This is Chapter 4/5 only. Refer to 20240818 dissertation outline and notes for complete outline.

Kaili's very official dissertation draft

For purposes of graduating in the near-ish future

4/5. Learning biases and optimal learning conditions

As per usual, my initial work will be done in Google Docs while the final draft(s) will be in LaTeX. So any formatting issues or nuances will be sorted out later.

Yellow highlighting indicates updates/follow-up needed.

Notes and/or specific requests for feedback highlighted in blue throughout.

Prior to this chapter, I will have introduced:

- The balto-finnic languages / vowel pattern typology
- Representative sample languages (patterns) for the above
- My constraint set & OT analysis of the sample languages
- Learning algorithms, in particular GLA-type learners
    - Definition of idempotence
    - Definition of θ-notation for ranking values
    - Explain my method of calculating "average frequency of correct results"
- My methods for simulations - python script etc
- Simulated learning data for the sample languages

4.2. F_{spec} >> F_{gen}

The constraint set that I use for this project includes only two faithfulness constraints, **ID(Bk)** and **ID(Bk)-σ_1**, the first applying more broadly and the second in a narrower context. When two such versions of a faithfulness constraint exist, it is possible to construct a grammar in which marked elements in underlying forms surface only in privileged contexts. For example, the ranking **ID(Bk)-σ_1** >> ***F_3** >> **ID(Bk)** bans vowels in set **F_3** in general, but permits them in initial syllables.

A specific-over-general faithfulness bias (F_{spec} >> F_{gen}) is a strategy that can help find the most restrictive grammar that accounts for the input data, avoiding a superset (overgenerating) grammar (Hayes 2004, Tessier 2007). I take two slightly different approaches to this idea, detailed in Sections 4.2.1 and 4.2.2.

4.2.1. *A priori* bias

One approach to the specific-over-general faithfulness bias is to ensure that the ranking value for the specific version of the constraint is a minimum specified distance higher than that of the general version [TODO citation?]. The satisfaction of this bias is checked persistently through the learning simulation, both in the initial state and after each individual learning update.

4.2.1.1. Rationale

Maintaining a minimum difference between the ranking values of a specific-general pair of faithfulness constraints ensures that the specific version of the constraint always has a better opportunity to claim credit for a particular output form than the general one does, corresponding to a more restrictive grammar overall.

4.2.1.2. Implementation

The F_{spec} >> F_{gen} bias between any specific-general pair of faithfulness constraints can be implemented by means of an *a priori* bias that ensures $\theta_{Fspec} - \theta_{Fgen} \geq d$, for some distance *d*. Practically, the learner adjusts the initial ranking values such that any two constraints in this type of relationship are at least *d* apart, and then does the same after each learning update. If the two constraints have a difference of less than *d*, then it is always the case that the specific one has its value increased rather than the general one having its value decreased. OTSoft (Hayes et al, 2013a) sets the default value of this difference to be *d = 20*, stating that it is "very close probabilistically to being an obligatory ranking" (Hayes et al, 2013b: 24).

In my learning simulations, I test the omission of this bias as well as a range of different *d* values: 0 (θ_{Fspec} must be no less than θ_{Fgen}), 10, 20, 30, and 40.

4.2.1.3. Simulation results - *a priori*

To demonstrate the effect of the *a priori* bias, I simulate acquisition of the three sample languages using Learner B, defined with the settings in Table 8. The selection of *d* = 20 for illustrative purposes is drawn from the OTSoft default as mentioned above. Results using learners with other values of *d* are summarized in Section 4.5.

Parameter	**Setting**
All basic parameters	Default
A priori bias	*d* = 20

Table 8. Parameter settings for Learner B.

With the *a priori* bias set to *d* = 20, learning simulations for all three sample languages fail to acquire the target grammars. The learner trained on North Estonian data produces a grammar that, while not correct, does have some promising characteristics. On the other hand, the learners trained on Finnish and North Seto data once again produce fully-faithful grammars. Test results are summarized in Table 9.

