How phonologically determined is lexically specific phonology?*

Hannah Sande
Georgetown University
hannah.sande@georgetown.edu

January 2020

1 Introduction

Observation: There are two vowel alternations that apply to the same subset of roots in Guébie (iso:gie), a Kru language spoken in Côte d'Ivoire.

1. **Vowel deletion:** The initial vowel in CVCV roots can optionally be unpronounced.

   (1) /jɪla$^{3.2}$/, 'ask' → [jra^{32}], [jɪla$^{3.2}$]

   (2) /bala$^{3.3}$/, 'hit' → [bra^{3}], [bala$^{3.3}$]

   – Vowel deletion can only occur in 33% of CVCV roots in Guébie.

2. **Vowel harmony:** All features of root vowels are replaced with those of particular affixes:

   (3) /jɪla$^{3.2}$+ɔ^{2}/ → [jɔlɔ$^{3.2}$], *[jɪla-ɔ$^{3.3.2}$], *[jɪlɔ$^{3.2}$]
   ask+3SG.HUM.OBJ → 'ask him'

   – Full vowel harmony only applies to the same 33% of roots as deletion.

- While there are phonological tendencies which hold amongst alternating roots, there is no set of phonological features that picks out all and only the roots that alternate.

Goals:

1. Determine whether or to what extent the lexically specific deletion and harmony processes are phonologically determined.

2. Explore whether lexically naive phonological learning models can successfully learn the distribution of deletion and harmony in the Guébie lexicon.

Preview: I show that based on phonotactic traits of surface forms alone, we can accurately model the distribution of types of roots in the lexicon that alternate, but not the specific frequency with which a given root alternates.

Road map:

*Many thanks to the Guébie community. Thanks to Georgetown PhD student Bernie O'Connor for his help in organizing the data and running models related to this work. Thanks also to Sharon Inkelas, Larry Hyman, Peter Jenks, Dasha Kavitskaya, and audiences at the Leipzig University, WOCAL 9, and Georgetown University for comments on various aspects of this work. This work is funded by NSF-DEL grant number 1760302.

2 The data

Two phonological alternations in Guébie affect vowels in the same subset of words (4a-d):

1. Optional vowel deletion
2. Morphologically conditioned vowel harmony

Other roots fail to undergo both alternations (4e).

- In a corpus of 1840 roots, 617 of them (33.5%) alternate.
- Speakers are remarkably consistent (97% agreement) in their judgements of which roots can alternate.

(4) **Alternating roots**

	CVCV	*CCV*	*Gloss*	*Verb+3sg.obj*	*Gloss*
a.	bala$^{3.3}$	bra^{3}	'hit'	bɔl^{3}+ɔ^{2}, br+ɔ^{32}, *bal^{3}+ɔ^{2}	'hit him/her'
b.	jɪla$^{3.3}$	jra^{3}	'ask'	jɔl^{3}+ɔ^{2}, jr+ɔ^{32}, *jɪl^{2}+ɔ^{32}	'ask him/her'
c.	pɛja$^{3.1}$	pja^{31}	'buy'	pɔj^{3}+ɔ^{12}, pj+ɔ^{312}, *pɛj^{3}+ɔ^{12}	'buy him/her'
d.	tulu$^{4.4}$	tru^{4}	'chase'	tɔl^{4}+ɔ^{2}, tr+ɔ^{42}, *tul^{4}+ɔ^{2}	'chase him/her'
e.	ɟʊla$^{3.2}$	*ɟra^{32}	'take, borrow'	ɟʊl+ɔ$^{3.2}$, *ɟɔl+ɔ$^{3.2}$	'take him/her'

- This section briefly presents the deletion and harmony patterns.

2.1 Vowel deletion

Certain CVCV roots can surface as CCV in Guébie.

- Deletion seems to be optional, but is more common in fast, casual speech than in slow, careful speech. It is not dependent on morphosyntactic or phonological environment.

(5) **CVCV reduced to CCV**[1] (syl_20161207)

	CVCV	*CCV*	*Gloss*
a.	bala$^{3.3}$	bra^{3}	'hit'
b.	tulu$^{4.4}$	tru^{4}	'chase'
c.	wʊlʊ$^{3.3}$	wrʊ^{3}	'granary'
d.	munu$^{3.3}$	mnu^{3}	'bite/sting'
e.	mana$^{3.3}$	mna^{3}	'meat'
f.	jɪla$^{3.3}$	jra^{3}	'ask'
g.	sija$^{2.3}$	sja^{23}	'be defeated'
h.	kuɓə$^{3.1}$	kɓə^{31}	'yesterday'

Not all CVCV roots allow deletion of the initial V.

