Malagasy OCP targets a single affix: implications for morphosyntactic generalization in learning

Investigators have uncovered evidence for phonological LEARNING BIASES: biases inherent in learners that favor certain language phonologies over others (Wilson 2006, Finley 2012, Moreton & Pater 2012, Hayes & White 2013, White 2014, McMullin & Hansson 2014, *a.o.*). To what extent can a learning bias be defied in language? This question bears directly on the theory of phonological learning, as it addresses the limits of learner capability.

A growing family of findings suggests that learners tend to favor phonological constraints that are morphosyntactically general—i.e., obeyed by at least several morphemes, or in multiple or all grammatical contexts. That phonological alternations are typically corroborated by the phonotactics of a given language was observed as early as Chomsky & Halle (1968) and Kenstowicz & Kisseberth (1977), but the generalizing tendency just mentioned has also been observed in a number of recent corpus studies: Martin (2007, 2011), Zuraw (2015), Breiss & Hayes (2018), and Shih & Zuraw (2018) observe cases of grammatical "leaking", in which strong phonotactic restrictions tend to manifest across word boundaries or compound boundaries, or affect the choice between grammatical constructions. Chong (2017) found that certain alternations purported to be apparent derived environment effects are just that—merely apparent. For example, though Korean *t*-palatalization triggered by high front vocoids was previously proposed to constitute a derived environment effect due to there existing *ti*-sequences in some roots (Kiparsky 1973, 1993; Iverson & Wheeler 1988; *et seq*), Chong showed that such sequences are highly underattested in the Korean lexicon. Generalization effects were also borne out in artificial language learning experiments: Myers & Padgett (2014) found that participants generalize a phrase-final devoicing pattern to the word-final domain without exposure to unambiguous evidence; Chong (2017) found that participants more readily learned a suffixal harmony alternation when they were exposed to higher rates of root harmony, corroborating proposals that phonotactic generalizations assist in acquiring alternations (Tesar & Prince 2003, Hayes 2004, Jarosz 2006, *a.o.*).

In light of their findings, Martin (2011) and Chong (2017) propose learning models whereby whenever the learner weights positively a structure-specific constraint (e.g., applying only stem-internally), it gives positive weight to an analogous structure-insensitive constraint, leading to the generalizing tendency. If there were to exist an alternation that applies consistently in a constrained morphosyntactic context without even an analogous *statistical tendency* in phonotactics to accompany it, then that would complicate our understanding of learners' generalizing bias, as it would suggest that wholly structure-specific patterns could indeed be learned. Theories of biased learning would then have to respond accordingly, spelling out a model that could acquire wholly structure-specific phonology despite the bias.

This squib presents one such alternation. Malagasy displays backness dissimilation, an alternation that has persisted across multiple generations that sends a back vowel to front in the presence of a nearby back vowel. The process applies very consistently to the passive imperative suffix, *–u*, and is blocked by intervening front vowels—a behavior typical of dissimilation—suggesting the working of an OCP constraint. But *–u* is the only affix in the language that undergoes dissimilation, and is the only suffix even eligible to undergo it. Stems in the lexicon, moreover, show no preference for dissimilation whatsoever; they in fact display a modest but highly significant opposing preference for harmony. This suggests that Malagasy learners induce a morphologically specific OCP constraint—specific either to –*u* alone or to the suffix domain as a whole—without the need for a corroborating phonotactic trend. These findings suggest that *no* degree of morphosyntactic generality is a *necessary* condition for learning. Though learners might be biased towards acquiring grammatically general constraints, the Malagasy system suggests that they are capable of overriding it. I present this system below, and discuss the problems it poses for a theory in which learners favor grammatically general constraints.

1. Backness dissimilation applying to the passive imperative suffix

Unless otherwise specified, the data below come from the Malagasy Dictionary and Encyclopedia of Madagascar (hereafter MDEM; malagasyword.org; de la Beaujardière 2004), an annotated online corpus containing ~92,000 Malagasy words. The Malagasy vowel inventory is composed of [i e a u] (Parker 1883, de la Beaujardière 2004). There are four suffixes: the passive suffixes –*ina* and –*ana*, the active imperative suffix –*a*, and the passive imperative suffix –*u* (Parker 1883, Richardson 1885).[1]

The passive imperative suffix conditionally undergoes backness dissimilation (Parker 1883, AUTHOR 2015, 2018): underlying –*u* (1a-b) surfaces as –*i* after stems containing

[1] Note that MDEM specifies that –*u* can also attach to verbs to form relative imperatives, in addition to passive imperatives. I nevertheless follow Parker (1883), Manaster-Ramer (1995), *inter alia* in calling it a passive imperative suffix.

u (2a-d) unless a front vowel intervenes (3a-b). The alternation conforms to patterns driven by the Obligatory Contour Principle (Leben 1973, Goldsmith 1976, *et seq*).

