | GMH(**B**) | GMH(ɤ) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| /i..e..**B**/ → | i..e..**F** ∼ i..ɤ..**B** | | | | | L | W | W | L | W | W | W |
| /i..**B**..e/ → | i..**B**..ɤ∼ i..**F**..e | | | | | W | L | L | W | L | L | L |

As described in Section 6, K&P do propose a faithfulness constraint IDENT-$F_1$(Back) that preserves the [back] feature of the first foot rather than just the first syllable; however, adopting such a constraint as a solution to this problem is easily overwhelmed by simply imagining an input with two initial /i/s rather than just one.

The general problem of ensuring that the first non-transparent vowel is the harmony trigger is independent of the learning approach used, whether LFCD, GLA, or any other learner. The issue of directionality is a challenge in the broader harmony literature as well (Baković, 2000; Jurgec, 2011), in this case specific to a context where harmony is characterized as being driven by a particular position.

### 5.2.2 Problematic markedness constraints

In Section 3.1 I present several constraints in addition to the ones proposed by K&P. Of particular interest for the purposes of this section of the discussion are the two markedness constraints *e (constraint 5) and ***B** (constraint 6), which are counterintuitive in that they are markedness constraints but that penalize unmarked segments. The inclusion of *e, as described in Section 3.1, is necessary if only to facilitate the inclusion of GMH(e), without which a KE grammar cannot differentiate between the transparent behaviour of front unmarked /i/ and the harmony behaviour of front unmarked /e/.

However, the inclusion of ***B** has a less clear motivation. In early runs of the learning simulations, it was not included in the constraint set and while LFCD was capable of acquiring a correct SE grammar without it, the GLA could not do the same. As Figure 6 depicts, ***F** (which would ideally be ranked between ID(Bk) and ID-$\sigma_1$(Bk) along with *ɤ) falls quite far and quite quickly early on in the simulation and does not have opportunity to recover. However, *ɤ, whose promoting or demoting inputs are similarly structured to those for ***F**, demonstrates a slower fall, with more oscillation in its descent. Inspection of the learning errors affecting the trajectories of these two constraints shows that while they are both demoted approximately the same number of times, *ɤ has the opportunity to be re-promoted more often because of its antagonistic relationship with *e. That is, anytime *e is demoted, *ɤ is promoted. Thus *ɤ benefits from having a symmetrical constraint in the set. Once ***B** was included to provide similar balance with ***F**, the GLA was able to successfully acquire the ranking in (23) and Table 6.

It is clear from the fact that LFCD succeeds without ***B** but GLA does not, that the underlying issue here is not in the constraint set but rather in its interaction with the gradual learning process. In the attempt to solve this learning problem with the addition of ***B** (and *e), it is not difficult to conceive of an unfortunate typological implication that arises from situating ID-$\sigma_1$(Bk) $\gg$ *e; ***B** at the top of a constraint ranking. This produces the problematic grammar in which only the *marked* vowels /F/ and /ɤ/ are permitted in non-initial syllables, rather than the unmarked ones /**B**/ and /e/. Thus upcoming work must explore other strategies for addressing this learning challenge that do not produce typologically unfortunate side effects.

### 5.2.3 Further typological implications

K&P frame their work in terms of the typology of vowel harmony patterns in many Balto-Finnic languages; however, as discussed below in Section 6, their typology does not include all related languages. In particular, 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 each of the four rankings presented in this paper (LFCD-learned and GLA-learned, for both KE and SE), the constraints GMH(ɤ) and GMH(**B**) are installed in