(6) **Non-alternating roots** (syl_20161207, syl_20170315)

	CVCV	*CCV*	*Gloss*
a.	ɟʊla$^{3.2}$	*ɟra^{32}	'take/borrow'
b.	tɛlɪ$^{3.3}$	*trɪ^{3}	'carve'
c.	sijo$^{2.3}$	*sjo^{23}	'wipe'
d.	ɲɛpɛ$^{3.1}$	*ɲpɛ^{31}	'sweep'

[1]Guébie has four tone heights, marked with numbers 1-4, where 4 is high. There is also an alternation between [l] and [r] in Guébie, where [r] is typically used in onset clusters (CCV), and [l] is used elsewhere. While all surface [l] and [r] consonants come from underlying /l/, I use [r] in clusters and [l] elsewhere to reflect production patterns.

The alternation between CVCV and CCV for a subset of roots is common across Kru languages (cf. Marchese (1979)), and is found in nearby Mande (Bearth 1971; Vydrine 2004, 2010; Green and Diakite 2008).

- In some Kru languages, where the first vowel is predictable given the second, CCV is claimed to be underlying, optionally produced CVCV in careful speech (Zogbo 2019, 9-10).
  - This /CCV/ → [CVCV] alternation could be seen as an instance of copy epenthesis (Dorsey's Law), where a vowel is inserted in a CCV word to break up the cluster (Miner and Dorsey 1979; Miner 1989; Hale and White Eagle 1980; Hayes 1995; Clements 1986, 1991; Halle et al. 2000; Kawahara 2007; Stanton and Zukoff 2018).
  - In copy epenthesis, the epenthetic vowel matches the quality of the following one:
    * Winnebago /*prás*/ → [*parás*] (Miner and Dorsey 1979, 27)
- In Guébie, unlike cases of copy epenthesis, the first vowel in a CVCV word is not predictable given its CCV counterpart.
  - *jɛla*$^{3.2}$, 'appear', and *jɪla*$^{3.2}$, ask', both surface as *jra*^{32} in their CCV form.
  - Given the surface form *jra*^{32}, the CVCV form is not predictable.
  - V1 and V2 are of identical quality in 329 of the 617 alternating CVCV roots.
  - Additionally, we do not find the same prosody on both syllables; each vowel in an alternating CVCV word can bear its own independent tone:
    * *jili*$^{2.3}$, 'steal'; *jɛla*$^{3.2}$, 'ask'
  - Only 270 of the 617 alternating roots have the same tone on both syllables.
- I assume that all surface CCV forms are underlyingly CVCV in Guébie, where the initial vowel can optionally be deleted (cf. Sande 2019).
- **Summary:** Initial vowel deletion applies optionally, only to a subset of lexical items.
- In other words, deletion is variable in multiple ways (cf. Linzen et al. 2013).
  - Deletion is *categorical* in that it does not involve partial reduction of a vowel (it is not *gradient*); the vowel is either present or not.
  - Deletion is *optional*.
  - Deletion is *lexically specific* in that it does not apply across the board to all lexical items equally.

2.2 Vowel harmony

The same CVCV roots that optionally surface as CCV show vowel harmony in certain morphosyntactic contexts.

- This process applies in the environment of object enclitics on verbs, and plural suffixes on nouns.
  - For brevity, I discuss only the object enclitic facts here.
- On monosyllabic roots, it is difficult to see the effect of vowel harmony, because all roots end in vowels, and when a vowel-initial suffix is added, the normal hiatus resolution process of the language is to delete to root-final vowel.
  - Ex: /ɓi^{2}-o^{2}/, *finish*-PASS, is pronounced [ɓo^{2}], *[ɓio$^{2.2}$].

- For this reason, I use disyllabic examples throughout.

- **Object enclitics on alternating verbs**
  - When an object enclitic is present on an alternating root,
    a. the final vowel of the root fails to surface (hiatus resolution),
    b. the initial vowel of a root surfaces with the same features as the object enclitic.