Underlying –u	(1a)	/bata+u/	[bata-u]	lift-PASS.IMP
	(1b)	/sava+u/	[sava-u]	inspect-PASS.IMP
Backness dissimilation	(2a)	/bab**u**+**u**/	[bab**u**-**i**]	plunder-PASS.IMP
	(2b)	/t**u**v+**u**/	[t**u**v-**i**]	fulfill-PASS.IMP
	(2c)	/s**u**av+**u**/	[s**u**av-**i**]	bless-PASS.IMP
	(2d)	/**u**^{n}dan+**u**/	[**u**^{n}dan-**i**]	bolster-PASS.IMP
Blocking by front vowels	(3a)	/t**uri**+**u**/	[t**uri**-**u**]	preach-PASS.IMP
	(3b)	/f**ule**s+**u**/	[f**ule**s-**u**]	thread-PASS.IMP

5,035 words in MDEM with the passive imperative suffix were extracted. The counts in Table 1 show that dissimilation is triggered by the presence of stem-internal *u*, applies regularly when the trigger is local and semi-regularly across *a*, and is regularly blocked by front vowels.

Context (ignoring consonants)	***–u***	***–i***	**Dissim. rate**	**Example**
No trigger	2649	8	0.0%	bata-u
Adjacent trigger	6	1365	99.6%	bab**u**-**i**
Intervening *a*	272	245	47.3%	t**u**da-**i**
Intervening front vowel	485	5	0.1%	t**u**r**i**-**u**

Table 1: *Counts for Malagasy backness dissimilation*

Multiple lines of evidence suggest that Malagasy speakers acquire this alternation. For one, the process applies very consistently, and in particular *essentially categorically* in the local case: the *–u* allomorph of the suffix surfaces after nearly all of the 2,657 stems that lack *u*, while the *–i* allomorph surfaces after nearly all of the 1,371 stems whose last vowel is *u*. Moreover, dissimilation is observed across at least two generations: it was reported as early as Parker (1883), and evidence for it appears in dictionaries since then (e.g., Abinal & Malzac 1888, Rajemisa 1985, de la Beaujardière 2004).

To further assess whether speakers extended the pattern to novel contexts, I conducted a study to determine whether dissimilation and its blocking is observed even when *–u* comes after loaned stems occurring in MDEM, as in (4a-l) below. Stems marked with (W) are indicated to be loans in the World Loanword Database (wold.clld.org; Adelaar 2009b); stems marked with (Bl) are indicated to be loans in Blench (2008); and the rest are indicated to be loans in MDEM itself. The study uncovered fourteen loaned stems whose last vowel is *u*; even in these stems, we observe *categorical* dissimilation when the closest vowel to the suffix is a trigger, reflecting the pattern observed across the corpus. The study uncovered five stems whose second-to-last vowel is *u*, and whose last vowel is *a*. These words, on the whole, vary in whether they undergo dissimilation, and therefore also reflect the corpus pattern. Finally, the study uncovered five stems with a triggering *u* being separated from the suffix by front vowel(s); in all cases, we observe categorical blocking, again following the corpus pattern. These data suggest that speakers consistently extended the dissimilatory patterns observed in the corpus when they encountered eligible novel words. Below I provide some of the loanwords—for the full set of them, see the Appendix.[2]

[2] A reviewer expresses concern that speakers may not extend the alternation in light of Becker, Ketrez & Nevins's (2011) findings regarding speaker knowledge of unnatural conditioning factors of a laryngeal alternation in Turkish. Becker et al.'s results suggest that speakers fail to acquire a modest contrast in the rate of so-called deneutralization

					Loaned from
Categorical	(4a)	/a^{m}pu^{m}b**u**+**u**/	[a^{m}pu^{m}b**u**-**i**]	husk-PASS.IMP	Bantu (Bl)
Local	(4b)	/kirar**u**+**u**/	[kirar**u**-**i**]	shoe-PASS.IMP	Bantu (W)
Dissimilation	(4c)	/mats**u**+**u**/	[mats**u**-**i**]	march-PASS.IMP	English
	(4d)	/puruf**u**+**u**/	[puruf**u**-**i**]	proof-PASS.IMP	English
	(4e)	/ikisukis**u**+**u**/	[ikisukis**u**-**i**]	pig call-PASS.IMP	French (W)
	(4f)	/pai^{ŋ}g**u**r+**u**/	[pai^{ŋ}g**u**r-**i**]	hairpin-PASS.IMP	French
Variable	(4g)	/a^{m}b**u**az+**u**/	[a^{m}b**u**az-**i**]	dog-PASS.IMP	Bantu
Nonlocal	(4h)	/ipatal**u**ha+**u**/	[ipatal**u**ha-(**u**~**i**)]	pants-PASS.IMP	French
Dissimilation	(4i)	/k**u**^{m}pa+**u**/	[k**u**^{m}pa-**u**]	compass-PASS.IMP	French
Blocking	(4j)	/bur**usi**+**u**/	[bur**usi**-**u**]	brush-PASS.IMP	French
	(4k)	/tur**usi**+**u**/	[tur**usi**-**u**]	gauge-PASS.IMP	French
	(4l)	/**u**^{m}b**i**+**u**/	[**u**^{m}b**i**-**u**]	OX-PASS.IMP	Bantu (Bl)