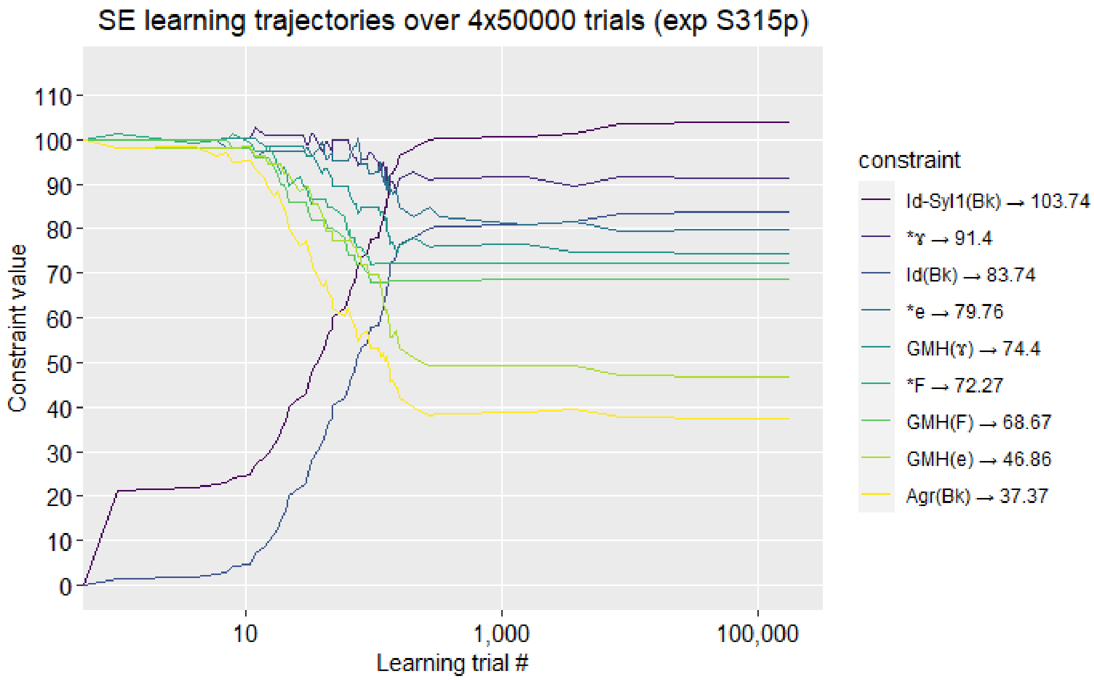


**Figure 6:** Changes in constraint values over 200,000 trials of SE learning data. The horizontal axis is truncated as there were no further learning errors after trial 100,000.

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 both ***F** and GMH(**F**) is clear through the work they do to identify winning candidates in SE and KE, respectively. *e is motivated by the need for GMH(e), in whose definition it is included. While *ɤ and ***B**, both necessary for different reasons, do not require the existence of their corresponding GMH constraints, I have included the harmony constraints for the sake of continuity and symmetry: if ***F** with GMH(**F**) and *e with GMH(e) are members of the set (that is, if CON is permitted to conjoin AGR(Bk) with some segmental markedness constraints), then ***B** along with GMH(**B**) and *ɤ along with GMH(ɤ) should be also. With that said, I have run simulations using both learning algorithms without these two GMH constraints, and both succeed (as predicted, given their inactivity in the rankings presented). This is worth noting, as the additional constraints discussed in Sections 5.2.2 and in this subsection may well add a much greater degree of complexity to this typology than is necessary or justified. The typological implications of including these constraints must be investigated and addressed in future work.

# 6 Adapting Kiparsky and Pajusalu's (2003) constraint set

K&P propose constraints and rankings to 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 Kihnu Estonian without also inadvertently characterizing /e/ as transparent as well. 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/.

K&P define three positional faithfulness constraints in their example (4): IDENT-$\sigma_1$(Back), IDENT-$F_1$(Back), and IDENTROOT(Back). The first privileges the initial syllable, resulting in that syllable being a position of maximal contrast and also being a harmony trigger. The second (faithfulness in the first foot) and third (faithfulness in the monomorphemic stem) result in harmony requirements being stricter in derived environments than in the root and/or initial feet. For this project, the EDC KE data have not yet been analyzed to the point where conclusions can be made about the obligatoriness of harmony in roots and initial feet vs elsewhere, though preliminary observations suggest that other than a privileged first syllable, harmony patterns are consistent through the entire word, independent of morpheme or foot. Thus I have omitted IDENT-$F_1$(Back) and IDENTROOT(Back), and for the purposes of learnability I have explicitly included IDENT(Back), a general faithfulness constraint that incurs a violation for each segment with a change in [back]'s parity. Tableau (25) demonstrates how specific and general faithfulness constraints facilitate maximal contrast in $\sigma_1$ while allowing harmony to be propagated through the rest of the word.