(7) **Object enclitics trigger vowel harmony on alternating roots**

	Bare verb	3SG.HUM =ɔ^{2}	3SG =ɛ^{2}	3PL =ɪ^{2}	*Gloss*
a.	jili$^{2.3}$	jɔl=ɔ$^{2.32}$	jɛl=ɛ$^{2.32}$	jɪl=ɪ$^{2.32}$	'steal'
b.	jɪla$^{3.3}$	jɔl=ɔ$^{3.2}$	jɛl=ɛ$^{3.2}$	jɪl=ɪ$^{3.2}$	'ask'
c.	bala$^{3.3}$	bɔl=ɔ$^{3.2}$	bɛl=ɛ$^{3.2}$	bɪl=ɪ$^{3.2}$	'hit'
d.	wʊla$^{3.1}$	wɔl=ɔ$^{3.12}$	wɛl=ɛ$^{3.12}$	wɪl=ɪ$^{3.12}$	'look at'

- Other phonologically identical suffixes do not trigger vowel harmony.

(8) **Passive suffix does not trigger vowel harmony**

	Verb	*Passive*	*Gloss*
a.	jili$^{2.3}$	jil-o$^{2.32}$, *jol-o$^{2.32}$	'be stolen'
b.	bala$^{3.3}$	bal-ɔ$^{3.2}$, *bɔl-ɔ$^{3.2}$	'be hit'
c.	jɪla$^{3.3}$	jɪl-ɔ$^{3.2}$, *jɪl-ɔ$^{3.2}$	'be asked'

- **Object enclitics on non-alternating verbs**
  - When an object enclitic is present on a non-alternating verb root,
    a. the final vowel of the root fails to surface (hiatus resolution),
    b. the initial root vowel retains its input features.

(9) **Non-replaceable roots in object contexts** (syl_20161207, syl_20170315)

	Root	*Root=ɔ^{2}*	*Gloss*
a.	sumu$^{2.2}$	sum=ɔ$^{2.2}$, *sɔmɔ$^{2.2}$	'boil him'
b.	ɟʊla$^{3.2}$	ɟʊl=ɔ$^{3.2.2}$, *ɟɔlɔ$^{3.2}$	'take him'
c.	tɛlɪ$^{3.3}$	tɛl^{3}=ɔ^{2}	'carve'
d.	sijo$^{2.3}$	sij^{2}=ɔ^{32}	'wipe'
e.	ɲɛpɛ$^{3.1}$	ɲɛp^{3}=ɔ^{12}	'sweep'

- **Summary:** Vowel harmony occurs in a subset of CVCV roots in object enclitic contexts (and plural contexts), but not elsewhere.
  - Harmony applies *categorically*, and never partially/gradiently.
  - Unlike deletion, harmony is not optional; alternating roots always show harmony in the relevant morphosyntactic contexts.
  - Harmony is *lexically specific*.

(10) **Variability in Guébie deletion and harmony**

	Gradient	*Optional*	*Lexically specific*
Deletion	–	✓	✓
Harmony	–	–	✓

3 Considering possible phonological factors

Across languages, some phonological alternations only apply to a subset of lexical items, or apply variably at different rates for different lexical items (cf. Itô and Mester 1995; Zuraw 2010; Smith 2011; Zymet 2018).

- **Question:** To what extent can alternating and non-alternating lexical items be distinguished based on phonology/phonotactics?

Given that both phenomena affect the first vowel of a CVCV root, we may expect the vowel quality of V1 to matter.

- High lax vowels are more prone to deletion than others.
- /o, e, a/ are least likely to be deleted.

(11) **Deletion given V1**

	i	ɪ	e	ɛ	ə	a	u	ʊ	o	ɔ
Alternating	86	20	15	70	24	89	98	22	32	85
Total	229	33	74	212	46	470	207	36	151	240
Percent	37.6	60.6	20.2	33	52.2	18.9	47.3	61.1	21.2	35.4
O/E	1.12	1.81	.60	.99	1.56	.56	1.41	1.82	.63	1.06

We also may expect the combination of V1 and V2 to predict alternation.

- In general, alternating roots are more common among those where V1 is the same as V2 (shaded cells), as discussed above.
  - For all other cells where we see more than 10% more alternating roots than the average (≥43%), we have fewer than 15 relevant data points.
- In a few cells, we see a drastically lower number than expected (ex: V1=o, V2=i).