Remarkably, it would seem that the passive imperative suffix is the *only* affix to undergo dissimilation, and, assuming the process sends back vowels to front but not vice versa, is the only suffix even eligible to undergo it, as far as I am aware (being the only one to contain *u*). Even if we assume that dissimilation sends back vowels to front *and* vice versa, it is still not displayed by any other affix, according to an MDEM search—see the Appendix for details. If there were other evidence for a dissimilatory tendency in the grammar, we would expect to find it in phonotactics. We now turn to a corpus study of roots to assess whether this is the case.

---

observed between two environments: after high vowels (~74% in Becker et al.'s corpus given in their supplementary materials, if we include alternators + vacillators) versus after non-high vowels (~43%). I see local dissimilation in Malagasy as fundamentally different: it applies nearly obligatorily after 1,371 stems in which *u* is the nearest stem-internal vowel. Even leaving aside the productivity of local dissimilation after loan stems, I am not aware of any literature suggesting that language learners could fail to extend a categorical alternation after such a great number of stems.

2. A backness harmony trend in Malagasy stem phonotactics

Surprisingly, roots display a modest but highly significant tendency toward backness *harmony*. MDEM gives numerous harmonic roots:

(5)	k**i**r**i**	'small hole'	sar**u**t^{r}**u**	'cape'	**u**z**u**na	'curse'
	l**u**f**u**	'persistence'	t**e**v**i**ka	'spasm'	ts**i**^{n}d^{r}**i**	'compression'
	g**e**g**i**	'indiscreet'	v**u**l**u**	'color'	d**u**k**u**	'identity'

Counts of tier-adjacent pairs involving only front or back vowels (*i*, *e*, and *u*) were enumerated across 4,514 roots that were extracted from MDEM. The counts reveal no preference for disharmonic sequences in roots, as Table 2 reveals below. Note that the majority of roots in the corpus are classified as nouns (2,737), adjectives (729), or adverbs (733); verbs are derived through affixation (*cf.* Keenan & Polinsky 1998).[3]

	# harmonic **V**C$_0$**V** *seq.s*	*# disharmonic* **V**C$_0$**V** *seq.s*	*# harmonic* **V**C$_0$aC$_0$**V** *seq.s*	*# disharmonic* **V**C$_0$aC$_0$**V** *seq.s*
Within noun roots	786	602	35	44
Within adj. roots	185	183	27	11
Within adv. roots	312	188	109	49
Within interj., conj., prep. roots	96	41	24	14
TOTAL	**1379**	**1014**	**205**	**118**

Table 2: *Raw counts of (dis/)harmonic sequences in roots*

[3] Some words displaying reduplication (*cf.* Lin 2005) were classified as roots in the corpus; in these cases, only the root involved in reduplication contributed to the counts, rather than the reduplicated stem as a whole. A conference reviewer points out that there could exist productive pseudoreduplication, with the first syllable being a copy of the second, potentially inflating the harmony rate. The corpus revealed that only 115 of the 4,514 roots have matching first and second syllables, with only 64 beginning with a front or back vowel ([diditra] = 'twisting', [vuvuka] = 'dust'). It is not at all obvious that the language possesses pseudoreduplication, considering how low the count is here.

There are around 350 more local harmonic sequences than local disharmonic sequences, and around 100 more nonlocal harmonic sequences than nonlocal disharmonic sequences. This makes backness dissimilation highly morphologically specific: it requires reference to the suffix domain or to the passive imperative suffix in particular, and lacks a counterpart generalization in stem phonotactics.

