Sources of (learning) evidence about transparent vowels in Finnic vowel harmony

Kaili Vesik and Anne-Michelle Tessier * *University of British Columbia*
MEDAL workshop - October 2025

am.tessier@ubc.ca
kvesik@gmail.com

L→R backness harmony in Finnic languages

Back harmony in West Votic	*kõɲvõ-ta* suit-3PL.PRS ɤ…ɤ…ɑ	*tanttsi-ma* dance-SPN ɑ…i…ɑ	*siga-lõõ* pig.SG-ALL i…ɑ…ɤ	*kirpittsa-ssa* brick.SG-ELA i…i…ɑ…ɑ
Front harmony in Kihnu Estonian	*müttä-gä* club.SG-COM y…æ…æ	*küüsi-mä* ask-SPN y…i…æ	*pidä-mä* must-SPN i…æ…æ	*kiibitse-mä* keep.at-SPN i…i…e…æ

Markus & Rozhanskiy (2022: 333); Lindström et al (2022)

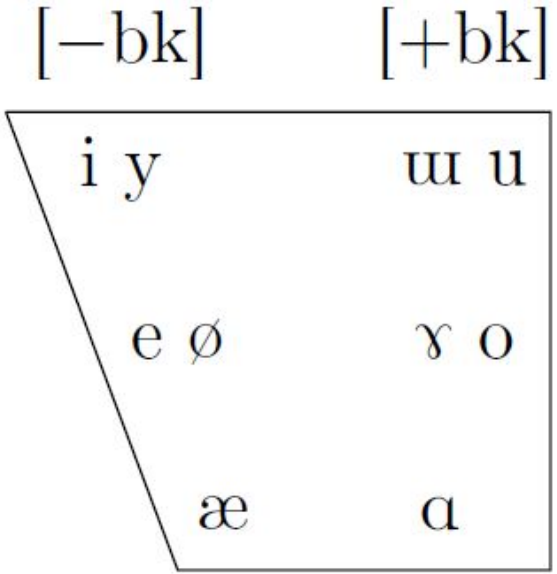
The roadmap

1) Data Bkgd: some established facts about Finnic vowel harmony
2) Theoretical Bkgd: grammars, learners, and problems of initial transparent vowels
3) New Data Results: corpus analysis of VH in five lesser-studied Finnic languages
    - How did we count, and what did we find?
    - How over/under represented are the key patterns?
    - Zooming in: how do alternating vs. fixed suffixes behave in longer words?
4) Interpreting our New Data Results
    - What does a grammar have to encode about initial transparent vowels in Finnic?
    - What kind of learner could notice these patterns?

Data Background: On Finnic VH and transparency

Vowels and phonological characteristics

- Main stress is on σ_1
- Full vowel contrast in σ_1
- No prefixes (σ_1 is always in the stem), but many potential suffixes

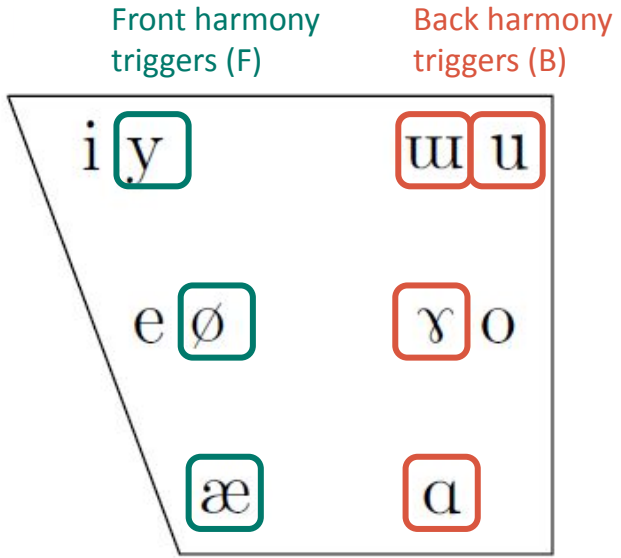


Vowel harmony

- Many (but not all) Finnic languages have VH, which is:
  - Progressive
  - Front/back
- Harmony extends through stem and (for the most part) into suffixes

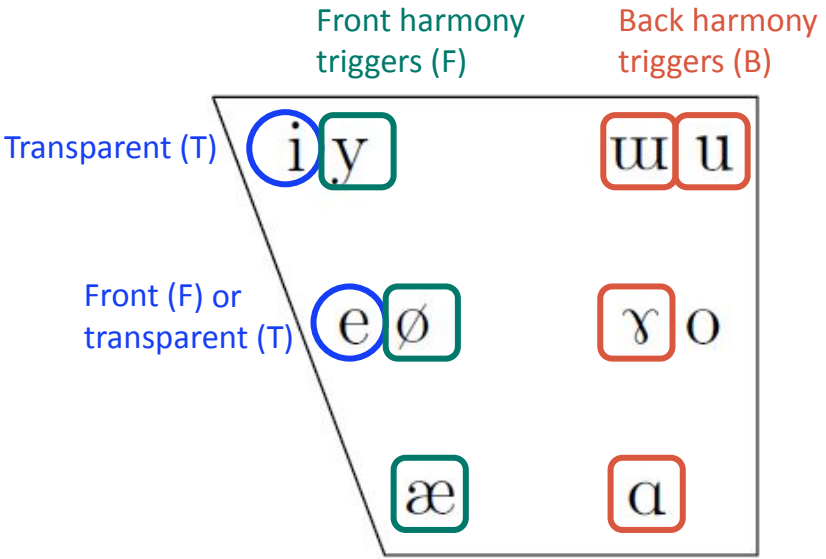
Vowel harmony

- Many (but not all) Finnic languages have VH, which is:
  - Progressive
  - Front/back
- Harmony extends through stem and (for the most part) into suffixes



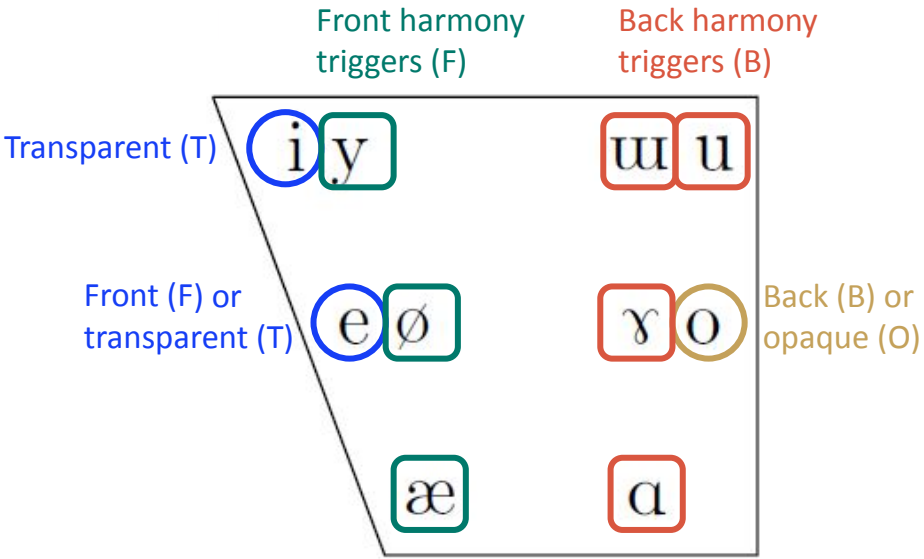
Vowel harmony

- Many (but not all) Finnic languages have VH, which is:
  - Progressive
  - Front/back
- Harmony extends through stem and (for the most part) into suffixes



Vowel harmony

- Many (but not all) Finnic languages have VH, which is:
  - Progressive
  - Front/back
- Harmony extends through stem and (for the most part) into suffixes



Vowel harmony

West Votic	Kihnu Estonian	South Seto
kõɲvõ-ta suit-3PL.PRS ɤ…ɤ…ɑ	*angu-ga* drift.SG-COM ɑ…u…ɑ	*umanõ* intrinsic u…ɑ…ɤ
tšülä village y…æ	*müttä-gä* club.SG-COM y…æ…æ	*päälüne* loft æ…y…e

Markus & Rozhanskiy (2022: 333); Lindström et al (2022); Pajusalu (2022: 369)

Vowel harmony

West Votic	Kihnu Estonian	South Seto
kõɲvõ-ta suit-3PL.PRS ɤ..ɤ..ɑ	*angu-ga* drift.SG-COM ɑ..u..ɑ	*umanõ* intrinsic u..ɑ..ɤ
tšülä village y..æ	*müttä-gä* club.SG-COM y..æ..æ	*päälüne* loft æ..y..e

Markus & Rozhanskiy (2022: 333); Lindström et al (2022); Pajusalu (2022: 369)

Transparency

West Votic	Kihnu Estonian	South Seto
tanttsi-ma dance-SPN ɑ…i…ɑ	*mõrsi-ga* string.bag.SG-COM ɤ…i…ɑ	*manitsõma* mention-SPN ɑ…i…ɤ…ɑ
läsi-mä be.sick-SPN æ…i…æ	*küüsi-mä* ask-SPN y…i…æ	*hämonõ* bleary æ…o…ɤ

Markus & Rozhanskiy (2022: 333); Lindström et al (2022); Pajusalu (2022: 369, 371)

Transparency

West Votic	Kihnu Estonian	South Seto
tanttsi-ma dance-SPN ɑ.ɑ.i.ɑ	*mõrsi-ga* string.bag.SG-COM ɤ.ɑ.i.ɑ	*manitsõma* mention-SPN ɑ.ɑ.i.ɑ.ɤ.ɑ
läsi-mä be.sick-SPN æ.æ.i.æ	*küüsi-mä* ask-SPN y.æ.i.æ	*hämonõ* bleary æ.o.ɑ.ɤ

Markus & Rozhanskiy (2022: 333); Lindström et al (2022); Pajusalu (2022: 369, 371)

More on transparency

- Consider Kihnu Estonian (more on this corpus data in a moment):
  - 10.84% of 3-syllable word types are B…T…B
    e.g. *olimõ* (to.be.1PL.PST), *juurikkad* (rhizome.PL.NOM)
- What if T is in σ1?
  - For some Finnic languages, VH is is determined by the vowel in the second syllable (e.g. Finnish; see Suomi et al, 2008)
  - What about others? T…B… is clearly *possible*; e.g. Kihnu *himustama* (desire.SPN)
- What if T is in both σ1 and σ2? That is, T…T…
  - No mention of how this affects harmony in the literature we reviewed.