(12) **Co-occurrence of V1 and V2**

	i	ɪ	e	ɛ	ə	a	u	ʊ	o	ɔ	Total
i	39/82 47.5	2/2 100	14/50 28	3/15 20	1/10 10	13/31 41.9	3/8 37.5	0/0 NA	16/41 39	3/15 20	86/229 37.6
ɪ	1/1 100	7/7 100	0/0 NA	3/10 30	0/0 NA	3/4 75	0/0 NA	0/2 0	4/5 80	4/6 66.7	20/33 60.6
e	1/7 14.3	0/3 0	11/24 45.8	2/8 25	1/5 20	4/11 36.3	0/11 0	0/1 0	0/11 0	1/7 14.3	15/74 20.2
ɛ	1/16 6.25	1/38 2.63	3/12 25	51/114 44.7	0/1 0	16/40 40	3/7 42.8	1/9 11.1	2/9 22.2	6/27 22.2	70/212 33
ə	1/10 10	1/1 100	5/13 38.4	1/3 33.3	19/26 73.1	3/6 50	0/4 0	0/0 NA	0/4 0	1/2 50	24/46 52.2
a	3/49 6.1	0/39 0	1/21 4.7	5/121 4.1	5/17 29.4	81/158 51.2	0/24 0	1/16 6.25	2/16 12.5	2/78 2.5	89/470 18.9
u	8/12 66.7	0/0 0	11/26 42.3	1/10 10	15/26 57.7	4/21 19.0	56/89 62.9	1/1 100	13/19 68.4	8/12 66.7	98/207 47.3
ʊ	4/9 44.4	6/9 66.7	0/2 0	3/7 42.8	0/0 0	2/3 66.7	1/1 100	12/13 92.3	1/1 100	4/9 44.4	22/36 61.1
o	0/45 0	0/6 0	1/29 3.4	3/12 25	3/11 27.3	2/19 10.5	10/25 40	1/1 100	23/54 42.6	5/15 33.3	32/151 21.2
ɔ	0/11 0	0/10 0	0/9 0	4/33 12.1	2/6 33.3	28/65 43.1	4/12 33.3	6/22 27.3	5/18 27.8	56/125 52	85/240 35.4
Total	57/272 20.9	16/90 17.8	36/140 25.7	66/290 22.7	40/89 44.9	157/360 43.6	73/168 43.4	16/51 31.4	47/137 34.3	76/246 30.9	

- We may also wonder whether the two vowels having the same backness or height features is relevant in determining alternation.
  - When V1=/ɔ/, it never alternates when followed by /i, ɪ, e/ and rarely before /ɛ/.
  - Similarly, when V1=/e/, it never alternates before /u, ʊ, o/, and rarely before /ɔ/.
- Adding constraints on backness and height to the models presented in section 4 does not improve results.
- I ran a logistic regression model coding for the following factors, to determine which factors are relevant in predicting whether a root is part of the alternating class:
  - V1, V2, C1, C2, Tone1, Tone2, V1/V2 interaction, C1/C2 interaction, T1/T2 interaction, V1/C1 interaction, V1=V2, T1=T2
  - The features in (13), V1=V2, T1=T2, and the identity of C2 (whether or not it is /l/), play the largest role in determining whether a root will alternate.
  - These factors will be built into the weighted constraint model of the Guébie facts in the following section.
- Alternating roots tend to, but do not always, show the following three phonotactic traits:
  - C2 (consonant) is /l/ ([n] in a nasal root)
  - V1 (vowel) and V2 are identical
  - T1 (tone) and T2 are identical
- The more of the above features a given root shows, the more likely it is to be alternating.

(13) **Factors influencing alternation**

	None	**T1=T2**	**C2=l**	**V1=V2**	**T&C2**	**T&V**	**C2&V**	**All 3**
Alternating	157	269	287	328	145	208	199	127
Total	751	614	536	611	244	339	244	154
Percent	20.9	43.8	53.5	53.7	59.4	61.4	81.6	82.5
O/E	.62	1.31	1.60	1.60	1.77	1.83	2.44	2.46

- Despite the phonotactic tendencies presented in this section, no phonological trait (or combination of traits) exhaustively and exclusively picks out alternating roots.
  - There are minimal pairs of roots where one alternates and the other does not.