Language	**Average frequency of correct outputs**
Finnish	0.2895
North Estonian	0.7796
North Seto	0.3133
Overall	0.4608

Table 9. Summary of results from simulations with Learner B.

Finnish: Table 10 shows the final ranking values for a selection of crucial constraints, after learning from simulated Finnish data. Both faithfulness constraints have risen to the top. Id(Bk)'s distance above $*B_2$ and the relevant no-disagreement constraints is small enough that evaluation noise might cause it to swap rankings with one of its neighbours. However, in order to meet the crucial rankings

$*B_2$, $*F_3\ldots\underline{B}_5$, $*F_3\underline{B}_5$, $*\underline{B}_5\ldots F_3$, $*\underline{B}_5F_3$ >> Id(Bk)σ_1 >> Id(Bk)

proposed in Section 3.X, such swaps would have to be guaranteed to occur at every evaluation, which is extremely unlikely given the final ranking values. Hence the final grammar produced by Learner B on Finnish inputs is more or less fully faithful, with a representative evaluation shown in Tableau (1).

Constraint	**Final ranking value**
Id(Bk)σ_1	136.000
Id(Bk)	116.000
$*\underline{B}_5\ldots F_3$	112.000
$*B_2$	110.000
$*F_3\underline{B}_5$	110.000
$*F_3\ldots\underline{B}_5$	110.000
$*\underline{B}_5F_3$	110.000

…	…

Table 10. Excerpt of final ranking values for Finnish after simulation with Learner B.

(1) Sample evaluation of input /o..æ/ in the Finnish grammar acquired by Learner B. The grammar selects the faithful candidate [o..æ] as optimal even though it is not harmonic.

	/o..æ/	Id(Bk)σ_1	Id(Bk)	*$\underline{B}_5 \ldots F_3$	… (constraints such as *B_2, *$F_3\underline{B}_5$, *$F_3 \ldots \underline{B}_5$, *$\underline{B}_5 F_3$, and others)
☞	o..æ			*	
	o..ɑ		*!		
	ø..æ	*!	*		
	ø..ɑ	*!	**		

[TODO possibly adjust the following paragraph / table after 20241128 discussion with AMT]

In theory it should have been reasonable for Id(Bk)σ_1 to end up with a final ranking value greater than or equal to the top-ranked markedness constraints with Id(Bk) lower down. However, at the time that Id(Bk)σ_1 approaches the highest-ranked markedness constraints (including *B_2 with θ=110), the other context-free markedness constraints all have values in [100, 106] and are therefore within a small enough window for evaluation noise to make (e.g.) *B_3 or *F_3 active in selecting the optimal candidate (see Table 10.1). This results in errors and therefore more updates which push the faithfulness constraints ever higher. It is only once Id(Bk) surpasses this clump of markedness constraints that errors taper off and the learner converges.

Trial number 69 i..e..ɑ ~ ɯ..ɤ..ɑ	Trial number 73 u..u..ɑ ~ u..y..æ	Trial number 90 u..y..ø ~ i..u..o
*B_2 110 Id(Bk)σ_1 108 *B_3 106 *F_3 104 *B_1 104 *$\underline{B}_5 \ldots F_3$ 104 *F_1 102 *$F_3\underline{B}_3$ 102 *$F_3 \ldots \underline{B}_3$ 102 *$F_3\underline{B}_5$ 102 *$F_3 \ldots \underline{B}_5$ 102 *$\underline{B}_3 \ldots F_3$ 102 *$\underline{B}_5 F_1$ 102	Id(Bk)σ_1 112 *B_2 110 *B_3 108 *$\underline{B}_5 \ldots F_3$ 106 *F_3 104 *B_1 104 *$F_3 \ldots \underline{B}_3$ 104 *$F_3\underline{B}_5$ 104 *$F_3 \ldots \underline{B}_5$ 104 *$\underline{B}_5 F_3$ 104 *$F_3\underline{B}_3$ 102 *$F_4 \ldots \underline{B}_3$ 102 *$\underline{B}_3 \ldots F_3$ 102	Id(Bk)σ_1 120 *B_2 110 *$\underline{B}_5 \ldots F_3$ 108 *B_3 106 *$F_3\underline{B}_5$ 106 *$F_3 \ldots \underline{B}_5$ 106 *$\underline{B}_5 F_3$ 106 *F_3 104 *B_1 104 *$F_3 \ldots \underline{B}_3$ 104 *$\underline{B}_5 F_1$ 104 *$\underline{B}_5 \ldots F_1$ 104 *F_1 102