Theoretical Background:
Driving Finnic VH from V1
How to learn?
And *what if V1 is* ***transparent***?

see e.g.: Bakovic, 2000; **Pulleyblank, 2002**; Kramer; 2003
Kiparsky and Pajusalu, 2003; **Vesik, 2025b**…

One theory of L→ R harmony

Why harmony?: Avoidance of V[α-bk] … V[β-bk]

		NO-DIS AGREEMENT		
/æ…ɑ/	[**æ**…**ɑ**]	*!		
	[æ…æ]			
	[ɑ…ɑ]			

see e.g.: Bakovic, 2000; **Pulleyblank, 2002**; Kramer; 2003
Kiparsky and Pajusalu, 2003; **Vesik, 2025b**…

One theory of L→ R harmony

Why harmony?: Avoidance of V[α-bk] … V[β-bk]
Why L→R harmony? Protection of privileged positions: stressed-V, root-V, initial-V…

		NO-DIS AGREEMENT	**IDENT [back]**	**IDENT[back]-ˈσ, IDENT[back]-V1**
/**æ**…ɑ/	[æ…ɑ]	*!		
	✓ [ˈ**æ**…æ]		*	
	[ˈ**ɑ**…ɑ]		*	*!

From the learner's perspective

Where it comes from?	***What they hear:***	*What to notice?*	
	"[æ…æ]"		
	"[ɑ…ɑ]"		

e.g. Hayes and Wilson (2008); Albright and Hayes, 2003; **Vesik (2025b)**...
c.f. Dresher (1999); Tesar and Smolensky (2000); Tessier (2009)...

From the learner's perspective

Where it comes from?	*What they hear:*	***What to notice?***	
/æ…ɑ/?	"[æ…æ]"	*æ…ɑ?	
/ɑ…æ/?	"[ɑ…ɑ]"	*[-bk]...[+bk]?	

e.g. Hayes and Wilson (2008); **Albright and Hayes, (2003)**; Vesik (2025b)...
C.f. Dresher (1999); Tesar and Smolensky (2000); Tessier (2009)...

From the learner's perspective

	What they hear:	*What to notice?*	***Relative strength of the generalization?***
	"[æ…æ]"	*æ…ɑ?	few ex., fewer exceptions
	"[ɑ…ɑ]"	*[-bk]...[+bk]?	
		…	
		*V[α-bk] … V[β-bk]?	many ex., many exceptions

e.g. van der Hulst (1985); Ringen and Heinämäki (1999); Välimaa-Blum (1999); Gafos (1999); Kiparsky and Pajusalu (2003), *interalia*

Complications from transparent vowels

Most common concern: ***How to ignore a word-medial T?***

	Surface Pattern	*What Harmony Sees:*
mõrsi-ga (KE) string.bag.SG-COM	ɤ..i..ɑ [bk]: [+ - +]	**ɤ**…i…**ɑ** [bk]: [+ +]
tanttsi-ma (WV) dance-SPN	ɑ..i..ɑ [bk]: [+ - +]	**ɑ**…i…**ɑ** [bk]: [+ +]

Complications from transparent vowels

Our concern: ***How to drive harmony after a word-initial T?***

		NO-DIS AGREEMENT	**IDENT [back]**	**IDENT[back]-ˈσ, IDENT[back]-V1**
/i…ɑ…æ/	i.**ɑ**…**ɑ**?		*	
	i.**æ**…**æ**?		*	
	i.**ɑ**…**æ**?	*		

Complications from transparent vowels

Our concern: ***How to drive harmony after a word-initial T?***

		What drives these choices?
/i…ɑ…æ/	i…**ɑ…ɑ**?	Spread L→R from first non-T vowel?
	i…**æ…æ**?	Default to V1 [-bk]?
	i…**ɑ…æ**	Give up? (Ruled out by NoDisagreement?)
/i…i…ɑ…æ/	???	

Complications from transparent vowels

Previous consensus: **How does harmony work after a word-initial T?**

/i…ɑ…æ/	**i…ɑ…ɑ**	**Spread L→R from first non-T vowel?**	
	i…æ…æ?		
	i…ɑ…æ		
/i…i…ɑ…æ/	???	*see eg Suomi et al (2008); Kiparsky (1982)....*	

Current Data Context:
Corpora from W/E Votic, N/S Seto, and Kihnu Estonian

Corpus of Estonian Dialects

- Spontaneous speech collected between 1938-2010 from speakers aged 42-101.
- Data identified by dialect group, dialect, and parish.
- Also includes data from Votic and Livonian.

Lindström et al (2022)

Selected subcorpora

Dialect	# wordforms	# single-root wordforms, native monophthongs only
Kihnu Estonian	21 595	16 385 tokens / 3 391 types
North Seto	21 127	18 894 tokens / 3 984 types
South Seto	27 416	24 817 tokens / 3 878 types
East Votic	9 572	7 728 tokens / 2 463 types
West Votic	24 744	19 967 tokens / 4 789 types

Corpus processing and coding

- (Native) monophthongs only (though cf. Vesik, 2023)
- 2, 3, and 4-syllable words
- Focusing on type frequency
- Predictions about pattern frequency are sensitive to vowel position
- Rates of occurrence are calculated out of the total number of n-syllable words

Corpus investigations of initial transparent vowels

Pt1: The extent of downstream VH

Guiding questions

- Is harmony driven from V2 if V1 is transparent?
- Is harmony driven from V3 if V1 AND V2 are transparent?

Predictions

IF *for example* /i…ɑ…æ/ →		THEN *we should see*
i…ɑ…ɑ	**/i/$_1$ is T, VH → from V2**	similar rates of Front, Back, and Disharmony in T_1 words as for V_1 harmonic words
i…æ…æ	**/i/$_1$ is F, VH → from V1**	no back harmony in T_1 words
i…ɑ…æ	**/i/$_1$ is T, no VH from V2**	rates of Front, Back, and Disharmony in T_1 words similar to chance

… and similar for words with T1 AND T2, moving along to V3

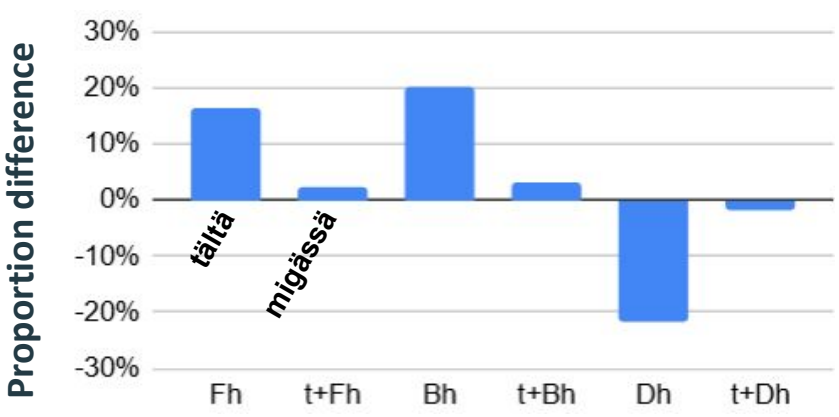
Testing predictions with West Votic

Compare actual vs predicted rates of front, back, and disharmony for:

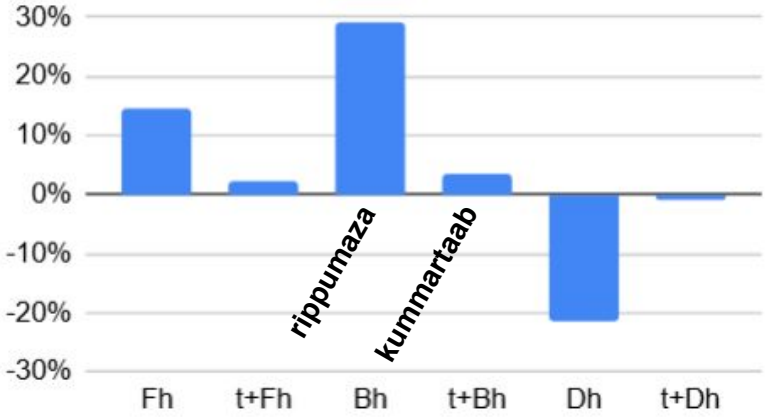
- T…V…V vs V…V words
  - e.g. *migässä* vs *tältä*
- T…V…V…V vs V…V…V words
  - e.g. *rippumaza* vs *kummartaab*
- T…T…V…(V) vs T…V…(V) words
  - e.g. *kirpittsassa* vs *viskaavad*