(25)

| /e..ɑ/ | ID-$\sigma_1$(Bk) | AGR(Bk) | ID(Bk) |
|---|---|---|---|
| a. e..ɑ | | $*!$ | |
| ☞ b. e..æ | | | $*$ |
| c. ɤ..ɑ | $*!$ | | $*$ |

K&P provide several examples of Balto-Finnic languages in which /e/ participates in harmony (N. Seto, S. Seto, N. Tarto; see K&P p. 219). However, they also state that "[u]nder every ranking of the proposed constraints, ... if even one front neutral vowel is transparent, *all of them must be*" (emphasis mine). They present featural markedness constraints in their example (2) that address only the marked segments /ɤ, æ, ø, y, ɨ/ in the languages they investigate. However, in order to distinguish /e/ from /i/, allowing it to participate in KE harmony rather than behave as a transparent neutral vowel, there will have to be additional constraints that deal with /e/; these are presented below.

Additionally, K&P propose a pair of Marked Harmony constraints. They first define the conjoined constraint

(26) Marked Harmony (MH): Agr(Bk) & ***F**[6] (K&P example (12))

and then continue to define both a specific and a general version (K&P examples (14-15)):

(27) Core MH: a vowel may not be both marked for a particular feature $f$ (in this case [back]) and disharmonic for $f$.

(28) Generalized MH: a domain may not contain both a vowel marked for $f$ (again, [back]) and a pair of syllable-adjacent vowels disharmonic for $f$.

For the purposes of analyzing KE, the approach presented has several shortcomings. First, “domain” - and therefore the violation profile embodied by GMH - is not precisely defined. I have addressed this in my definitions in Section 3.1. Second, conjoining Agr(Bk) with only one of the featural markedness constraints (here ***F**) means that even though a vowel sequence like /e..ɤ/ should be ungrammatical in KE, it will not be penalized by a MH constraint (it is penalized by Agr(Bk), but the transparency of /i/ requires that constraint to be ranked quite low). Hence my inclusion of Marked Harmony for /ɤ/ (conjunction of Agr(Bk) with K&P’s *ɤ), and my proposal of the additional segmental markedness constraint *e in order to facilitate Marked Harmony for /e/ as well.

The inclusion of GMH(e) (and therefore *e) is justified by the arguments presented via the ERC matrix in (29). Each row implies one ERC for KE which, in the absence of GMH(e), form a contradiction when considered together. For input /e..**B**/, preference of candidate [e..**F**] over candidate [e..**B**] requires Agr(Bk) to outrank both ***F** and Id(Bk), whereas for input /i..**B**/, preference of [i..**B**] over [i..**F**] requires either ***F** or Id(Bk) to outrank Agr(Bk). This inconsistency is circumvented by the inclusion of GMH(e). Beginning with the requirement (from the second ERC) that either ***F** or Id(Bk) must outrank Agr(Bk), it is then crucial for GMH(e) to outrank that constraint, in order to satisfy the first ERC. Thus at least one of GMH(e) $\gg$ ***F** $\gg$ Agr(Bk) or GMH(e) $\gg$ Id(Bk) $\gg$ Agr(Bk) must hold. Without the additional Marked Harmony constraint, the grammar would be unable to differentiate between the behaviour of sequences involving harmonic /e/ vs transparent /i/.

(29)

| input | candidates | Agr(Bk) | ***F** | Id(Bk) | GMH(**F**) | GMH(e) |
|---|---|---|---|---|---|---|
| /e..**B**/ → | e..**F** ∼ es..**B** | W | L | L | | W |
| /i..**B**/ → | i..**B** ∼ i..**F** | L | W | W | | |

[6] Referred to as ***æ, *ø, *y** by K&P.

Finally, for the sake of simplicity I have omitted K&P's COREMH constraints from these simulations.[7]

# 7 Future work

As is true for many aspects of this project, future work will have two main facets: some of the next steps will focus on the approach to learning, whereas others will probe the theoretical framework, independent of the learning models themselves.