In fact, as we will see, Malagasy displays a significant tendency toward backness *harmony* in roots—these counts are unlikely to have arisen by chance alone. The observed rates of local and nonlocal harmony are 1379/(1379 + 1014) = 57.3% and 205/(205+ 108) = 63.5%, respectively. We can calculate the expected local harmony rate given the frequencies of front and back vowels by extracting from roots all V_1V_2 sequences in which each vowel belongs to [i e u], and calculating $[p(V_1 = u) \times p(V_2 = u)] + [p(V_1 = i \text{ or } e) \times p(V_2 = i \text{ or } e)]$, where e.g. $p(V_1 = u)$ is the number of instances of *u* in V_1 position divided by the number of instances of *i*, *e*, and *u* in V_1 position. The expected rate of nonlocal harmony is computed analogously over V_1aV_2 sequences. Doing this, we obtain 51.6% and 57.7% as expected rates of local and nonlocal harmony. Comparing the observed and expected rates, we find that observed rates (local: 57.3%; nonlocal: 63.5%) are higher than expected (local: 51.6%; nonlocal: 57.7%). To determine whether harmonic sequences occur *significantly* more than chance would predict, we can run a Monte Carlo simulation (Kessler 2001). To run a simulation for local vowel sequences, we gather pairs of tier-adjacent vowels belonging to [i e u], shuffle the second vowels of each pair and randomly concatenate each of them to a first vowel, calculate the new harmony rate, and then repeat 10,000 times. The simulation for nonlocal sequences

(V_1aV_2) can be computed analogously. Figures 1a-b below show histograms of (non/)local harmony rate frequencies after the 10,000 trials.

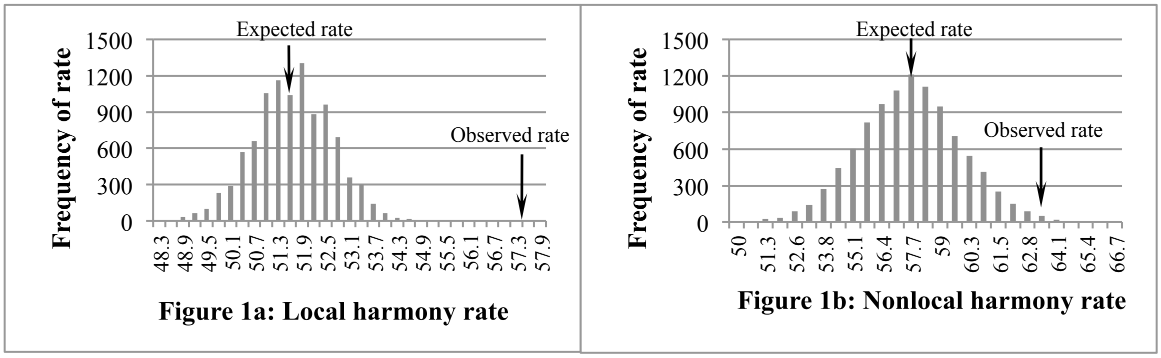

Figure 1a: Local harmony rate

Figure 1b: Nonlocal harmony rate

Figures 1a & 1b: *distribution of non/local harmony rates yielded by Monte Carlo trials*

For local harmony, the observed rate of 57.3% is greater than *any rate* yielded by 10,000 trials, and is thus significantly greater than chance would predict (est. $p < \frac{1}{10000} = 0.0001$). For nonlocal harmony, the observed rate of 63.5% is greater than 9,834 of the trials, and is thus significantly above chance as well (est. $p = \frac{10000-9834}{10000} = 0.008$). The results suggest that overrepresentation is not coincidental, but rather reflects a backness harmony preference in phonotactics. Note that there exists some evidence of a harmony alternation as well. The *–in–*/*–un–* infix is used to create passive verbs ([sava] = clear, [s-in-ava] = clear-PASS; [fidi] = choice, [f-in-idi] = choice-PASS). MDEM gives 288 words with *–in–* and 14 with *–un–*. *–in–* can surface before any vowel, and in particular surfaces before *u* in 56 forms. But in the 14 forms with *–un–*, the following vowel is always *u* ([b**u**ri] = round; [b-**u**n-**u**ri] = round-PASS), suggesting that *–un–* is selected to satisfy a (weak) harmony drive.[4] That an infix can harmonize while another suffix dissimilates is reminiscent of Yucatec Maya, in which two suffixes harmonize for backness and height, but another dissimilates for backness, and yet another for backness and height (Blair 1964; see Krämer 2001 for an account). Altogether, these cases suggest that *contradictory* markedness pressures can target different morphemes or domains.

To summarize, Malagasy backness dissimilation applies consistently to the passive imperative suffix and displays blocking behavior typical of OCP. Roots, however, show no dissimilatory tendency, but rather a modest but highly significant harmony preference.