West Votic: rates of harmony after initial T* (actual - predicted)

(T)…V…V:



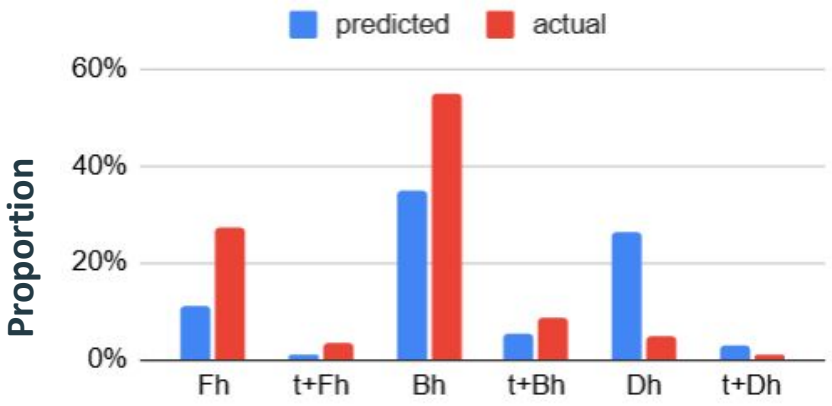
(T)…V…V…V:



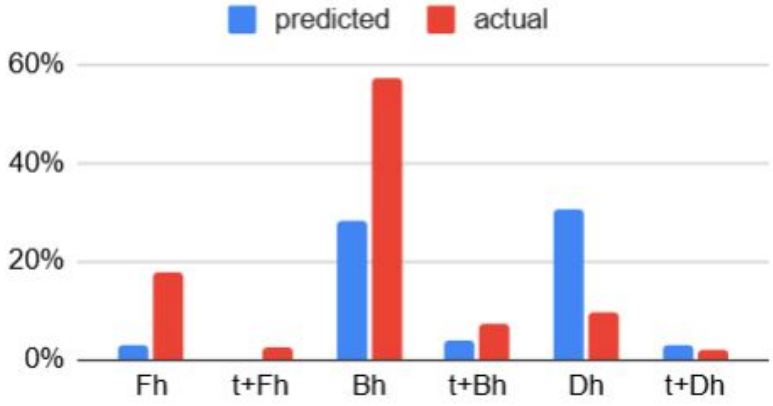
* T = /i/; /e/ is variable

West Votic: rates of harmony after initial T (direct comparison)

(T)…V…V

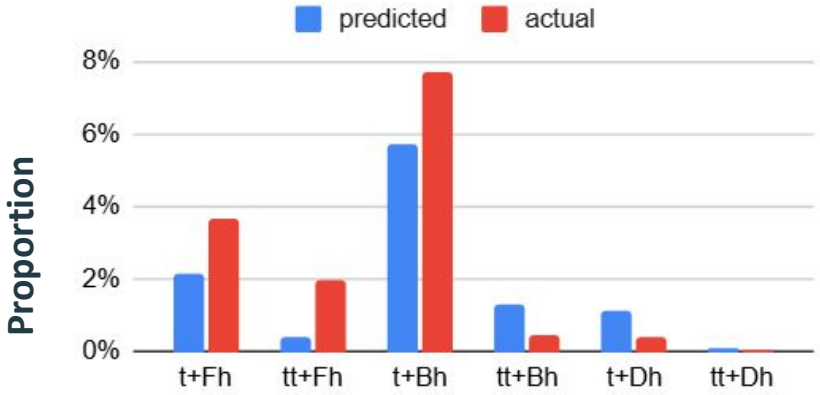


(T)…V…V…V



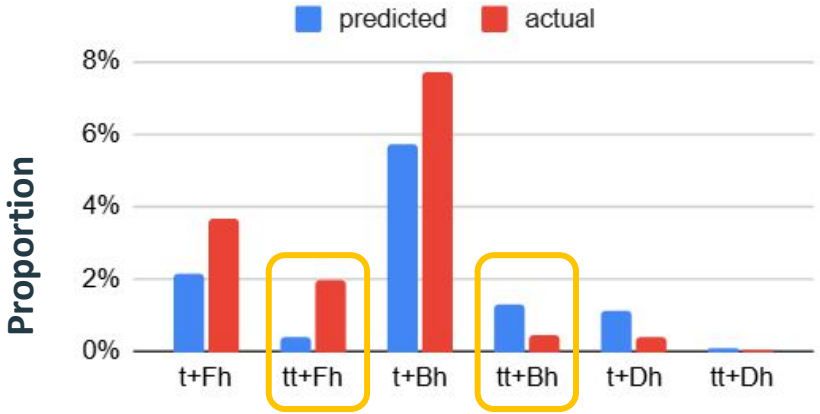
West Votic: rates of harmony after initial T..T (direct comparison)

(T)…T…V…(V) **proportions**



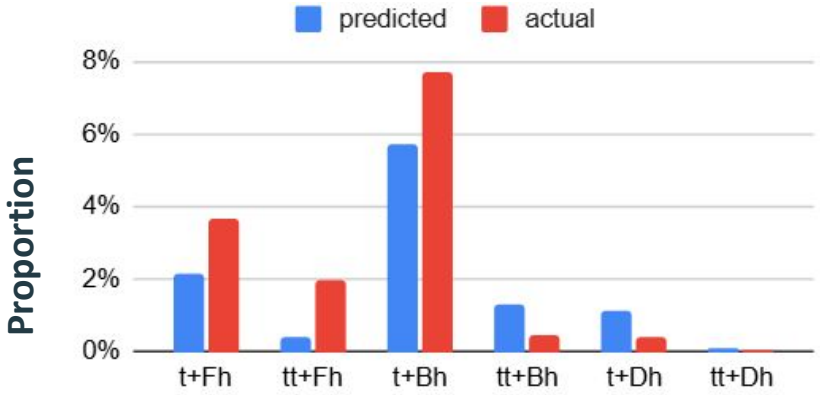
West Votic: rates of harmony after initial T..T (direct comparison)

(T)…T…V…(V) **proportions**

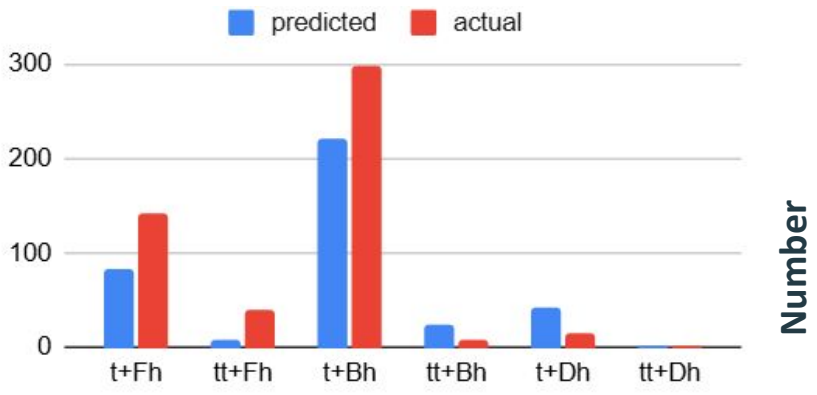


West Votic: rates of harmony after initial T..T (direct comparison)

(T)…T…V…(V) **proportions**



(T)…T…V…(V) **raw counts**



Summary of trends, after T…(T)

- T… words:
  - Rate of disharmony < predicted
  - Rate of front harmony > predicted

- T…T… words:
  - Rate of disharmony < predicted
  - Rate of front harmony > predicted

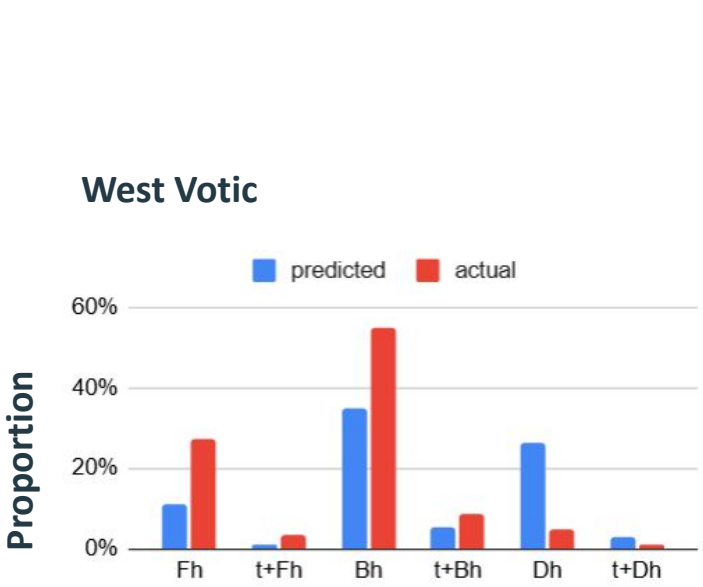
Summary of trends, after T…(T)