Two primary learning goals for this work are to (a) incorporate variation in KE and (b) model the process of second-dialect acquisition in the context of KE and SE. Introducing variation to the learning data for KE is straightforward with the GLA; relative frequencies of multiple potential output forms for a single input can be presented as part of the input to the algorithm, which will sample the learning data according to the specified proportions. This, in turn, facilitates the refinement of the distance between constraint values, which enables the noisy evaluation of test data to replicate the proportion of each variant. In particular I intend to test the GLA's ability to learn the /e/-/ɤ/ variation in KE. However, more subtle variation and noise in the corpus can also be accommodated with this kind of approach, bringing the learning model even closer to the kind of reality that a human learner faces when acquiring a language.

Second-dialect acquisition (e.g. a SE speaker learning KE) could be modelled by using the constraint values from first-dialect acquisition (Section 4.3.2 results) as the initial values for second-dialect learning. This poses a challenge because harmony constraints are ranked relatively low in SE, with no way to promote them for a target KE grammar using only positive evidence. However, for a learner of a second dialect that is as closely related to the first as KE is to SE, there is no reason that such a learner should not take advantage of shared lexical items in the learning process. Unfaithful inputs could be considered as long as they are of a type such that a SE speaker, uttering a word in their first dialect, might hear the word restated (corrected, in some sense) by a KE speaker. This potential for negative as well as positive evidence could help shift constraint rankings from already-learned orderings to novel ones.

In terms of the theoretical underpinnings of the project, future work needs to address the issues raised in Section 5.2; that is, (a) the general theoretical stumbling block of non-initial harmony triggers in a positionally-driven system, and (b) the typological implications of adding not only *e and ***B**, but also GMH(ɤ), GMH(e), and GMH(**B**) to the constraint set proposed by K&P. Alternate approaches to account for what are apparently

[7] K&P provide examples of languages (e.g. Vepsian, Khanty) that require Core and Generalized marked harmony constraints to be distinguished from each other, in order to account for languages in which harmony in adjacent vowels is obligatory but harmony in vowels separated by a neutral vowel is not. However, this distinction does not exist in SE or KE. See K&P for further discussion.

positionally-driven harmony patterns must be investigated and evaluated with respect to the KE patterns. Constraints that were proposed in this paper to help account for the vowel patterns described must be analyzed in terms of their effect on the predicted typology for related vowel harmony phenomena.

# 8 Conclusion

The constraint set proposed by K&P to account for the vowel harmony patterns of Balto-Finnic languages is insufficient to describe all such languages, in particular KE. Adaptations to this set were necessary 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 rankings that correspond to correct and restrictive grammars for both the vowel harmony of KE and the positional restrictions of SE. An existing corpus (EDC) 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, three biases were necessary in order for acquisition to be successful: (1) specific faithfulness must outrank general faithfulness, (2) faithfulness constraints must start low compared to markedness constraints, and (3) the Magri (2012) update rule must be used. This finding shows that even with the simple patterns as those found in KE and SE and a capable set of constraints proven by LFCD, there are challenges inherent to the use of GLA that require a delicately balanced learning environment and a nuanced approach on the part of the learner.

## References


Asu, E. L., Lippus, P., Niit, E., & Türk, H. (2012). The acoustic characteristics of monophthongs and diphthongs in the Kihnu variety of Estonian. *Linguistica Uralica*, *48*(3), 161–170. https://doi.org/10.3176/lu.2012.3.01

Asu, E. L., & Teras, P. (2009). Estonian. *Journal of the International Phonetic Association*, *39*(3), 367–372. https://doi.org/10.1017/S002510030999017X

Baković, E. (2000). *Harmony, dominance, and control* (Doctoral dissertation) [ROA-360]. Rutgers University. New Brunswick. http://roa.rutgers.edu

Boersma, P., & Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistic inquiry*, *32*(1), 45–86.

Dresher, B. E. (1999). Charting the learning path: Cues to parameter setting. *Linguistic inquiry*, *30*(1), 27–67.

Hayes, B. (2004). Phonological acquisition in optimality theory: The early stages. Cambridge University Press.