3. Discussion and potential analytical directions

The Malagasy system complicates the discussion of morphosyntactic generality bias in phonological learning. Martin (2011) finds that phonotactic generalizations can "leak" into the cross-boundary domain: in Navajo sibilant harmony and English geminate avoidance, a categorical generalization within roots is mirrored by a statistical tendency across compound boundaries. Moreover, Chong (2017)'s corpus study of Korean palatalization and Turkish velar deletion, alternations previously purported to constitute derived environment effects, suggests that the effect is merely apparent in that the alternations are either accompanied by a statistically significant phonotactic generalization, or not productive. To account for these facts, Martin and Chong propose learning models that cannot acquire wholly structure-specific principles (see esp. Chong 2017, Section 5.4)—in particular, MaxEnt learning systems whereby whenever the

[4] An anonymous reviewer points out that these data suggest that infixes are treated as part of the stem in a way that other affixes are not. It may be that only prefixes, suffixes and other morphemes located at the edge depart from language-wide phonotactics.

learner weights positively a structure-specific constraint (e.g., applying only across a boundary), it gives positive weight to an analogous structure-insensitive constraint, leading to the generalizing tendency. Perhaps a variety morphophonological frameworks that do not encode the generalizing bias would be able to capture the Malagasy data; but models that cannot learn an alternation without phonotactic evidence, or that are biased to generalize the driving force for the alternation across domains, cannot obviously be applied to Malagasy, at least without further elaboration.

Here we briefly speculate upon a few potential solutions to the problem. First, perhaps when the passive imperative suffix was adopted in Malagasy, dissimilation arose to distinguish the suffix boundary (Robert Daland and Keren Rice, p.c.)—a drive that conflicts with the generalizing bias. Another possibility is to say that Malagasy OCP targets the passive imperative suffix *–u* rather than the entire suffix domain, and that while any particular *morpheme* can depart from typical phonological behavior in a language, whole *domains* must respect the generality property to some degree. Of course, we cannot be certain that OCP targets the passive imperative exclusively in Malagasy: it may be that OCP scopes over the whole domain, with the one suffix violating it, *–u*, undergoing dissimilation. Alternatively, it could be that a generalizing bias applies even in Malagasy learners, but that they make use of a phonotactic harmony constraint that counteracts leaking of dissimilation into roots (see AUTHOR 2018 for a model of this). That the same grammar could possess contradictory pressures is arguably unsurprising in the case of backness restriction: backness harmony *and* dissimilation are observed crosslinguistically (Parker 1883, Esztergár 1971, Campbell 1977, Clements & Sezer

1982, Itô 1984, Harrison 1999, *a.o.*), and the two alternations can apply to different affixes in the same language (as in Yucatec Maya and Malagasy).[5] Finally, a reviewer suggests that nonlocal patterns may inherently require domains, and therefore one might expect that only these patterns could be completely structure-specific. A better understanding of the typology of strictly local restrictions—e.g., whether wholly structure-sensitive local restrictions exist—should address this.

4. Conclusion

Several findings now suggest that learners tend to favor morphosyntactically general phonological constraints. This squib argues that this bias, if it exists, can be overridden. Malagasy backness dissimilation applies very consistently to the passive imperative suffix *–u*, and displays blocking behavior typical of OCP. But *–u* is the only affix in the grammar that undergoes it, and is the only suffix even eligible to undergo it. Stems, on the other hand, display a modest but significant harmony trend. This suggests that Malagasy learners induce a morphologically specific OCP constraint—specific either to *–u* or to the suffix domain as a whole—without the need for a corroborating phonotactic trend. These findings suggest that *no* degree of morphosyntactic generality is a *necessary* condition for learning. Though learners might favor grammatically general constraints, the Malagasy system suggests that they are capable of overriding this bias.

[5] A reviewer points out that this predicts the existence of systems with harmony applying across a boundary and roots showing dissimilation. Whether we find such systems is a question left to future investigation.

Appendix

Given below is the full set of loanwords with *–u*. Stems marked with (W) are indicated to be loans in the World Loanword Database (wold.clld.org; Adelaar 2009b); stems marked with (A) are indicated to be loans in Adelaar (2009a); stems marked with (Bl) are indicated to be loans in Blench (2008); and the rest are indicated to be loans in MDEM. It is again noted that, even in these stems, we observe *categorical* dissimilation when the closest vowel to the suffix is a trigger, variable—but extant—dissimilation when *a* intervenes, and categorical blocking when a front vowel intervenes, following the corpus-wide patterns illustrated in Table 1. These data suggest that speakers extended the dissimilatory patterns when they encountered eligible novel words.