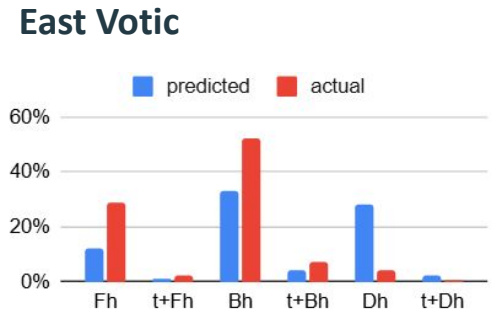
- T… words:
  - Rate of disharmony < predicted
  - Rate of front harmony > predicted
  - **Rate of back harmony > predicted**
  - *Rate of back harmony > rate of front harmony*
- T…T… words:
  - Rate of disharmony < predicted
  - Rate of front harmony > predicted
  - **Rate of back harmony < predicted**
  - *Rate of back harmony < rate of front harmony*

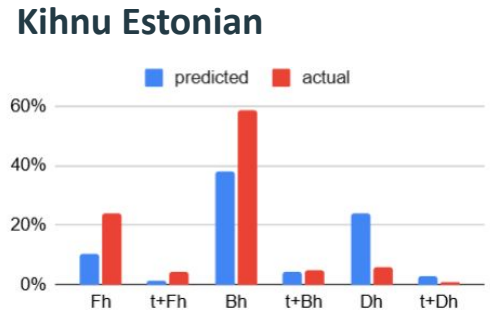
Checking predictions for West Votic

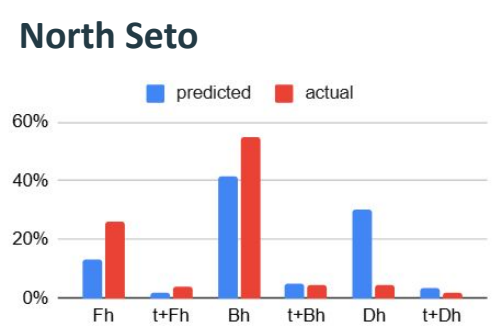
IF *for example* /i…ɑ…æ/ →		THEN *we should see*
i…ɑ…ɑ	**/i/$_1$ is T, VH → from V2**	similar rates of Front, Back, and Disharmony in T_1 words as for V_1 harmonic words 
i…æ…æ	**/i/$_1$ is F, VH → from V1**	no back harmony in T_1 words 
i…ɑ…æ	**/i/$_1$ is T, no VH from V2**	rates of Front, Back, and Disharmony in T_1 words similar to chance 

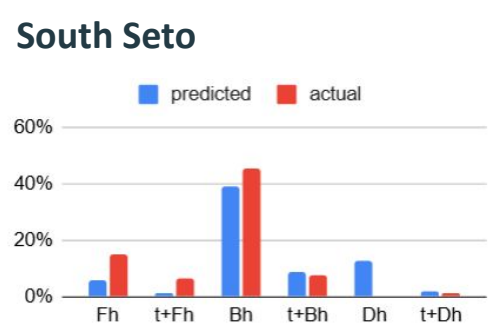
T…V…V vs V…V words in all five languages











Summary of trends, after T…(T)

True for all five languages, except…

- T… words:
  - Rate of disharmony < predicted
  - Rate of front harmony > predicted
  - Rate of back harmony > predicted
  - Rate of back harmony > rate of front harmony
- T…T… words:
  - Rate of disharmony < predicted
  - Rate of front harmony > predicted
  - Rate of back harmony < predicted
  - Rate of back harmony < rate of front harmony

> predicted only for longer words in North & South Seto

> predicted (barely) in Kihnu Estonian

Corpus investigations of initial transparent vowels
Pt2: A bit of the behaviour of suffixes

Morphology of T…T… words

- In what ways does the morphology of 3- and 4-syllable words beginning with T…T interact with vowel harmony?
- Is there any need for harmony to be driven from the third syllable?

Predictions

IF /i…i… V…V/ →	THEN *we should see*
/i/$_1$ & /i/$_2$ are T, VH → from V3	both T…T…B(h) and T…T…F(h) AND
	B-suffixes preferred for T..T..B stems, F-suffixes preferred for T..T..F stems
/i/$_1$ + /i/$_2$ are F(?), VH → from V1 or V2	F-suffixes preferred overall for T..T..(V) stems AND
	higher rates of Disharmony

West Votic: suffixes on T…T-initial words

TT+B(h)	TT+F(h)	TTBF	TTFB
nīvītta (4) this.way	*viimittää* (16) finally		*niittijee-kaa* (cf *-kää) (1) thread.PL-COM
lintii-kaa (cf *-kää) (3) ribbon.SG-COM	*niitti-väd* (15) shear-3PL.PRES		
kirpittsa-ssa (1) brick.SG-ELA	*pillittä-mā̄* (5) play-SPN		

Total of 1950 3- and 4-syllable words

West Votic: suffixes on T…T-initial words

TT+B(h)	TT+F(h)	TTBF	TTFB
nīvītta (4) this.way	*viimittää* (16) finally		*niittijee-kaa* (cf *-kää) (1) thread.PL-COM
lintii-kaa (cf *-kää) (3) ribbon.SG-COM	*niitti-väd* (15) shear-3PL.PRES		
kirpittsa-ssa (1) brick.SG-ELA	*pillittä-mā* (5) play-SPN		

Total of 1950 3- and 4-syllable words

West Votic: suffixes on T…T-initial words

TT+B(h)	TT+F(h)	TTBF	TTFB
nīvītta (4) this.way	*viimittää* (16) finally		*niittijee-kaa* (cf *-kää) (1) thread.PL-COM
lintii-kaa (cf *-kää) (3) ribbon.SG-COM	*niitti-väd* (15) shear-3PL.PRES		
kirpittsa-ssa (1) brick.SG-ELA	*pillittä-mā* (5) play-SPN		

Total of 1950 3- and 4-syllable words

Checking predictions for West Votic

IF /i…i… V…V/ →	THEN *we should see*	
/i/$_1$ & /i/$_2$ are T, VH → from V3	both T…T…B(h) and T…T…F(h) AND	?
	B-suffixes preferred for T..T..B stems, F-suffixes preferred for T..T..F stems	?
/i/$_1$ + /i/$_2$ are F(?), VH → from V1 or V2	F-suffixes preferred overall for T..T..(V) stems AND	✓
	higher rates of Disharmony	?

Interpreting our Results

How does the grammar/learner drive harmony after T1 (& T2)?

The problem: driving L→R harmony from V2?

		NO-DISAGREEMENT	**IDENT [back]**	**IDENT[back]-ˈσ, IDENT[back]-V1**
/i…ɑ…æ/	✓ i…**ɑ …ɑ** ?		*	
	i…æ…æ?		*	
	i…ɑ…æ?	*!		

Option 1:
Initial FOOT faith? (Kiparsky & Pajusalu, 2003)

		NO-DIS AGREEMENT	**IDENT [back]**	**IDENT[back]-ˈFt**	**IDENT[back]-ˈσ, IDENT[back]-V1**
/i…ɑ…æ/	✓ (ˈi…**ɑ) …ɑ**		*		
	(ˈi…æ)…æ		*	*!	
	i…ɑ…æ	*!			

Note: not trivial to rule out *(ɑ…æ) and *(æ…ɑ), with both Vs in the initial foot!

Upshot: Qs about T1 (+ T2) transparency

1) What kind of grammar can treat
   a) T… fully Transparent to downstream harmony AND
   b) T_T… as moderately Front-harmonic?

2) What kind of learner will get enough input evidence to learn *anything* about West Votic's T_T harmony?

   … and are humans that kind of learner?

Ongoing work and Future Directions

Thank you!

References

- Bakovic, E. (2000) *Harmony, Dominance and Control.* Doctoral dissertation, Rutgers University.
- Dresher, B. E. (1999). Charting the learning path: cues to parameter setting. *Linguistic Inquiry* 30(1): 27-67.
- Gafos, Adamantios. (1998) Eliminating long-distance consonantal spreading. *NLLT* 16: 223–78.
- Hayes, B. & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39(3): 379-440.
- van der Hulst, H. (1985). Vowel harmony in Hungarian: a comparison of segmental and autosegmental analyses. In H. van der Hulst & N., Smith (eds.) *Advances in non-linear phonology.* Dordrecht: Foris 267–303.
- Jesney, K. & Tessier, A.-M. (2011). Biases in Harmonic Grammar: the road to restrictive learning. *NLLT:* 29(1).
- Jurgec, P. (2011).*Feature Spreading 2.0 A Unified Theory of Assimilation*. Doctoral dissertation, University of Tromsø.
- Kiparsky, P. (1982). Lexical morphology and phonology. In *Linguistics in the morning calm: selected papers from SICOL-1981.*
- Kiparsky, P., & Pajusalu, K. (2003). Towards a typology of disharmony. *The Linguistic Review, 20*, 217-241.
- Kirchner, R. (1993). Turkish vowel harmony and disharmony: An Optimality Theoretic account. Ms, UCLA.
- Krämer, M. (2003). *Vowel harmony and Correspondence Theory.* Walter de Gruyer.
- Lindström, L., Todesk, T., & Pilvik, M.-L. (2022). *Corpus of Estonian dialects*. Institute of Estonian and General Linguistics, University of Tartu. Retrieved from https://datadoi.ee/handle/33/492