Hayes, B., Tesar, B., & Zuraw, K. (2013). OTSoft [software package] [Version 2.6 beta]. Retrieved April 11, 2021, from http://www.linguistics.ucla.edu/people/hayes/otsoft/

Jurgec, P. (2011). *Feature spreading 2.0: A unified theory of assimilation* (Doctoral dissertation). University of Tromsø. Tromsø.

Kiparsky, P., & Pajusalu, K. (2003). Towards a typology of disharmony. *Linguistic Review*, *20*, 217–241. https://doi.org/10.1515/tlir.2003.009

Léonard, J. L. (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*.

Lindström, L. (2013). Eesti murdekorpus [Estonian dialect corpus]. https://doi.org/10.15155/TY.0007

Lombardi, L. (1999). Positional faithfulness and voicing assimilation in optimality theory. *Natural Language & Linguistic Theory*, *17*(2), 267–302. https://doi.org/10.3176/lu.2012.3.01

Magri, G. (2012). Convergence of error-driven ranking algorithms. *Phonology*, *29*(2), 213–269.

Pajusalu, K., Pomozi, P., Juhász, D., & Viitso, T.-R. (2012). Sociolinguistic comparison of the development of Estonian and Hungarian dialect areas. *Linguistica Uralica*, *48*(4), 241–264. https://doi.org/10.3176/lu.2012.4.01

Prince, A. (2002). Entailed ranking arguments [ROA-500]. http://roa.rutgers.edu

Sang, J. (2009). Ühest fonotaktilisest kollisioonist (Kihnu näitel) [On a phonotactic collision (on the example of Kihnu subdialect)]. *Keel Ja Kirjandus*, *52*(11), 809–817.

Simonet, M. (2016). The phonetics and phonology of bilingualism. In D. EIC (Ed.), *Oxford handbooks online: Scholarly research reviews* (pp. 1–25). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199935345.013.72

Van Rossum, G., & Drake, F. L. (2009). Python [programming language] [Version 3.8]. Retrieved November 1, 2021, from http://www.python.org/

# Appendices

## A Corpus

The Estonian Dialect Corpus (EDC; Lindström, 2013) was used as the data source for the SE and KE learning portions of this project. The EDC 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. SE, unlike KE, does not have a precise geographic association, but is associated with the Central Dialect area and Tallinn in particular (Pajusalu et al., 2012). Therefore the subset of the EDC that was used to inform the SE learning data was extracted by restricting entries to those from the Central dialect group, Keila and Kose parishes (the EDC does not contain any recordings made in Tallinn itself; however, these two parishes are direct neighbours).

Both subsets include over 20,000 word tokens. They have been analyzed for word type counts, identification of monophthongs, and grammaticality with respect to the dialect in question; see Table 10.

Maybe clarify that by this you (presumably) mean consistency with the descriptive generalizations for these dialects as per the literature.

| | KE corpus | SE corpus |
|---|---|---|
| Total entries (word tokens) | 21,599 | 23,971 |
| Total entries (word types) | 5052 | 5017 |
| Types with $\geq 2$ monophthongs | 3375 | 3493 |
| ... of which $\geq 3$ | 1200 | 1335 |
| Word-initial monophthong bigrams | 1973 | 946 |
| Word-initial monophthong trigrams | 1173 | 1010 |
| *Faithful* word-initial monophthong bigrams | 1935 | 940 |
| *Faithful* word-initial monophthong trigrams | 1052 | 1006 |

**Table 10:** Corpus analysis for KE (vowel harmony) and SE (positional restrictions).

Vowels in sets {æ, ø, y} and {ɑ, o, u} were collapsed and replaced with their representative symbols **F** (front marked vowels) and **B** (back unmarked vowels). Then, in order to inform relative frequencies of learning data as described in Section 4, the number of instances of each word-initial vowel bigram or trigram was counted in each dialect subset of the corpus; see Table 11 for excerpts.