$^*\underline{B}_5 \ldots F_1$	102	$^*\underline{B}_5 F_1$	102	$^*F_3\underline{B}_3$	102
$^*\underline{B}_5 F_3$	102	$^*\underline{B}_5 \ldots F_1$	102	$^*\underline{B}_3 \ldots F_3$	102
$^*F_4$	100	$^*F_1$	100	Id(Bk)	100
$^*F_5$	100	$^*F_4$	100	$^*F_4$	100
$^*B_5$	100	$^*F_5$	100	$^*F_5$	100
…		$^*B_5$	100	$^*B_5$	100
Id(Bk)	88	…		…	
		Id(Bk)	92		

Table 10.1. The highest of Finnish Learner B's constraint ranking values after three different learning updates. Although the crucial constraints (ideally $^*B_2$ >> Id(Bk)σ_1 >> $^*\underline{B}_5 \ldots F_3$, $^*F_3\underline{B}_5$, $^*F_3 \ldots \underline{B}_5$, $^*\underline{B}_5 F_3$) are in reasonably good positions, constraints such as $^*B_3$, $^*B_5$, $^*F_1$, $^*F_3$, $^*F_4$, $^*F_5$ are near enough to be potentially disruptive.

North Estonian: The final ranking values for a selection of crucial constraints, after learning from simulated North Estonian inputs, are shown in Table 11. Several of the crucial relative rankings $^*B_1$ >> Id(Bk)σ_1 >> $^*F_3$, $^*B_2$ >> Id(Bk), proposed in Section 3.X, are met by this grammar. However, one of the key elements – the full ban on vowels from set B_1 – is missing, by virtue of that fact that $^*B_1$'s final value is not only not at the top, but below even Id(Bk). Thus the acquired grammar will incorrectly permit B_1 vowels in initial syllables.

Constraint	Final ranking value
Id(Bk)σ_1	129.220
$^*B_2$	115.000
…	…
$^*F_3$	110.220
Id(Bk)	109.220
$^*B_1$	108.000
$^*\underline{B}_5 \ldots F_3$	108.000
…	…
$^*F_5 \ldots \underline{B}_2$	106.000
…	…
$^*F_5\underline{B}_2$	104.600
…	…
$^*\underline{B}_5 F_3$	104.000
…	…

Table 11. Excerpt of final ranking values for North Estonian after simulation with Learner B. A selection of no-disagreement constraints are included here for the purpose of comparison with results of simulations discussed in subsequent sections.

The ranking acquired by this learner does generally follow the required positional restrictions by ranking Id(Bk)σ_1 >> $^*F_3$, $^*B_2$ >> Id(Bk); however, the ranking values are close enough together that the stochastic nature of evaluation results in somewhat variable adherence to these positional restrictions. For example, ungrammatical test input /y..æ/ would be expected to surface as [y..ɑ], neutralizing the restricted vowel in the second syllable. However, during testing, this grammar selects output candidates with the frequencies shown in Table 12.

/y..æ/	Output frequency
y..æ	0.35
y..ɑ	0.65
u..æ	0.00
u..ɑ	0.00

Table 12. Frequency of candidate selection for input /y..æ/ with North Estonian grammar acquired by Learner B. Number of sample evaluations = 100.