References

- Markus, E., & Rozhanskiy, F. (2022). Votic. In M. Bakró-Nagy, J. Laakso, & E. Skribnik (Eds.), *The Oxford Guide to the Uralic Languages* (p. 330-349). Oxford University Press.
- McCarthy, J. (2004). Headed Spans and Autosegmental Spreading. Ms, UMass Amherst.
- Pajusalu, K. (2022). Seto South Estonian. In M. B Bakró-Nagy, J. Laakso, & E. Skribnik (Eds.), *The Oxford Guide to the Uralic Languages* (p. 367-379). Oxford University Press.
- Pulleyblank, D. (2002). Harmony drivers: no disagreement allowed. *Proceedings of BLS 28: http://dx.doi.org/10.3765/bls.v28i1.3841*
- Tesar, B. & Smolensky, P. (2000). *Learnability in Optimality Theory.* MIT Press.
- Ringen, Catherine and Orvokki Heinämäki (1999). Variation in Finnish Vowel Harmony: An OT account. *NLLT* 17(2): 303-337.
- Suomi, K., Toivanen, J., & Ylitalo, R. (2008). *Finnish Sound Structure. Phonetics, phonology, phonotactics and prosody*. (Studia Humaniora Ouluensia 9). Oulu: University of Oulu.
- Välimaa-Blum, Riitta (1999). A feature geometric description of Finnish vowel harmony covering both loans and native words. Lingua 108(4): 247-268.
- Vesik, K. (2023). Vowel harmony in the Kihnu variety of Estonian: A corpus study. *Linguistica Uralica, 59* (3), 181-199.
- Vesik, K. (2025b). *Gradual error-driven learning and typology of Finnic vowel interactions* (Doctoral dissertation, University of British Columbia, Vancouver, BC). http://hdl.handle.net/2429/92155

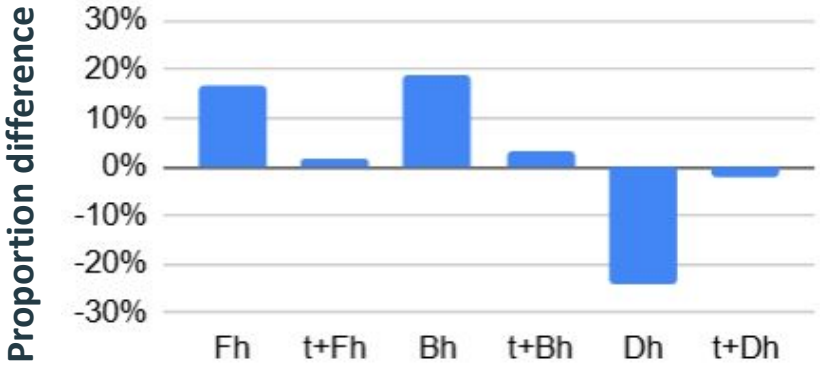
Appendix

Summary of Finnic vowel patterns

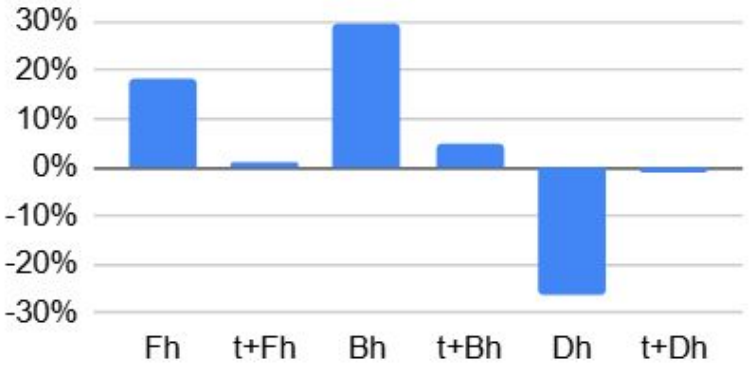
Languages	Inventory gaps	Restrictions outside of σ_1	Vowel harmony			
			Back	Front	Transp	Opaque
North Est	*{ɯ}	gaps + *{æ,ø,y,ɤ}	—	—	—	—
Livonian	*{y,ø}	gaps+ *{æ,ɯ,ɤ,o}	—	—	—	—
Finnish Karelian Ingrian	*{ɯ,ɤ}	—	ɑ,u,o	æ,y,ø	i,e	—
Votic Kihnu Est	*{ɯ}	—	ɤ,ɑ,u,o	(e),æ,y,ø	i,(e)	—
N Seto	—	gaps + *{ɯ}	ɯ,ɤ,ɑ,u,o	e,æ,y,ø	i	—
S Seto	—	gaps + *{ø,ɯ}	ɯ,ɤ,ɑ,u,(o)	e,æ,y,ø	i,(e)	o
Veps	*{ɯ,ɤ}	gaps + (*{æ,y,ø})	ɑ,u,o	æ,y,ø	i,e	—

East Votic: rates of harmony after initial T* (actual - predicted)

(T)…V…V



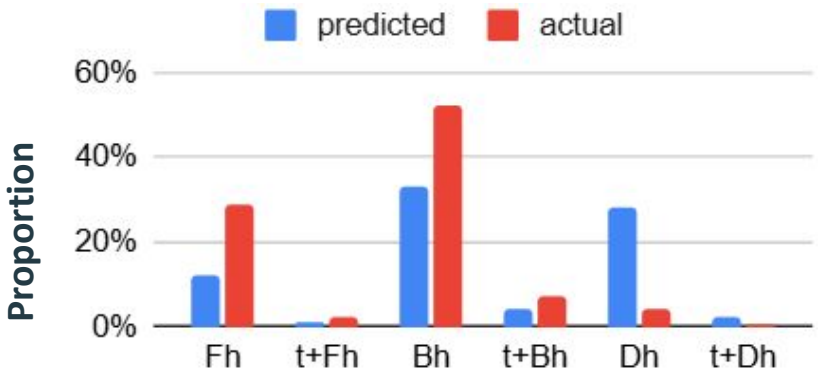
(T)…V…V…V



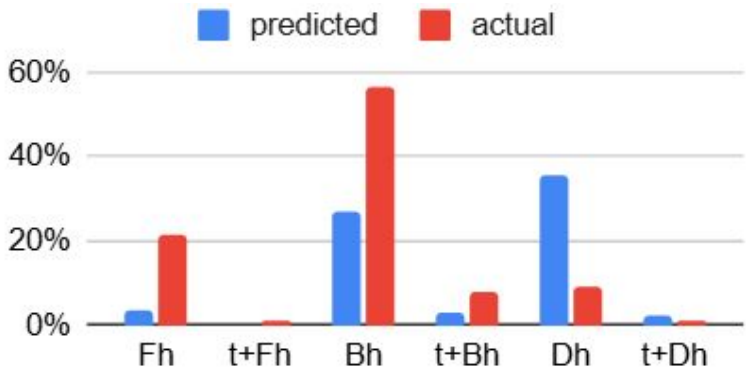
* T = /i/; /e/ is variable

East Votic: rates of harmony after initial T (direct comparison)

(T)…V…V

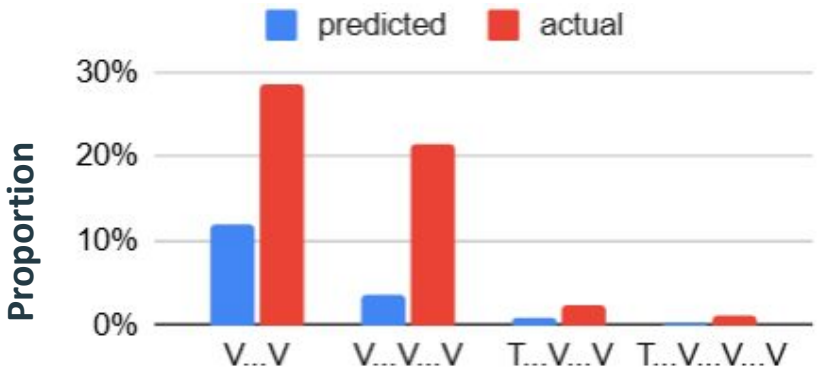


(T)…V…V…V

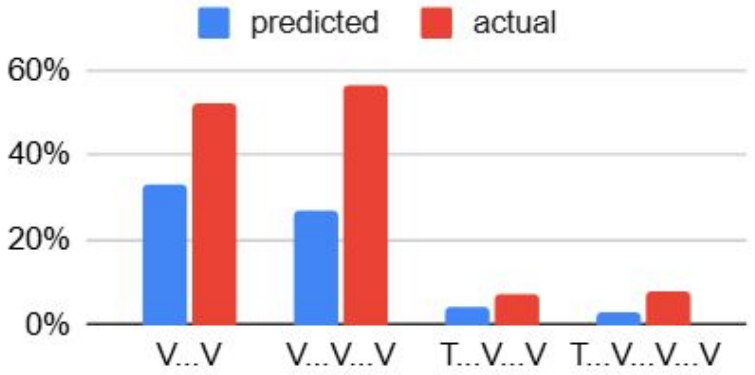


East Votic: rates of front and back harmony (direct comparison)