					Loaned from
Categorical	(6a)	/aka^{n}d^{z}**u**+**u**/	[aka^{n}d^{z}**u**-**i**]	coat-PASS.IMP	Bantu (W)
Local	(6b)	/a^{m}pu^{m}b**u**+**u**/	[a^{m}pu^{m}b**u**-**i**]	husk-PASS.IMP	Bantu (Bl)
Dissimilation	(6c)	/a^{ŋ}gan**u**+**u**/	[a^{ŋ}gan**u**-**i**]	story-PASS.IMP	Bantu
	(6d)	/bab**u**+**u**/	[bab**u**-**i**]	plunder-PASS.IMP	Bantu
	(6e)	/kirar**u**+**u**/	[kirar**u**-**i**]	shoe-PASS.IMP	Bantu (W)
	(6f)	/kuhukuh**u**+**u**/	[kuhukuh**u**-**i**]	poultry call-PASS.IMP	Bantu (W)
	(6g)	/kulukul**u**+**u**/	[kulukul**u**-**i**]	turkey sound-PASS.IMP	Bantu (Bl)
	(6h)	/aref**ut**+**u**/	[aref**ut**-**i**]	rifle sound-PASS.IMP	English (A)
	(6i)	/mats**u**+**u**/	[mats**u**-**i**]	march-PASS.IMP	English
	(6j)	/pisupis**u**+**u**/	[pisupis**u**-**i**]	puss-PASS.IMP	English
	(6k)	/puruf**u**+**u**/	[puruf**u**-**i**]	proof-PASS.IMP	English
	(6l)	/ibila**ur**+**u**/	[ibila**ur**-**i**]	blot-PASS.IMP	English
	(6m)	/ikisukis**u**+**u**/	[ikisukis**u**-**i**]	pig call-PASS.IMP	French (W)
	(6n)	/pai^{ŋ}g**ur**+**u**/	[pai^{ŋ}g**ur**-**i**]	hairpin-PASS.IMP	French
Variable	(6o)	/a^{m}b**u**az+**u**/	[a^{m}b**u**az-**i**]	dog-PASS.IMP	Bantu
Nonlocal	(6p)	/ipatal**u**hav+**u**/	[ipatal**u**hav-**i**]	pants-PASS.IMP	French
Dissimilation	(6q)	/ipatal**u**ha+**u**/	[ipatal**u**ha-(**u**~**i**)]	pants-PASS.IMP	French
	(6r)	/av**u**na+**u**/	[av**u**na-**u**]	knot-PASS.IMP	Bantu (Bl)
	(6s)	/k**u**^{m}pa+**u**/	[k**u**^{m}pa-**u**]	compass-PASS.IMP	French

Blocking	(6t)	/bur**usi**+**u**/	[bur**usi-u**]	brush-PASS.IMP	French
	(6u)	/tur**usi**+**u**/	[tur**usi-u**]	gauge-PASS.IMP	French
	(6v)	/tab**uri**+**u**/	[tab**uri-u**]	round-PASS.IMP	Bantu
	(6w)	/**u**^{m}b**i**+**u**/	[**u**^{m}b**i-u**]	ox-PASS.IMP	Bantu (Bl)
	(6x)	/p**ui**z**i**n+**u/**	[p**ui**z**i**n-**u**]	poison-PASS.IMP	English

Given below are all affixes in MDEM occurring with at least 20 stems and that can place a front/back vowel tier-adjacent to a front/back root vowel. None of them alternate based on the root vowel (see http://malagasyword.org/bins/derivLists?form#longScroll), except for *–in–/–un–*, which displays some evidence of a harmony alternation (see Section 3).

Pref.	# forms w/ pref.	Circumf.	# forms w/ circumf.	Inf.	# forms w/ inf.	Suff.	# forms w/ suff.
fi- 'manner of doing X'	2618	fi-…-ana 'instance of X'	2144	-in-/-un- -PASS-	288+14	-ina/-na[6] -PASS	1700+32
ki-[7] 'act of doing/ state of being X'	78	i-…-ana renders X into relative verb	1991				
mi- ACTIV-	4312	a^{m}pi-…-ina renders X into passive verb	31				
^{m}pi- 'one who provides X'	1975						
t͡si- 'instance of X'	46						
ku- 'that which is X'	44						
fa^{m}pi- PASS-	53						
ma^{m}pi- ACTIV-	693						
^{m}pa^{m}pi-	20						

[6] The counts of the *–na* allomorph might be inaccurate, as it also serves as the allomorph to another passive suffix, *–ana* (Richardson 1885). Regardless *–na* surfaces as a result of hiatus repair in the language (*cf.* Albro 2005, Lin 2005, O'Neill 2015, *a.o.*).