Although there is some crowding, at least one success the North Estonian learner achieves that the Finnish one does not is that the learner converges with several markedness constraints between the two faithfulness constraints. [TODO add some explanation here after 20241128 discussion with AMT]

North Seto: Table 13 shows the final ranking values for a selection of crucial constraints, after learning from simulated North Seto data. Similar to the Finnish results, both faithfulness constraints have risen to the top. Though the markedness constraints are correctly ordered relative to each other, the relative positions of the faithfulness vs the markedness constraints are not correct with respect to the crucial rankings proposed in Section 3.X:

$*F_4 \ldots \underline{B}_5$, $*F_4\underline{B}_5$, $*\underline{B}_5 \ldots F_4$, $*\underline{B}_5F_4$, Id(Bk)syl1 >> *B1 >> Id(Bk)

Again, the final grammar produced by Learner B on North Seto inputs is essentially fully faithful, with similar learning challenges as described for Finnish.

Constraint	Final ranking value
Id(Bk)σ_1	136.000
Id(Bk)	116.000
$*F_4 \ldots \underline{B}_5$	110.000
$*\underline{B}_5 \ldots F_4$	110.000
$*F_4\underline{B}_5$	108.000
$*\underline{B}_5F_4$	108.000
…	…
$*B_1$	104.000
…	…

Table 13. Excerpt of final ranking values for North Seto after simulation with Learner B.

Results from both Section 4.2.1 and Section 4.2.2 are discussed in Section 4.2.4.

4.2.2. Favour specificity

The underlying idea for the favour-specificity bias is to allow the specific faithfulness constraint to rise independently of the general one, similar to the *Favour Specificity* principle that Hayes (2004) introduces for the Low-Faithfulness Constraint Demotion algorithm. Although that proposal focuses on a different algorithm, the same principle can be adapted to apply to the GLA as well.

4.2.2.1. Rationale

As discussed in 4.2.1, setting an *a priori* bias helps specific faithfulness constraints stay above their general counterparts. However, because each violation of a first-syllable faithfulness constraint is also necessarily a violation of a general faithfulness constraint, there is no opportunity for the specific constraint to ever rise any further above the general version than the *a priori* bias specifies. That is, it is always the case that either the pair of constraints is moving in tandem (if there is an error in the first syllable only) or the general constraint is “pushing” the specific one up form below (if there is at least one error in a non-initial syllable). Both of these scenarios have the same effect: the specific constraint does not ever move independently of the general one. Recall the ERC matrix in Table 7 for an illustration of this phenomenon.

This type of movement, where specific and general constraints are separated by what is effectively a constant distance, can cause a challenge for the learner in that the *d* value that is specified for the *a priori* bias may or may not be large enough for other necessary constraints and/or interactions to “fit” between the two faithfulness constraints, depending on the target grammar. For instance, suppose the target grammar has crucial rankings $M_1 >> F_1 >> F_2 >> M_2$, and the learner is set to its task with a fixed value (e.g., $d = 20$) assigned to the *a priori* bias. The $F_1 >> F_2$ relationship will be effectively categorical, which is sufficient for this grammar. However, suppose the target grammar has instead crucial rankings $F_1 >> M_1 >> F_2 >> M_2$. In this case, $d = 20$ does not create enough space: the constraints in either of the crucial rankings $F_1 >> M_1$ or $M_1 >> F_2$ (or both) will have ranking values close enough that evaluation noise will create some variability in surface forms. Conversely, attempting to solve this problem by arbitrarily setting the a priori bias to be larger can cause other issues instead (for example, it would prevent the learning of a target grammar where $F_1 >> M_1 >> F_2$ but M_1 must be variably interchangeable with both F_1 and F_2).

To address this challenge, I take an approach that allows the space between specific and general counterparts to change, depending on the kinds of errors that are made.