T+Fh sequences

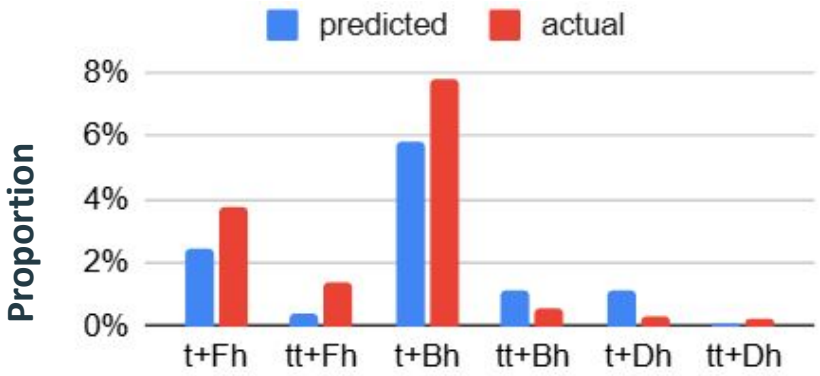


T+Bh sequences

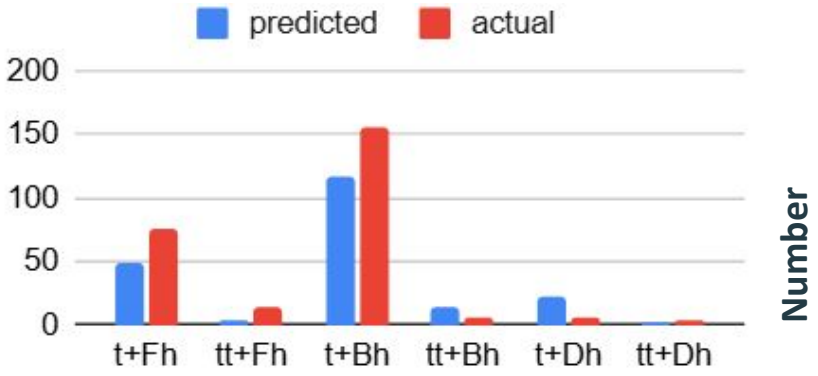


East Votic: rates of harmony after initial T..T (direct comparison)

T…(T)…V…V **proportions**

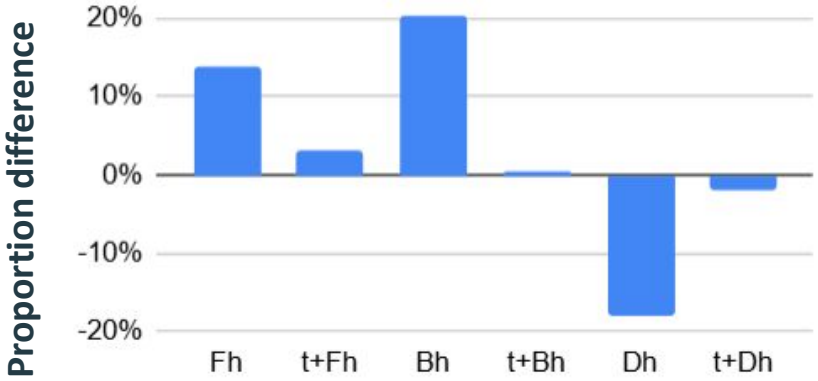


T…(T)…V…V **raw counts**

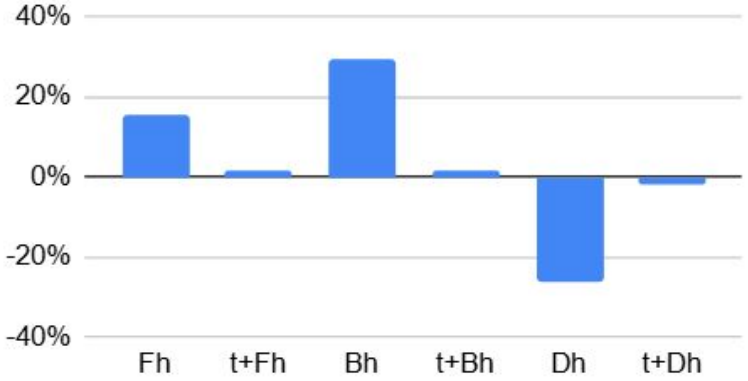


Kihnu Estonian: rates of harmony after initial T* (actual - predicted)

(T)…V…V



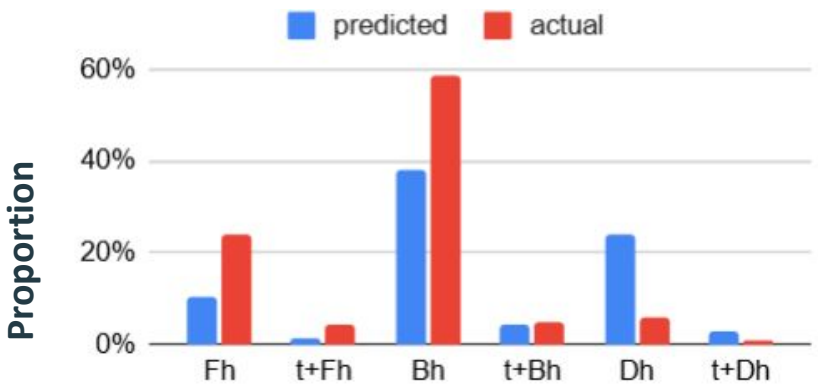
(T)…V…V…V



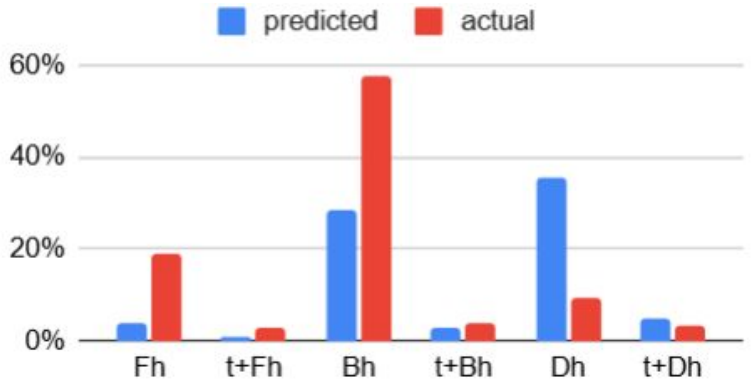
* T = /i/; /e/ is variable

Kihnu Estonian: rates of harmony after initial T (direct comparison)

(T)…V…V

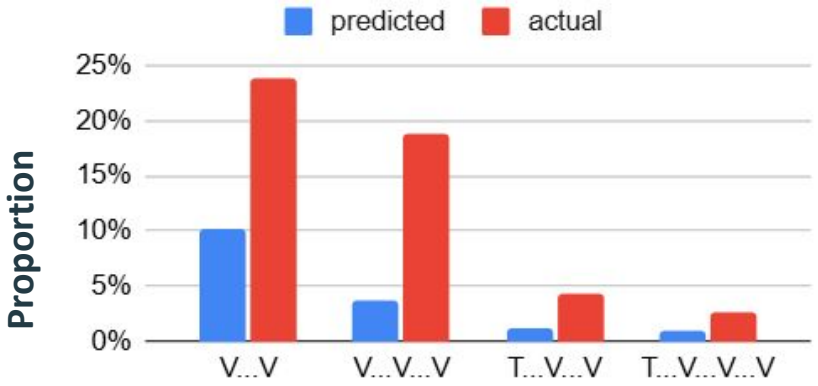


(T)…V…V…V

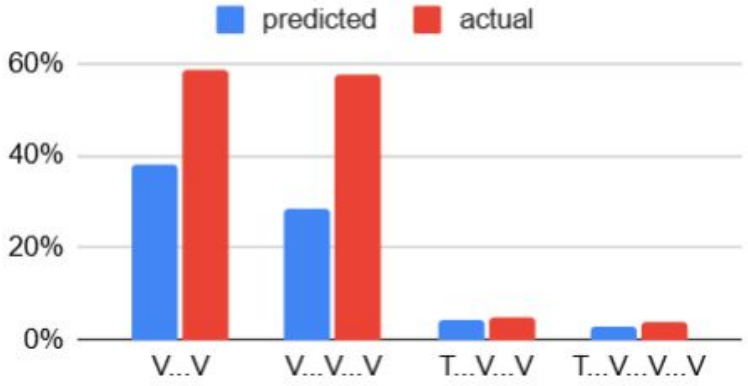


Kihnu Estonian: rates of front and back harmony (direct comparison)

T+Fh sequences

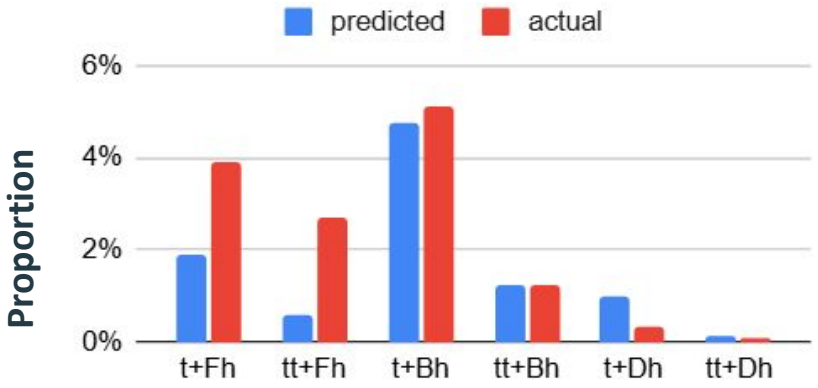


T+Bh sequences

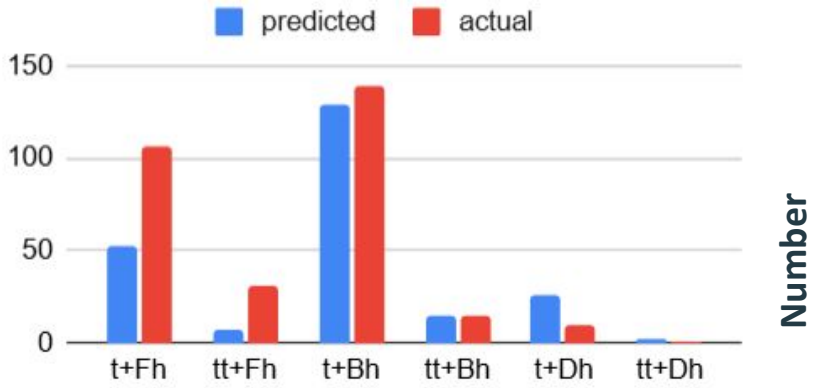


Kihnu Estonian: rates of harmony after initial T..T (direct comparison)