| Vowel sequence | KE corpus | SE corpus |
|---|---|---|
| i..i | 100 | 100 |
| i..e | 81 | 85 |
| i..ɤ | 5 | 0 (ungrammatical) |
| i..**F** | 53 | 0 (ungrammatical) |
| i..**B** | 159 | 314 |
| | … | |
| **F**..i | 107 | 141 |
| **F**..e | 117 | 151 |
| **F**..ɤ | 0 (ungrammatical) | 0 (ungrammatical) |
| **F**..**F** | 185 | 11 (ungrammatical) |
| **F**..**B** | 31 (ungrammatical) | 194 |
| | … | |
| e..**F**..i | 11 | 0 (ungrammatical) |
| e..**F**..e | 9 | 0 (ungrammatical) |
| e..**F**..ɤ | 0 (ungrammatical) | 0 (ungrammatical) |
| e..**F**..**F** | 12 | 0 (ungrammatical) |
| e..**F**..**B** | 1 (ungrammatical) | 0 (ungrammatical) |
| | … | |

**Table 11:** Samples of relative frequencies for ngrams in KE and SE corpora.

Many of the vowel sequences that are ungrammatical in my idealization of these dialects were unattested (as expected), as is evident from Table 11; however, there were also some that did nevertheless appear in the corpus. The incidence of unexpected forms in SE is easily attributed to the general noisiness of spontaneous speech data. In KE, however (and especially in KE trigrams), the proportion of forms that don't fit the expected patterns is somewhat larger. Some of this might certainly fall under the umbrella of noise, but further investigation of these tokens must be undertaken to determine their underlying causes (whether the previously described variability, their status as a loanword, variation in speakers, as-yet-hidden morphological effects, or other dialectal phenomena). However, the incidence of such ungrammatical forms was relatively low, especially for SE; see Table 12.

| Dialect / ngram | Number ungrammatical | Total number | Proportion ungrammatical |
|---|---|---|---|
| KE bigrams | 38 | 1973 | 0.019 |
| KE trigrams | 121 | 1173 | 0.103 |
| SE bigrams | 6 | 946 | 0.006 |
| SE trigrams | 4 | 1010 | 0.004 |

**Table 12:** Rates of incidence of ungrammatical forms in KE and SE subsets of EDC.

# 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 4) and three biases (low faithfulness, specific over general faithfulness, and Magri update) described in Section 4.2. Most of these options can be selected and varied independently of but in concert with the others. However, when both *a priori* rankings and Magri update are selected, the OTSoft ranking history file shows that only the *a priori* ranking is maintained; the Magri update is not simultaneously employed.

In order to address this gap, I wrote a Python (Van Rossum & Drake, 2009) script[8] to run the GLA, ensuring that the specifics of the implementation would facilitate the simultaneous application of all three necessary biases. I describe below how my implementation differs from that of OTSoft, and thus allows for simultaneous application of learning biases.

The low-faithfulness bias is treated the same way by both OTSoft and my own implementation. Each simply assigns particular values to faithfulness and markedness constraints (0 and 100 respectively, in this case). However, the difference lies in the application of various combinations of the other two biases. For example, in OTSOft:

1. (a) If neither *a priori* ranking nor Magri update are employed, each error occupies one line in the ranking history and all of the resulting updates to constraint values are done in one sweep. See Table 13a.

2. (b) If only the Magri update (but not *a priori* ranking) is involved, each error occupies one line in the ranking history and all of the resulting updates to constraint values are done in one sweep. This is necessary for the Magri update, since the actual sizes of the promotions depend on the numbers of promotions and demotions involved in the update. Many of the promotions in this ranking history file

[8] Available at https://github.com/kvesik/learning-Estonian-dialects

have fractional values as compared to the demotions, as per the Magri update rule calculations. See Table 13b.

3. (c) If only *a priori* rankings (but not the Magri update) are involved, each error occupies several lines in the ranking history file. Each constraint that needs updating gets its own line, and if after any of these individual constraint updates the *a priori* ranking is violated, the higher constraint's (ID-$\sigma_1$(Bk)) value is adjusted (on yet another line) to ensure that it is at least 20 greater than the lower one's (ID(Bk)) value. See Table 13c.

4. (d) If both *a priori* rankings and Magri update are selected, the ranking history file has the same structure as in scenario (c) above, with each constraint's update on a separate line. All promotions have their full, unscaled value; that is, the Magri update rule calculations are not applied.