(14) **Minimal pairs of alternating and non-alternating roots** (syl_20161207)

	CVCV	**CCV**	**Gloss**
a.	jili$^{2.2}$	jri^{2}	'be fat'
b.	jili$^{2.2}$	*jri^{2}	'fish'
c.	gɔlɔ$^{3.3}$	grɔ^{3}	'pain'
d.	gɔlɔ$^{2.3}$	*grɔ^{23}	'canoe'
e.	kpolo$^{3.1}$	kpro^{31}	'be clean'
f.	kpoke$^{2.4}$	*kpke^{24}	'crocodile'
g.	ɟulu$^{3.3}$	ɟru^{3}	'salt'
h.	ɟʊla$^{3.2}$	*ɟra^{32}	'take/borrow'

  - Thus, ultimately some information about subjectivity to alternation must be encoded lexically.
  - I confirm this conclusion with the results from learning models in the next section.

4 Analysis

To determine to what extent the alternating class of roots is phonologically predictable, I ran two sets of tests.

1. The MaxEnt Grammar Tool (Hayes and Wilson 2008) was used to determine whether the distribution of phonotactic types of words that alternate is predictable.
   – The model predictions were found to closely match the observed distribution of phonotactic word types that fall into the alternating class.
2. The MaxEnt Grammar Tool and the Gradual Learning Algorithm (Boersma and Hayes 2001) were used to determine whether novel individual lexical items could correctly be classified as alternating or not, based on a training set of data.
   – Neither model correctly predicts the output frequencies of individual novel roots.

- **Preview of findings:** I show that MaxEnt-HG can adequately model the distribution of phonotactic word types that alternate, but learning models have a harder time determining the rate at which any specific given lexical item will alternate. Since learning model predictions cannot mirror the data based on phonological information alone, I conclude that there must be some *lexical* difference between alternating and non-alternating roots.

4.1 Modeling the distribution of alternating roots in the lexicon

Can the distribution of alternating roots in the lexicon can be modeled phonologically?

- For the purposes of this talk, I assume the following four constraints.[2]

(15) **Reduce(T1=T2)**
Assign one violation if the tone on two consecutive syllables is identical (reduce if T1=T2).

(16) **Reduce(C2=l)**
Assign one violation if a vowel intervenes between [l] and a preceding consonant (reduce if C2=l).

(17) **Reduce(V1=V2)**
Assign one violation if vowels in two consecutive syllables are identical (reduce if V1=V2).

(18) **Max**
Assign one violation for every input segment that lacks a corresponding output segment.

– A candidate violates one of the Reduce constraints if it has not deleted V1, and shows the specified surface property, T1=T2, C2=l, or V1=V2.
– All candidates that are reduced, CCV, violate Max, because the initial input vowel fails to surface.
– A similar set of constraints in a Harmonize series account for the harmony alternations.

- These constraints, with weights determined by the MaxEnt Grammar Tool (Hayes et al. 2009), provide a model of the distribution of alternation in the Guébie lexicon.

[2]Other constraints such as Max specific to certain vowel qualities and constraints penalizing different height, backness, and rounding values of vowels within the same word were considered but did not improve the model.

- In this model, candidates are groups of roots that share the same subset of the three phonotactic properties in (13).
- In the MaxEnt model below, the predicted values quite closely mirror the observed amount of alternation for each word type.
- The fact that the predicted amount of reduction for each type of root in the MaxEnt analysis in (19) so closely mirrors the observed pattern supports the analysis of the proposed parameters (T1=T2, C2=l, V1=V2) as those most relevant in determining whether a given root is reducible.

(19) **MaxEnt HG weights: Vowel deletion**

	Reduce(T)	Reduce(C2)	Reduce(V)	Max			
	.662	1.02	1.23	1.15	**H**	**Obs (%)**	**Pred (%)**
T1=T2							
CVCV	1				.662	57.2	61.9
CCV				1	1.15	43.8	38.1
C2=l							
CVCV		1			1.02	46.5	53.1
CCV				1	1.15	53.5	46.9
V1=V2							
CVCV			1		1.23	46.3	48.0
CCV				1	1.15	53.7	52.0
T, C2							
CVCV	1	1			1.682	40.6	36.9
CCV				1	1.15	59.4	63.1
T, V							
CVCV	1		1		1.892	38.6	32.3
CCV				1	1.15	61.4	67.7
C2, V							
CVCV		1	1		2.25	18.4	24.9
CCV				1	1.15	81.6	75.1
T, C2, V							
CVCV	1	1	1		2.912	17.5	14.6
CCV				1	1.15	82.5	85.4
None							
CVCV					0	79.1	75.9
CCV				1	1.15	20.9	24.1

- Note that this model *cannot* be used to make predictions about how a particular root will surface.
  - 61.9% of /CVCV/ roots with the same tone on both syllables will categorically, always surface with harmony in the relevant contexts, and can always optionally surface with deletion.
  - Nothing about this model predicts which T1=T2 roots will alternate and which will not, or the frequency with which a specific alternating root will surface as CCV.