T…(T)…V…V **proportions**

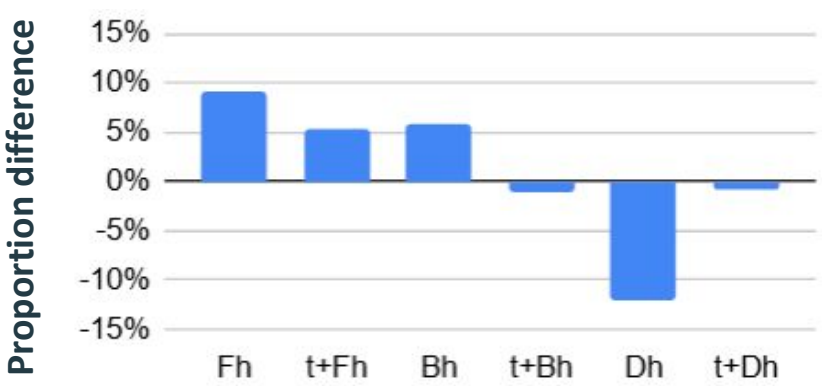


T…(T)…V…V **raw counts**

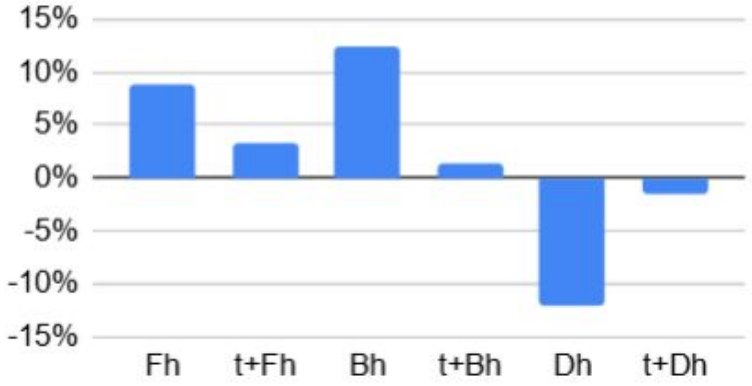


South Seto: rates of harmony after initial T (actual - predicted)

(T)…V…V

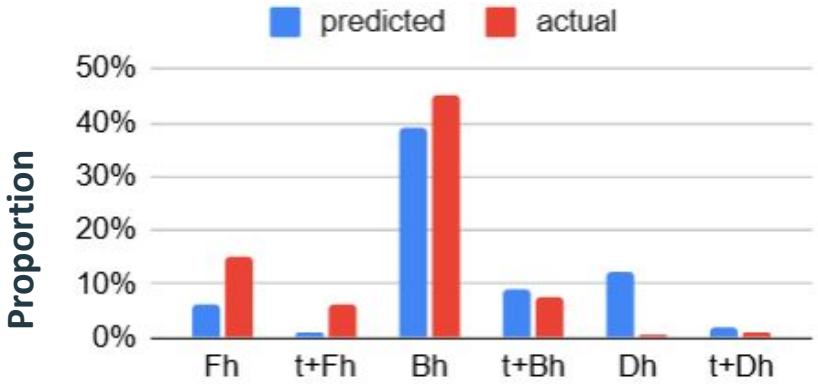


(T)…V…V…V

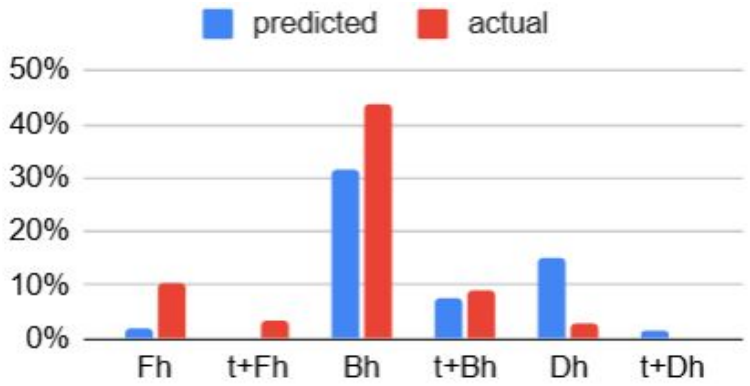


South Seto: rates of harmony after initial T (direct comparison)

(T)…V…V

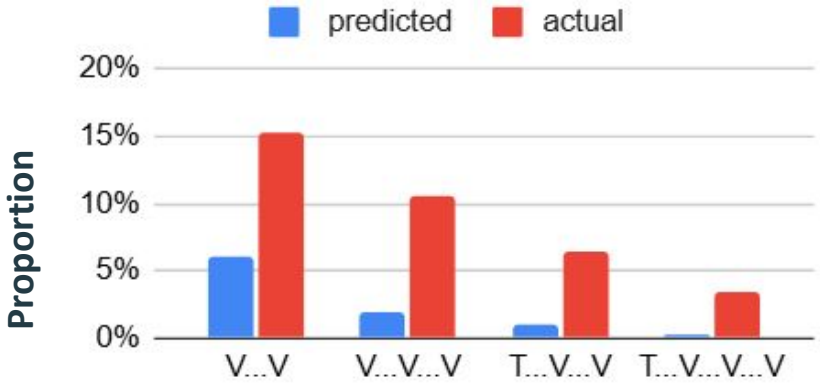


(T)…V…V…V

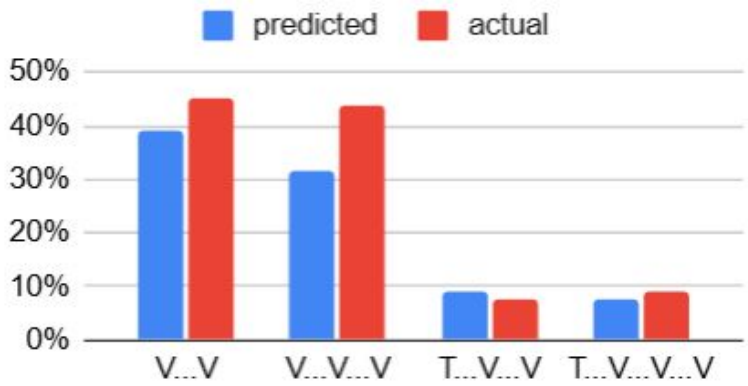


South Seto: rates of front and back harmony (direct comparison)

T+Fh sequences

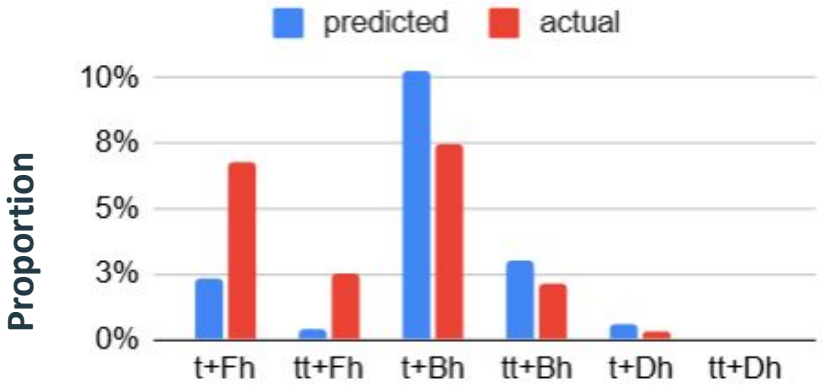


T+Bh sequences

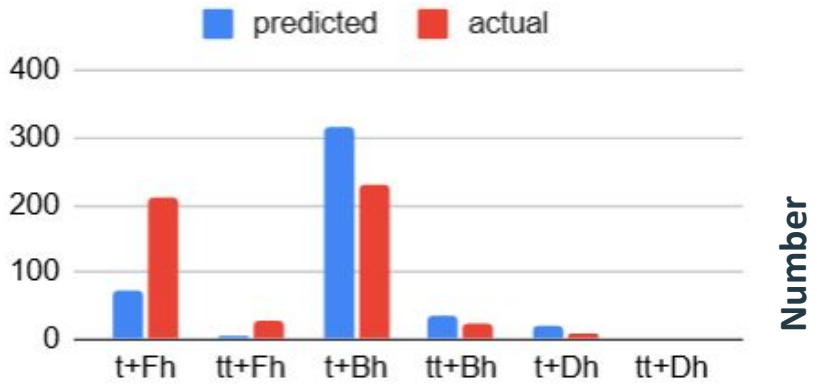


South Seto: rates of harmony after initial T..T (direct comparison)