[7] ki-/ku- could be allomorphs of the same morpheme—but even if this is were true, their distributions do not appear to be conditioned by neighboring vowels (ki-: http://malagasyword.org/bins/derivLists?form=ki~#longScroll; ku-: http://malagasyword.org/bins/derivLists?form=ko~#longScroll)

'one who provides X'	

Table 3: *different frequently occurring affixes and their counts*

References

Abinal, Antoine & Victorin Malzac. 1888. *Dictionnaire malgache-français.* Ambozoantany.

Adelaar, Alexander. 2009a. Loanwords in Malagasy. *Loanwords in the World's Languages: A Comparative Handbook.* Mouton de Gruyter.

Adelaar, Alexander. 2009b. Malagasy vocabulary. In: Haspelmath, Martin & Tadmor, Uri (eds.) *World Loanword Database*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Albro, Daniel. 2005. *Studies in Computational Optimality Theory, with special reference to the phonological system of Malagasy.* Ph.D. dissertation, UCLA.

Anttila, Arto. 2006. Variation and opacity. *Natural Language & Linguistic Theory* 24, 893–944.

AUTHOR. 2015. Distance-based decay in long-distance phonological processes. In U. Steindl et al. (eds.), *Proceedings of the 32nd West Coast Conference on Formal Linguistics*, pp. 72 – 81. Somerville, MA: Cascadilla Press.

AUTHOR. 2018. Contradictory Markedness Preferences across Morphological Domains. In *Proceedings of the 35th West Coast Conference on Formal Linguistics*, ed. Wm. G. Bennett et al., 479-488. Somerville, MA: Cascadilla Proceedings Project.

Blair, Robert. 1964. *Yucatec Maya noun and verb morpho-syntax.* Ph.D dissertation, Indiana University.

Blench, Roger. 2008. The Austronesians in Madagascar and Their Interaction with the Bantu of the East African Coast: surveying the Linguistic Evidence for Domestic and Translocated Animals. *Studies in Philippine Languages and Cultures* 18: 18-43.

Breiss, Canaan & Bruce Hayes. 2018. More on leaking grammars: Sentence construction respects phonological markedness constraints. Talk given at the 92nd *Annual Meeting of the Linguistic Society of America*.

Burzio, Luigi. 1997. Cycles, non-derived-environment blocking, and correspondence. Ms., Johns Hopkins University.

Campbell, Lyle. 1977. *Quichean Linguistic Prehistory*. University of California Press, Berkeley, CA.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English.* New York, NY: Harper and Row.

Chong, Adam J. 2017. On the relation between phonotactic learning and alternation learning. Ph. D. Dissertation, University of California, Los Angeles.

Clements, George & Engin Sezer. 1982. Vowel and consonant disharmony in Turkish. In Harry van der Hulst and Norval Smith (eds.), *Structure of phonological representations (Part II)*, 213-255. Dordrecht: Foris Publications.

de la Beaujardière, Jean Pierre. 2004. *Malagasy Dictionary and Encyclopedia of Madagascar.*

Esztergár, Marianne. 1971. *A generative phonology of nouns and vowel harmony in Hungarian.* Ph.D. dissertation, UC San Diego.

Finley, Sara. 2012. Typological asymmetries in round vowel haramony: Support from artificial grammar learning. *Language and Cognitive Processes* 27: 1550-1562.

Goldsmith, John. 1976. *Autosegmental phonology*. IULC and Garland Press, New York.

Hallanger, Frederik S. 1974. *Diksionera Malagasy-Frantsay*. Trano Printy Loterana.

Hansson, Gunnar Ólafur. 2001. *Theoretical and typological issues in consonant harmony*. Ph.D. Dissertation, University of California, Berkeley.

Harrison, K. David. 1999. Vowel harmony and disharmony in Tuvan and Tofa. *Proceedings of the Nanzan GLOW*.

Hayes, Bruce. 2004. Phonological acquisition in Optimality Theory: the early stages. In Kager, Rene, Joe Paterand Wim Zonneveld, (eds.), *Fixing Priorities: Constraints in Phonological Acquisition*. Cambridge University Press.

Hayes, Bruce & James White. 2013. Phonological naturalness and phonotactic learning. *Linguistic Inquiry* 44:45-75.

Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 39, 379–440.

Itô, Junko. 1984. Melodic dissimilation in Ainu. *Linguistic Inquiry* 15, 505-13.

Iverson, Gregory K. & Deirdre W. Wheeler. 1988. Blocking and the Elsewhere condition. In Michael Hammond & Michael Noonan (eds.), *Theoretical morphology: Approaches in modern linguistics*, 325 – 338. San Diego: Academic Press.

Jarosz, Gaja. 2006. *Rich lexicons and restrictive grammars: Maximum likelihood learning in Optimality Theory*. Ph.D. Dissertation, Johns Hopkins University.