4.2 Modelling the behavior of specific lexical items

Recall that no set of phonotactic traits picks out all and only the alternating roots, and there are (near) minimal pairs where one alternates and another does not.

- This section shows why a learning model cannot make good predictions about the frequency with which a given root will show deletion or harmony, relying purely on phonological information.

- Constraint-based models can be trained on a subset of data points to determine relevant constraint weights/rankings, and their success can then be tested on the remaining data.
  - MaxEnt-HG (using the MaxEnt Grammar Tool)
  - Gradual Learning Algorithm (using Stochastic OT in OT Soft)

(20) **Mini dataset**

	CVCV		*CCV*		*Gloss*
a.	jili$^{2.2}$	50%	jri^{2}	50%	'be fat'
b.	jili$^{2.2}$	100%	*jri^{2}	0%	'fish'
c.	gɔlɔ$^{3.3}$	50%	grɔ^{3}	50%	'pain'
d.	gɔlɔ$^{2.3}$	100%	*grɔ^{23}	0%	'canoe'
e.	kpolo$^{3.1}$	50%	kpro^{31}	50%	'be clean'
f.	kpoke$^{2.4}$	100%	*kpke^{24}	0%	'crocodile'
g.	ɟulu$^{3.3}$	50%	ɟru^{3}	50%	'salt'
h.	ɟʊla$^{3.2}$	100%	*ɟra^{32}	0%	'take/borrow'

- The roots in (20a,b) will always incur the same violations, since they are phonologically identical. However, their distributions of output forms differ.
- If we were to train a model using the data in (20), the model would learn that roots like /jili$^{2.2}$/ show deletion 25% of the time.
  - It predicts that both roots /jili$^{2.2}$/ in the input training data will surface as CCV 25% of the time and as CVCV 75% of the time.
  - It also predicts that if there was a novel root /jili$^{2.2}$/, or any novel CVCV root where V1=V2, T1=T2, and C2=L, it would surface as CCV 25% of the time and as CVCV 75% of the time.[3]
- The output distribution predicted by the model does not match that of either /jili$^{2.2}$/ input form.
- A lexically naive learning model will never be able to account for such differences, and instead will make bad predictions about the output probabilities of all roots.
- I trained two different learning models (using the MaxEnt Grammar Tool and GLA) on half of the Guébie CVCV data.
  - Including additional factors as constraints in the model did not improve it (tested on 8 different possible grammars, each including a different subset of constraints based on the possible contributing phonological factors discussed in section 3).
  - There was a mean difference of more than 20% between observed and predicted probabilities of output candidates in all 16 (8 grammars x 2 learning models) models.
- Perhaps unsurprisingly, MaxEnt and other learning models cannot capture phonologically unpredictable variation without building in lexically-specific constraints, or weights sensitive to lexical item.

[3] I ran a psycholinguistic experiment in the Guébie community in Summer 2019 to test this prediction. I worked with 22 participants, who were each introduced to 44 nonce singular words and asked for the plural form. 9 of the participants' data were usable. Of the 9, 4 never extended alternation to nonce words, suggesting that alternation is lexical. 3 extended to 1/44 words, 1 to 2/44 words, and 1 to 4/44 words. Of the 9 total words across all 9 participants that showed an alternation (2% of the data), 8 had V1=V2 and 7 had C2=L. Two of the words were produced as alternating by more than one speaker: jɔlɔ (2), wɛlɛ (3). There is too little data on any given word type to run stats, but it seems that some speakers extend alternation variably but minimally to nonce words, suggesting a small amount of phonological determinedness, while other speakers do not extend, suggesting lexical determinedness.