4.2.2.2. Implementation

As always, when a learning error triggers an update to the constraint ranking values, the relevant ERC is inspected for winner-preferring vs loser-preferring constraints. In this case, if both the specific and the general version of a particular faithfulness constraint are eligible for promotion (i.e., both prefer the winner), then only the specific one gets promoted.[2]

There is also a set of optional variations to this implementation, in which the *a priori* bias (if any; see Section 4.2.1) increases if the current θ_{Fspec} - θ_{Fgen} difference is greater than d. This increase is applied after all constraints involved in a learning update have been adjusted, and sets the new $d' = d + \Delta$. The three variations are (*only* occurring if $\theta_{Fspec} - \theta_{Fgen} > d$):

- Variation 1: $\Delta = (\theta_{Fspec} - \theta_{Fgen}) - d$ — expand the entire current difference
- Variation 2: $\Delta = ((\theta_{Fspec} - \theta_{Fgen}) - d)/2$ — expand half of the current difference
- Variation 3: $\Delta = ((\theta_{Fspec} - \theta_{Fgen}) - d)/i$, where i is the current learning trial number — expand a decreasing fraction of the current difference

[Have I done a poor job of explaining the variations? That last paragraph feels sort of convoluted. Also I'm not sure how far I want to actually go on this front, since it was not useful.]

4.2.2.3. Simulation results - Favour Specificity

To demonstrate the effect of the Favour Specificity bias, I simulate acquisition of the three sample languages using Learner C, defined with the settings in Table 14.

Parameter	Setting
All basic parameters	Default
Favour Specificity bias	Active

Table 14. Parameter settings for Learner C.

With the Favour Specificity bias applied, learning simulations for all three sample languages fail to acquire the target grammars. Once again, the grammar acquired by the learner trained on North Estonian is a significant improvement over the one acquired by Learner A, but the Finnish and North Seto grammars are essentially full faithful. Test results are summarized in Table 15.

Language	Average frequency of correct outputs
Finnish	0.2931
North Estonian	0.8998
North Seto	0.4260
Overall	0.5397

Table 15. Summary of results from simulations with Learner C.

[2] The general one still counts toward *W*, the total number of winner-preferring constraints, but it will not be affected by the update.

Based solely on its average frequency of correct outputs, North Seto Learner C appears to have shown some improvement over Learner A. However, this is in fact a statistically convenient side effect of final ranking values that are no better from a theoretical perspective; further explanation is provided below.

Finnish: Table 16 shows the final ranking values for a selection of crucial constraints, after learning from simulated Finnish data. Once again, both faithfulness constraints have risen to the top and the final grammar does not meet the crucial rankings proposed in Section 3.X:

$*B_2$, $*F_3 \ldots \underline{B}_5$, $*F_3\underline{B}_5$, $*\underline{B}_5 \ldots F_3$, $*\underline{B}_5F_3$ >> Id(Bk)σ_1 >> Id(Bk)

Rather, the final grammar produced by Learner B on Finnish inputs is more or less fully faithful.

Constraint	Final ranking value
Id(Bk)σ_1	116.000
Id(Bk)	112.000
$*F_3 \ldots \underline{B}_5$	110.000
$*\underline{B}_5 \ldots F_3$	110.000
$*B_2$	108.000
$*\underline{B}_5F_3$	108.000
…	…
$*F_3\underline{B}_5$	104.000
…	…

Table 16. Excerpt of final ranking values for Finnish after simulation with Learner C.

North Estonian: The final ranking values for a selection of crucial constraints, after learning from simulated North Estonian inputs, are shown in Table 17. As for Learner A, several of the crucial relative rankings $*B_1$ >> Id(Bk)σ_1 >> $*F_3$, $*B_2$ >> Id(Bk), proposed in Section 3.X, are met by this grammar. The issue of the full ban is still relevant - $*B_1$ is still not at the top of the rankings - but at least it is above Id(Bk).

Constraint	Final ranking value
Id(Bk)σ_1	124.000
$*F_3$	112.002
$*B_2$	112.000
$*\underline{B}_5 \ldots F_3$	110.000
…	…
$*B_1$	106.000
$*F_5 \ldots \underline{B}_2$	106.000
$*\underline{B}_5F_3$	106.000
…	…
$*F_5\underline{B}_2$	104.000
…	…
Id(Bk)	70.002
…	…

Table 17. Excerpt of final ranking values for North Estonian after simulation with Learner C. A selection of no-disagreement constraints are included here for the purpose of comparison with results of simulations discussed in subsequent sections.