In my Python implementation, it is always the case that all necessary updates associated with one error are done in one sweep (including any Magri-adjusted promotion values) and recorded on the same line of the ranking history file. The only exception to this is if the *a priori* rankings are violated after those updates are complete. In that case, there will be a second line to reflect the necessary adjustment to the higher-ranked constraint (ID-$\sigma_1$(Bk)) in the relationship. See Table 14. This facilitates the application of both the Magri update rule and the *a priori* ranking bias.

| Generated | Heard | ***F** | now | ***B** | now | *γ | now | … | ID-$\sigma_1$(Bk) | now | ID(Bk) | now | … |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (Initial) | | | 100 | | 100 | | 100 | … | | 0 | | 0 | … |
| Fe | Bõ | 2 | 102 | -2 | 98 | -2 | 98 | … | 2 | 2 | 2 | 2 | … |
| FF | BB | 2 | 104 | -2 | 96 | | | … | 2 | 4 | 2 | 4 | … |
| iõ | ie | | | | | 2 | 100 | … | | | 2 | 6 | … |
| õõõ | eee | | | | | 2 | 102 | … | 2 | 6 | 2 | 8 | … |

**(a)** OTSoft ranking history when neither *a priori* ranking nor Magri update is selected.

| Generated | Heard | ***F** | now | ***B** | now | *γ | now | … | ID-$\sigma_1$(Bk) | now | ID(Bk) | now | … |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a priori Id(Bk)Syl1>>Id(Bk) | | | | | | | | … | 20 | 20 | | | … |
| (Initial) | | | 100 | | 100 | | 100 | … | | 0 | | 0 | … |
| Fi | Bi | 2 | 102 | | | | | … | | | | | … |
| Fi | Bi | | | -2 | 98 | | | … | | | | | … |
| Fi | Bi | | | | | | | … | 2 | 22 | | | … |
| Fi | Bi | | | | | | | … | | | 2 | 2 | … |
| iei | iõi | | | | | -2 | 98 | … | | | | | … |
| iei | iõi | | | | | | | … | | | 2 | 4 | … |
| a priori Id(Bk)Syl1>>Id(Bk) | | | | | | | | … | 2 | 24 | | | … |

**(b)** OTSoft ranking history when *a priori* ranking but not Magri update is selected. The output is the same even if both biases are selected.

| Generated | Heard | ***F** | now | ***B** | now | *γ | now | … | ID-$\sigma_1$(Bk) | now | ID(Bk) | now | … |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (Initial) | | | 100 | | 100 | | 100 | … | | 0 | | 0 | … |
| Fee | Bõõ | 2 | 100.8 | -2 | 98 | -2 | 98 | … | 2 | 0.8 | 2 | 0.8 | … |
| iBõ | iFe | -2 | 98.8 | 2 | 98.57 | 2 | 98.57 | … | | | 2 | 1.37 | … |
| BBB | FFF | -2 | 96.8 | 2 | 99.07 | | | … | 2 | 1.3 | 2 | 1.87 | … |
| Bõ | Fe | -2 | 94.8 | 2 | 99.87 | 2 | 99.37 | … | 2 | 2.1 | 2 | 2.67 | … |

**(c)** OTSoft ranking history when Magri update but not *a priori* ranking is selected.

**Table 13:** OTSoft ranking history file excerpts for KE inputs.

| Generated | Heard | ***F** | now | ***B** | now | *ɣ | now | … | Id-$\sigma_1$(Bk) | now | Id(Bk) | now | … |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (Initial) | | | 100 | | 100 | | 100 | … | | 0 | | 0 | … |
| a priori Id(Bk)Syl1>>Id(Bk) | | | | | | | | … | 20 | 20 | | | … |
| õB | eF | -2 | 98 | 0.8 | 100.8 | 0.8 | 100.8 | … | 0.8 | 20.8 | 0.8 | 0.8 | … |
| Fii | Bii | 1.5 | 99.5 | -2 | 98.8 | | | … | 1.5 | 22.3 | 1.5 | 2.3 | … |
| iei | iõi | | | | | -2 | 96.8 | … | | | 2 | 4.3 | … |
| a priori Id(Bk)Syl1>>Id(Bk) | | | | | | | | … | 2 | 24.3 | | | … |

**Table 14:** Python script ranking history for KE when both *a priori* ranking and Magri update are selected.