4.3 Lexically specific phonology in Cophonologies by Phase

This section briefly recaps an analysis of the Guébie facts sensitive to the lexical and morphological context, set in Cophonologies by Phase (Sande 2019).

- See Linzen et al. (2013) for a similar analysis of lexically specific scaling factors of constraint weights in Russian lexically specific phonology.

CBP background

- Phonology is spelled out at syntactic phase boundaries.
- Phonological evaluation involves weighted constraints.
- There is a default weighted constraint grammar for a language.
- Individual vocabulary items (morphemes) may be associated with a constraint weight adjustment, which is added to the default weights, only in the syntactic phase containing that morpheme.

(21) *Relevant constraints*

- IDENT-IO(V): Assign one violation if an output vowel's features differ from the corresponding input segment.
- $*[\alpha F][\beta F]_{[+syllabic]}$ (Abbreviated VHARM(ONY))
  A segment with a given set of feature values may not directly precede another segment with a different set of feature values in the ordered set of output segments that are [+syllabic]. Assign one violation for each output form where at least one pair of vowels meets these criteria.
- *HIATUS: Assign one violation for every pair of two consecutive vowels in the output.

(22) *Default weights for suffix-triggered harmony*

Constraint	Weight
IDENT-V	12
VHARM	.5
*HIATUS	9

- In the presence of an object marker or plural suffix, constraint weights are adjusted (the weights in $\mathcal{R}$ are added to the default constraint weights, only in the syntactic phase domain containing the relevant vocabulary item).

(23) *Object marker vocabulary item*

$$[3sg.hum.acc] \longleftrightarrow \left\{ \begin{array}{ll} \mathcal{F}: & /\text{ɔ}^2/ \\ \mathcal{P}: & [= X]_\omega \\ \mathcal{R}: & \text{VHARM}^{+6.5}, \text{IDENT-V}^{-4} \end{array} \right\}$$

- On its own, this weight adjustment does not affect the output; the faithful, non-harmonizing output is still predicted to be most frequent, accounting for why non-alternating roots do not alternate.
- Alternating roots are associated with a particular n head, which is also associated with a constraint weight readjustment.

(24) *Alternating n vocabulary item*

$$[n_{alternating}] \longleftrightarrow \left\{ \begin{array}{ll} \mathcal{F}: & \emptyset \\ \mathcal{P}: & [X_\omega] \\ \mathcal{R}: & \text{VHARMONY}^{+2.5}, \text{IDENT-V}^{-6} \end{array} \right\}$$

- When both an alternating n and object marker or plural suffix are present in the same phase domain, the effects of their $\mathcal{R}$ adjustments combine.

(25) *Cumulative effects of morpheme-specific cophonologies*

Grammar	IDENT-IO(V)	VHARM
Default	12	.5
Obj/Pl	-4	+6
Alt. root	-6	+2.5
Total weight	2	9

(26) *Alternating root + object enclitic: Harmony*

[$_\omega$ bala$^{3.3}$]=ɔ^2	VHARMONY 8	IDENT-V 2	*HIATUS 9	**H**	**Obj**	**Pred**
a. [$_\omega$ bal^3=ɔ^2]	1			8	0	0
b. ☞[$_\omega$ bɔl^3=ɔ^2]		1		2	1	1
c. [$_\omega$ bala$^{3.3}$=ɔ^2]	1		1	17	0	0
d. [$_\omega$ bɔlɔ$^{3.3}$=ɔ^2]		2	1	13	0	0
e. [$_\omega$ balɔ$^{3.3}$=ɔ^2]	1	1	1	19	0	0

Sande (2019) shows that this is a doubly morphologically conditioned phonological alternation.

- The combined effect of two subweightings results in full vowel harmony only when both of the following are present:
  1. A plural suffix or object enclitic
  2. An alternating root

5 Implications and conclusions

33% of Guébie CVCV roots undergo harmony in certain morphosyntactic contexts, and can optionally surface as CCV.

- Phonological factors alone cannot categorically determine alternating from non-alternating roots.
- Guébie speakers, on the other hand, agree 97% of the time on which roots can alternate.

This discrepancy (and initial experimental results discussed in footnote 3) suggests that roots are lexically specified as alternating or not.

- I model this lexical specification using morpheme-specific constraint weight adjustments associated with vocabulary items in Cophonologies by Phase (CBP) (Sande and Jenks 2018; Sande 2019).
- CBP allows for doubly morphologically conditioned phonological processes like full harmony in Guébie.
- CBP makes predictions about the locality of interacting morpheme-specific phonological requirements and the locality domain of phonological alternations: they are restricted by syntactic phase boundaries.