There is also a great deal more space between Id(Bk)σ_1 and Id(Bk), allowing for more-categorical relationships between the constraints of interest. For example, when given ungrammatical test input /y..æ/, Table 18 shows that this grammar selects the intended output [y..ɑ] in 100% of test evaluations (compare with only 65% for Learner B as shown in Table 12).

/y..æ/	Output frequency
y..æ	0.00
y..ɑ	1.00
u..æ	0.00
u..ɑ	0.00

Table 18. Frequency of candidate selection for input /y..æ/ with North Estonian grammar acquired by Learner C. Number of sample evaluations = 100.

North Seto: Table 19 shows the final ranking values for a selection of crucial constraints, after learning from simulated North Seto data. Similar to the Finnish results, both faithfulness constraints have risen to the top. As for Learner B, though the markedness constraints are correctly ordered relative to each other, the relative positions of the faithfulness vs the markedness constraints are not correct with respect to the crucial rankings proposed in Section 3.X:

$*F_4…\underline{B}_5$, $*F_4\underline{B}_5$, $*\underline{B}_5…F_4$, $*\underline{B}_5F_4$, Id(Bk)syl1 >> *B1 >> Id(Bk)

Constraint	Final ranking value
Id(Bk)σ_1 Id(Bk) $*F_4…\underline{B}_5$ $*\underline{B}_5…F_4$ $*F_4\underline{B}_5$ $*\underline{B}_5F_4$ … $*B_1$ …	118.000 114.220 112.000 112.000 110.000 110.000 … 108.020 …

Table 19. Excerpt of final ranking values for North Seto after simulation with Learner C.

As noted earlier in this section (with reference to Table 15), the average frequency of correct results is higher than for the grammar acquired by Learner B, even though the final ranking values show a constraint ordering that appears to be fully faithful. This difference is due to the spacing between constraints– in particular, Id(Bk) vs $*B_1$ and the relevant VH constraints. In the final grammar acquired by Learner C, all of the constraints are much closer together and therefore the stochastic evaluation is more likely to result in Id(Bk) swapping places with one or more of the markedness constraints, generating outputs that are more likely to obey markedness (vowel harmony and/or positional restrictions) rather than faithfulness pressures.

4.2.3. Simulation results - *a priori* and Favour Specificity

[Not sure if I should include this section or not, with results from a learner that has both of the specific-faithfulness biases applied. The results look pretty similar to what we get from the learners with either of the biases above, depending on the language. Is it useful to talk about *why* that's the case? I don't really think so… On the other hand, I would like to use both of them combined in the later sections of this chapter, since the combination often seems to produce better results (when also combined with whatever other bias) than just one or the other. Wondering if it's appropriate to include both of the spec-F biases as a base for those more complex simulations if I haven't shown the results here.]

To demonstrate the combined effects of both F_{spec} >> F_{gen} biases (*a priori* and Favour Specificity), I simulate acquisition of the three sample languages using Learner D, defined with the settings in Table 20.

Parameter	**Setting**
All basic parameters	Default
A priori bias	*d* = 20
Favour Specificity bias	Active

Table 20. Parameter settings for Learner D.

With both of these biases applied, learning simulations for all three sample languages fail to acquire the target grammars, producing similar results to those from Learners B and C. Test results are summarized in Table 21.

Language	**Average frequency of correct outputs**
Finnish	0.3367
North Estonian	0.9000
North Seto	0.2931
Overall	0.6222

Table 21. Summary of results from simulations with Learner D.

Finnish: Table 22 shows the final ranking values for a selection of crucial constraints, after learning from simulated Finnish data. The results are not meaningfully different from those of Learner B; the grammar acquired by Learner D is more or less fully faithful.