Karlsson, Fred. 1983. *Suomen kielen äänne- ja muotorakenne* [*The Phonological and Morphological Structure of Finnish*]. Helsinki: Werner Söderström Osakeyhtiö.

Keenan, Edward & Maria Polinsky. 1998. Malagasy Morphology. In A. Zwicky and A. Spencer (eds.), *Handbook of Morphology*, Oxford University Press, Oxford, pp. 563-623.

Kenstowicz, Michael & Charles Kisseberth. 1977. *Topics in phonological theory.* New York: Academic Press.

Kessler, Brett. 2001. *The significance of wordlists*. Cambridge, MA: MIT Press.

Kiparsky, Paul. 1973. Phonological representations. In Osamu Fujimura (ed.), *Three dimensions of linguistic theory*, 1–135. Tokyo: TEC.

Kiparsky, Paul. 1993. Blocking in nonderived environments. In Ellen Kaisse & Sharon Hargus (eds.), *Phonetics and Phonology 4: Studies in lexical phonology*, 277–313. San Diego: Academic Press.

Kitto, Catherine & Paul de Lacy. 1999. Correspondence and epenthetic quality. *Proceedings of AFLA 4*: 181-200.

Krämer, Martin. 2001. Yucatec Maya Vowel Alternations: Harmony as Syntagmatic Identity. *Zeitschrift für Sprachwissenschaft* 20: 175-217.

Leben, Will. 1973. *Suprasegmental phonology*. Ph.D. dissertation, MIT. Published by Garland Press, New York.

Lewis, Geoffrey. 1967. *Turkish grammar*. Oxford: Oxford University Press.

Lin, Ying. 2005. Two perspectives on Malagasy reduplication: derivational and OT analyses. *UCLA Working Papers in Linguistics 12*.

Lubowicz, Ania. 1998. Derived environment effects in Optimality Theory. In: Susan J. Blake, Eun-Sook Kim, and Emary Shahin (eds.), *Proceedings of the 17th West Coast Conference on Formal Linguistics*, Vancouver, British Columbia.

Manaster-Ramer, Alexis. 1995. On the Subject of Malagasy Imperatives. *Oceanic Linguistics* 34: 203–210.

Martin, Andrew. 2007. The evolving lexicon. Ph.D. Dissertation, UCLA.

Martin, Andrew. 2011. Grammars leak: Modeling how phonotactic generalizations interact with in the grammar. *Language* 87(4): 751–770.

McMullin, McMullin and Gunnar Ólafur Hansson. 2014. Locality in long-distance phonotactics: evidence for modular learning. In *Proceedings of the 44th Meeting of the North East Linguistic Society*. GLSA Publications, UMass, Amherst.

Moreton, Elliott & Joe Pater. 2012. Structure and Substance in Artificial-phonology Learning, Part I: Structure. *Language and Linguistics Compass*, 6(11), 686-701.

Myers, Scott & Jaye Padgett. 2014. Domain generalization in artificial language learning. *Phonology* (31): 399 433.

O'Neill, Timothy. 2015. *The phonology of Betsimisaraka Malagasy.* Ph.D. dissertation, University of Delaware.

Parker, George Williams. 1883. *A concise grammar of the Malagasy language*. London, Trübner & Co.

Rajemisa-Raolison, Régis. 1985. *Rakibolana Malagasy*. Ambozoantany.

Richardson, James. 1885. *A New Malagasy-English Dictionary.* The London Missionary Society, Antananarivo.

Rose, Sharon, & Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language* 80: 475–531.

Sezer, Engin. 1981. The k/ø alternation in Turkish. In G. N. Clements (Ed.), *Harvard Studies in Phonology*, 354–382. Bloomington: IULC

Shih, Stephanie & Kie Zuraw. 2018. Phonological conditions on variable adjective-noun word order in Tagalog. *Phonological Data and Analysis*, an online section of *Language* 94: e317–e352.

Stanton, Juliet & Sam Zukoff. 2016. Prosodic effects in segmental correspondence. *Proceedings of CLS 51.* Chicago Linguistics Society. 501-515.

Tesar, Bruce & Alan Prince. 2003. Using phonotactics to learn phonological alternations. *Proceedings of the Thirty-Ninth Conference of the Chicago Linguistics Society, Vol. II: The Panels*.

Wilson, Colin. 2006. Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science* 30.5:945-982.

White, James. 2014. Evidence for a learning bias against saltatory phonological alternations. *Cognition*, 130(1): 96–115.

Zuraw, Kie. 2015. Allomorphs of French de in coordination: a reproducible study. *Linguistics Vanguard.*