MaxEnt-HG can be used to accurately model the distribution of phonotactic word types in the lexicon that alternate.

- However, it cannot correctly predict the frequency with which a given root will alternate in a lexically-specific process, based on phonological factors alone.
- This finding adds to a line of recent work showing what kinds of data MaxEnt models can and cannot handle (cf. Anttila et al. 2019).

References

Anttila, Arto, Scott Borgeson, and Giorgio Magri. 2019. Equiprobable mappings in weighted constraint grammars. In *Sigmorphon 2019: Proceedings of the 16th sigmorphon workshop on computational research in phonetics, phonology, and morphology*.

Bearth, Thomas. 1971. L'énoncé toura (côte d'ivoire). *Summer Institute of Linguistics* .

Clements, Nick. 1986. Syllabification and epenthesis in the Barra dialect of Gaelic. In *The phonological representation of suprasegmentals*, ed. Koen Bogers, Harry van der Hulst, and Marten Mous, 317–336. Dordrecht: Foris publications.

Clements, Nick. 1991. Place of articulation in consonants and vowels: A unified theory. In *Working papers of the Cornell phonetics library*, volume 5, 677–123.

Green, Christopher R, and Boubacar Diakite. 2008. Emergent syllable complexity in Colloquial Bamana. *The Journal of West African languages* 35:45.

Hale, Kenneth, and Josie White Eagle. 1980. A preliminary metrical account of Winnebago accent. *International Journal of American Linguistics* 46:117–132.

Halle, Morris, Bert Vaux, and Andrew Wolfe. 2000. On feature spreading and the representation of place of articulation. *Linguistic Inquiry* 31:387–444.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. University of Chicago Press.

Hayes, Bruce, Colin Wilson, and Anne Shisko. 2009. MaxEnt grammar tool. URL `http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/`.

Itô, Junko, and Armin Mester. 1995. Japanese phonology. In *The handbook of phonological theory*, ed. John Goldsmith, 816–838. Oxford: Blackwell.

Kawahara, Shigeto. 2007. Copying and spreading in phonological theory: Evidence from echo epenthesis. In *University of Massachusetts occasional papers in linguistics 32: Papers in Optimality Theory iii*, ed. Leah Bateman, Michael O'Keefe, Ehren Reilly, and Adam Werle. GLSA Publications.

Linzen, Tal, Sofya Kasyanenko, and Maria Gouskova. 2013. Lexical and phonological variation in Russian prepositions. *Phonology* 30:453–515.

Marchese, Lynell. 1979. *Atlas linguistique kru*. Abidjan: ILA.

Miner, Kenneth L. 1989. Winnebago accent: The rest of the data. *Anthropological Linguistics* 148–172.

Miner, Kenneth L, and Dorsey. 1979. Dorsey's law in Winnebago-Chiwere and Winnebago accent. *International journal of American linguistics* 45:25–33.

Sande, Hannah. 2019. A unified account of conditioned phonological alternations: Evidence from Guébie. *Language* 95(3):456–497.

Sande, Hannah, and Peter Jenks. 2018. Cophonologies by phase. *NELS 48 Proceedings* .

Smith, Jennifer L. 2011. Category-specific effects. *The Blackwell companion to phonology* 4:2439–2463.

Stanton, Juliet, and Sam Zukoff. 2018. Prosodic identity in copy epenthesis. *Natural Language & Linguistic Theory* 36:637–684.

Vydrine, Valentin. 2004. Areal and genetic features in West Mande and South Mande phonology: In what sense did mande languages evolve. *Journal of West African Languages* 30:113–126.

Vydrine, Valentin. 2010. Le pied métrique dans les langues mandé. *Essais de typologie et de linguistique générale. Mélanges offerts à Denis Creissels* 53–62.

Zogbo, Lynell. 2019. Central vowels in the Kru language family: Innovation and areal spreading. *ACAL 47 Proceedings* 725–750.

Zuraw, Kie. 2010. A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language & Linguistic Theory* 28:417–472.

Zymet, Jesse. 2018. Lexical propensities in phonology: corpus and experimental evidence, grammar, and learning. Doctoral Dissertation, UCLA.