Constraint	Final ranking value
Id(Bk)σ_1	134.006
Id(Bk)	114.006
$*B_2$	112.002
$*F_3\ldots\underline{B}_5$	110.002
$*\underline{B}_5\ldots F_3$	108.000
$*\underline{B}_5F_3$	108.000
…	…
$*F_3\underline{B}_5$	106.000
…	…

Table 22. Excerpt of final ranking values for Finnish after simulation with Learner D.

North Estonian: The final ranking values for a selection of crucial constraints, after learning from simulated North Estonian inputs, are shown in Table 23. The results are not meaningfully different from those of Learner C; the grammar is still lacking a full ban against vowels from set B_1.

Constraint	Final ranking value
Id(Bk)σ_1	128.000
$*B_2$	116.000
$*F_5\ldots\underline{B}_2$	116.000
$*F_3$	112.000
…	…
$*F_5\underline{B}_2$	110.000
$*\underline{B}_5\ldots F_3$	108.000
…	…
$*B_1$	106.000
$*\underline{B}_5F_3$	104.000
…	…
Id(Bk)	80.000
…	…

Table 23. Excerpt of final ranking values for North Estonian after simulation with Learner D. A selection of no-disagreement constraints are included here for the purpose of comparison with results of simulations discussed in subsequent sections.

North Seto: Table 24 shows the final ranking values for a selection of crucial constraints, after learning from simulated North Seto data. The results are not meaningfully different from those of Learner B; the grammar acquired by Learner D is more or less fully faithful.

Constraint	**Final ranking value**
Id(Bk)σ_1	136.060
Id(Bk)	116.060
$*F_4\ldots\underline{B}_5$	116.000
$*\underline{B}_5\ldots F_4$	116.000
$*\underline{B}_5F_4$	114.000
$*F_4\underline{B}_5$	112.000
…	…
$*B_1$	102.000
…	…

Table 24. Excerpt of final ranking values for North Seto after simulation with Learner D.

4.2.4. Discussion

The application of the *a priori* bias enables the learner to produce grammars in which specific faithfulness constraints are ranked higher than their general counterparts. In the context of Balto-Finnic languages, such a bias facilitates the kind of first-syllable privilege that languages in this typology require– whether to ensure that neutralization only occurs in non-initial syllables (as for North Estonian) or to allow for harmony to be driven by the first syllable (as for Finnish and North Seto).

The Favour Specificity bias has a similar effect, but is more flexible, allowing specific faithfulness constraints to rise independently of their general counterparts. However, without a minimum required distance between the specific and general faithfulness constraints, there are situations in which it results in a more crowded sequence of final ranking values (for example, North Seto Learner C). This means a more variable final grammar, in a learning context where variability is not a desired characteristic of the target grammar.

Even with the degree of success shown by the North Estonian Learners B and C, it is clear that neither of these biases is enough for successful learning of the sample languages. A particular obstacle that recurs consistently throughout the simulations discussed throughout Section 4.2 is that most of the markedness constraints do not shift away from their initial values to any great degree (Figure 1). There are a number of reasons for this, which are discussed below, but the overarching consequence is that most of the relative rankings between the various markedness constraints do not have the opportunity to become anywhere near categorical. With constantly shifting markedness pressures and steadily rising faithfulness constraints, the learner cannot determine which markedness constraints to credit with any successful outputs and is only able to start selecting the intended winners as optimal once the faithfulness constraints have surpassed the chaos of the markedness constraints. After that point, since the learners receive only positive evidence, the faithfulness constraints continue to get credit for any winners, rendering the markedness constraints effectively useless. [TODO update this once I discuss the issue from Section 4.2.1.3 with AMT.]