T…(T)…V…V **proportions**

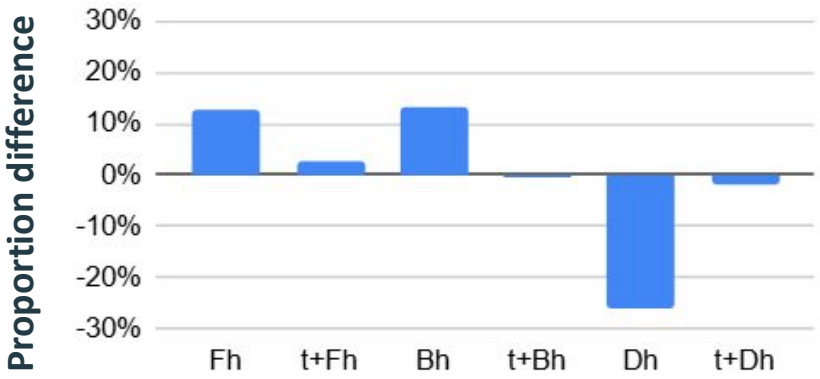


T…(T)…V…V **raw counts**

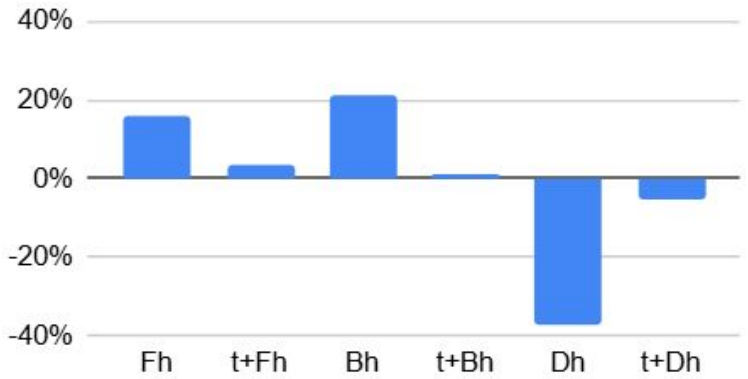


North Seto: rates of harmony after initial T (actual - predicted)

(T)…V…V

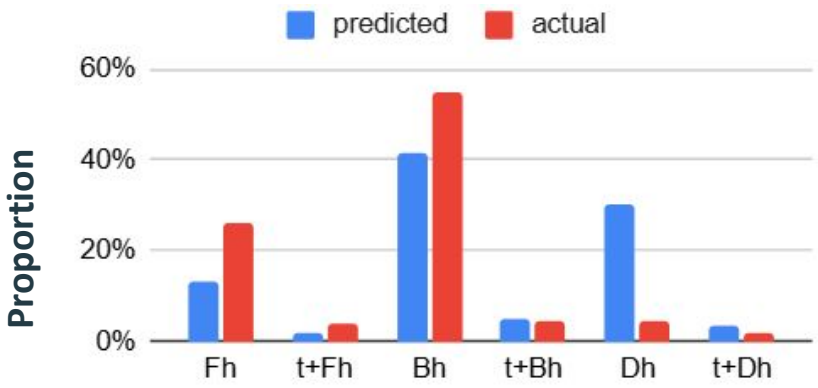


(T)…V…V…V

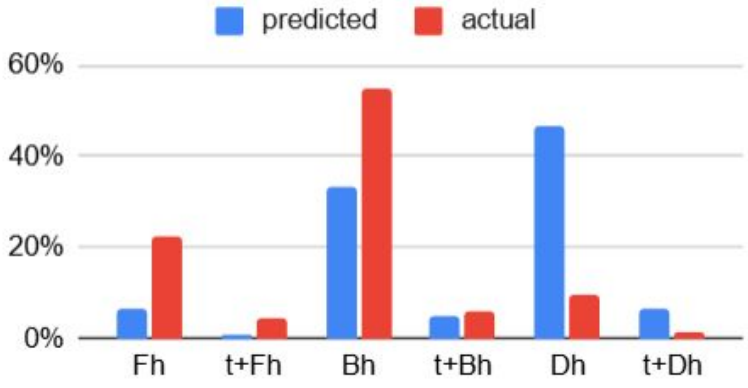


North Seto: rates of harmony after initial T (direct comparison)

(T)…V…V

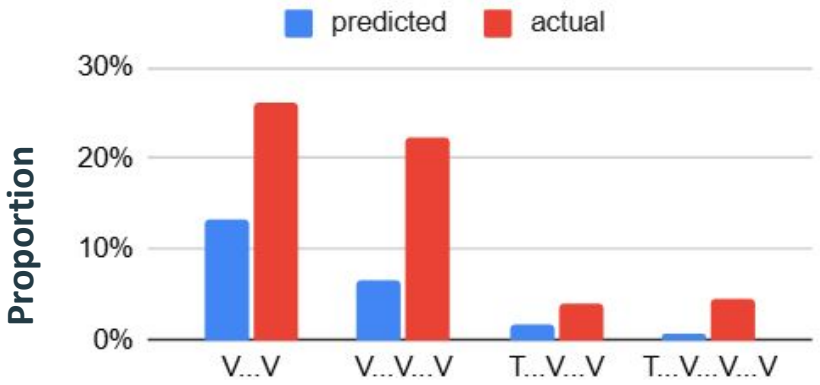


(T)…V…V…V

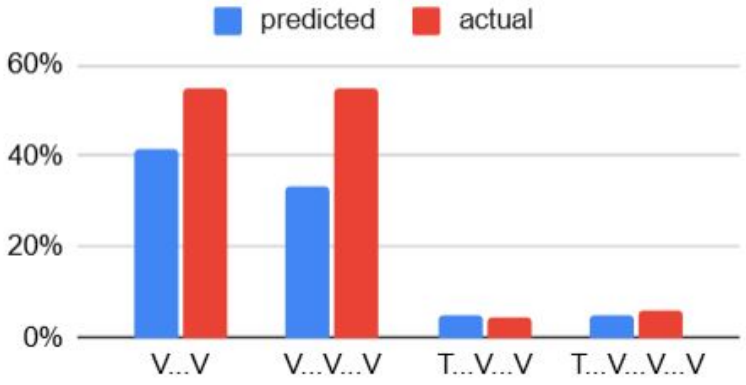


North Seto: rates of front and back harmony (direct comparison)

T+Fh sequences

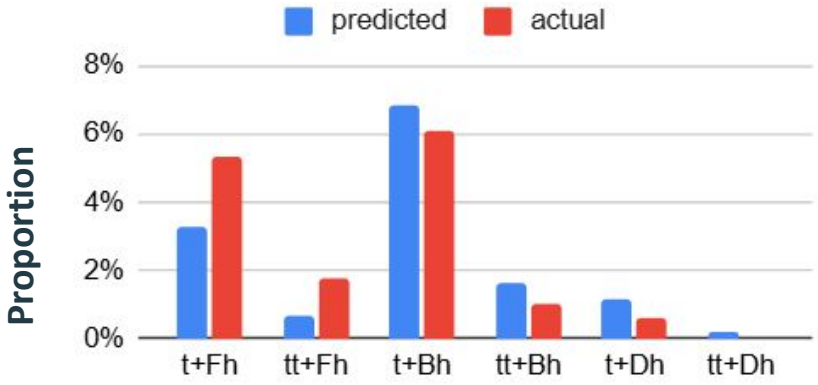


T+Bh sequences

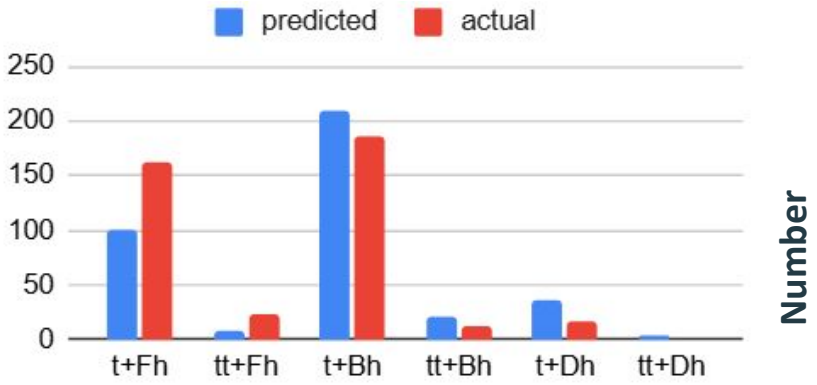


North Seto: rates of harmony after initial T..T (direct comparison)

T…(T)…V…V **proportions**



T…(T)…V…V **raw counts**



West Votic: suffixes on T…T-initial words

TT+B(h)	TT+F(h)	TTBF	TTFB
nīvītta (4) this.way	*viimittää* (16) finally		*niittijee-kaa* (cf *-kää) (1) thread.PL-COM
lintii-kaa (cf *-kää) (3) ribbon.SG-COM	*milti-sē* (9) what.kind.SG-ILL		
	niitti-väd (6)* shear-3PL.PRES		
kirpittsa-ssa (1) brick.SG-ELA	*pillittä-mā* (5) play-SPN		

INF. *niittää*

Total of 1950 3- and 4-syllable words