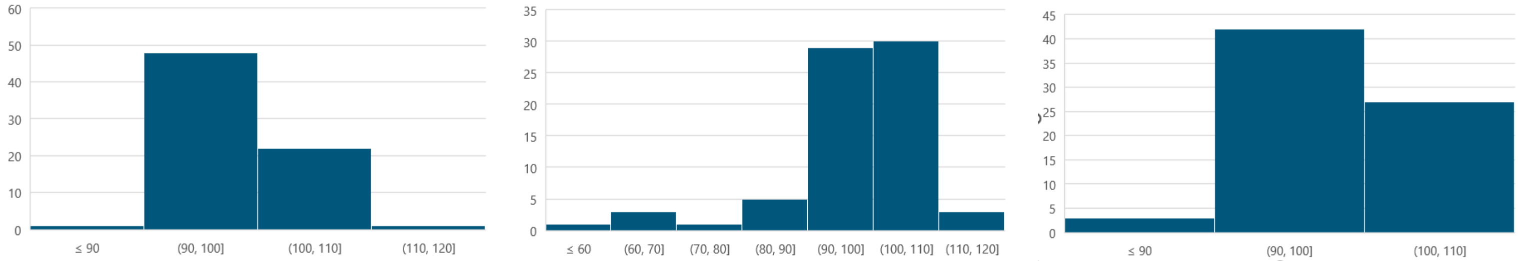

Figure 1. Distribution of final markedness constraint ranking values for Finnish, North Estonian, and North Seto Learners B. [These sad, sad plots need improving.]

Due to the nearsightedness of the GLA and the use of positive evidence only, markedness constraints that are never violated by the learning data (e.g., $*B_5…F_3$ in Finnish) are highly unlikely to ever be violated by a generated output; the only way this would happen is due to evaluation noise. Thus they have negligible opportunity to be promoted as a result of such an error, and ideally the other markedness constraints *fall* from their initial values. However, that turns out to be unlikely as well. For example, the symmetrical properties of $*B_5$ and $*F_5$ result in these two constraints 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 (2). The rest of the context-free scale referring constraints $*B_m$ and $*F_n$, while not perfectly symmetric, are antagonistic enough to result in approximately similar behaviour.

(2) Violation profile for a sample learning error with North Seto input.

/u..o/	$*B_5$	$*F_5$	…	$Id(Bk)\sigma_1$	Id(Bk)
✔ u..o	L** →		…		
☞ u..ø	*	W* ←	…		W* ←

As for the no-disagreement constraints, there is a slightly different issue at play. The vowel harmony constraints that are often violated and should be inactive in the target grammar do get demoted as errors are made in which they prefer the intended losers. But, given the complexity of this constraint set, there are other forces re-promoting these types of constraints. For example, consider $*F_5B_5$ in Finnish. Since /i/ and /e/, both in set F_5, are transparent in Finnish, such a constraint is violated by many of the learning inputs and therefore is demoted when it contributes to an error. On the other hand, it is also promoted often enough as a side effect of updates related to other errors, that much of the downward movement is cancelled out. (3a) and (3b) show examples of such errors, assuming faithfulness to the first syllable.

(3) ERC excerpts for errors “accidentally” promoting $*F_5B_5$ in Finnish.

	input	**candidates**	**$*F_5$**	**$*B_5$**	**$*F_3B_5$**	**$*F_5B_5$**

a.	/i..ø/	i..ø ~ i..o	L	W		W
b.	/æ..ø/	æ..ø ~ æ..o	L	W	W	W

In ERC (3a), both candidates are grammatical but only the intended winner is faithful to the input. Therefore when the learner selects the loser as optimal in order to avoid violating the currently-highly-ranked $*F_5$, the update promotes $*F_5\underline{B}_5$ even though it had nothing to do with the selection of the winner and is in fact very reasonable to violate in Finnish.

In ERC (3b), the learner again selects the loser in order to avoid violating $*F_5$. In this case, the resulting promotion of $*F_5\underline{B}_3$ is desired. But due to the no-disagreement constraints being built up from the nested stringency sets, the loser's violation (and resulting promotion) of $*F_5\underline{B}_3$ also necessarily means promotion of superset-referring $*F_5\underline{B}_5$.

Whether considering the context-free markedness constraints or the no-disagreement constraints, either way 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 need for more space between the (ideally) higher-ranked markedness constraints and the lower ones is hindered by their oscillation, but can be facilitated by defining a learner with asymmetry between its promotion vs demotion amounts. This adaptation - a modified update rule - is presented in Section 4.3.