The breakdown of abstract underlying representations in Maga Rukai

Jennifer Kuo

1 Introduction

In generative phonology, alternations are explained by positing abstract underlying forms (URs), with rules or constraints that derive the correct surface forms. Once we allow URs to deviate from surface forms, it also becomes important to ask how abstract can URs be. In other words, how much can URs diverge from their surface representations? Early work in generative phonology, centered around SPE (Chomsky and Halle, 1968), allowed for high degrees of abstractness in underlying forms. Since then, however, various scholars have questioned how abstract URs can be (e.g. Kiparsky, 1968; Derwing, 1974; Tranel, 1981). In other words, it could be that speakers cannot learn overly abstract URs, and morphophonological patterns which require positing highly abstract URs will either be unproductive, or restructured by learners into more concrete patterns. Kiparsky's (1968; 1973; 1982) work was particularly influential in highlighting possible limits on abstractness, drawing on data from language change.

Maga Rukai (a dialect of Rukai) is an Austronesian language spoken in Southern-Central Taiwan (Li, 1977a). It is particularly suitable for addressing issues around representational abstractness, because it exhibits morphophonemic alternations that require positing relatively abstract URs. The pattern of interest is rhythmic vowel deletion (syncope), where broadly speaking, every odd medial vowel counting from the left is deleted. Representative examples are given in (1), using the stem and negative forms, which are marked by both the prefix /i-/ and final vowel lengthening; nouns and stative verbs are additionally prefixed with /k-/. For ease of reading, the deleted vowel is indicated with an underline. As shown in these examples, prefixation shifts which vowels get deleted, resulting in ∅∼V alternations between the stem and negative allomorphs.

As a consequence of rhythmic syncope, many stems require positing a **composite UR**, which are abstract in the sense that the UR is not identical to any one surface allomorph, but must instead combine information from multiple allomorphs. For example, the word [tmúsu]∼[i-k-timsúː] 'salt' would typically be analyzed as having the UR /timusu/; this UR must take its initial vowel from the negative form and its second vowel from the stem form.

(1) *Examples of ∅∼V alternations in Maga*

UR	STEM	NEGATIVE	GLOSS
/timusu/	t_musu	i-k-tim_suː	'salt'
/ɖamarɨ/	ɖ_marɨ	i-k-ɖam_rɨː	'moon'
/subuku/	s_buku	i-k-sub_kuː	'eyeball'
/takasuluɖu/	t_kas_luɖu	ik-tak_sul_ɖuː	'shrimp'

Maga's rhythmic syncope pattern allows us to test the learnability of abstract URs. If the syncope pattern is stable over time, this would suggest that speakers are able to learn composite URs. On the other hand, restructuring of the pattern over time would suggest that composite URs are unlearnable (or highly dispreferred), and will be mis-learned in such a way that allows learners to posit less abstract URs. In this paper, I track how the syncope pattern has been restructured over time, by comparing three sets of data: i) Proto-Rukai, which is used to infer what Maga would have looked like before rhythmic syncope resulted in a need for composite URs; ii) Maga Rukai field data from Hsin (2000); iii) my own field data (collected June 2024), which is with a younger generation of speakers than represented by Hsin (2000).

To preview the results, I find that the majority of 'composite UR' paradigms in Maga have been restructured in such a way that results in more concrete URs. Interestingly, two patterns of restructuring are observed: first, some paradigms have undergone so-called **vowel-matching** changes, which maintain the rhythmic syncope pattern but make paradigms less abstract. Other forms have been **leveled**, resulting in a loss of ∅∼V alternations. Additionally, the newer generation of speakers have leveled paradigms more than older generations; from these findings, I propose a frequency-based explanation, where leveling (rather than vowel-matching) occurs when speakers receive less input for the syncope pattern.

The rest of this paper is organized as follows: Section 2 introduces the basics of Maga phonology and describes the rhythmic syncope pattern. Section 3 then formalizes degrees of UR abstractness. In the following sections, I then compare Proto-Rukai forms against a corpus of Maga forms (Hsin, 2000), and show that two types of paradigm restructuring have occurred; Section 4 describes vowel-matching changes, while Section 5 describes cases of leveling. Following this, Section 6 presents data from a younger generation of speakers, who have leveled paradigms to a greater degree than was found in older speakers. Finally, Section 7.2 proposes a frequency-based account for why two types of restructuring are observed, and Section 8 concludes the paper.

2 Rhythmic syncope in Maga Rukai

Maga is one of the six dialects of Rukai (the other five being Tanan, Budai, Tona, Labuan and Mantauran) (Li, 1977b). The Maga dialect is spoken in Maolin Township; the population of the township is around 866 (Kaohsiung City Government Civil Affairs Bureau, 2023), but the number of fluent speakers is much less than that. The name Maga comes from the historic name of Maolin Township. The local community identify themselves and the dialect by the name *tiḍilka*, but I use the term Maga to be consistent with prior literature.

Fig. 1 shows the geographic distribution of the Rukai dialects. The Maga, Tona, and Mantauran dialects closely neighbor each other and pattern together linguistically. However, the dialects of Rukai are generally agreed to be very divergent (Li, 1977a; Zeitoun, 1995); Maga is mutually intelligible with Tona to some extent, but not with the other dialects.

Earlier descriptive work on the Maga dialect includes Li (1975) and Saillard (1995). Hsin

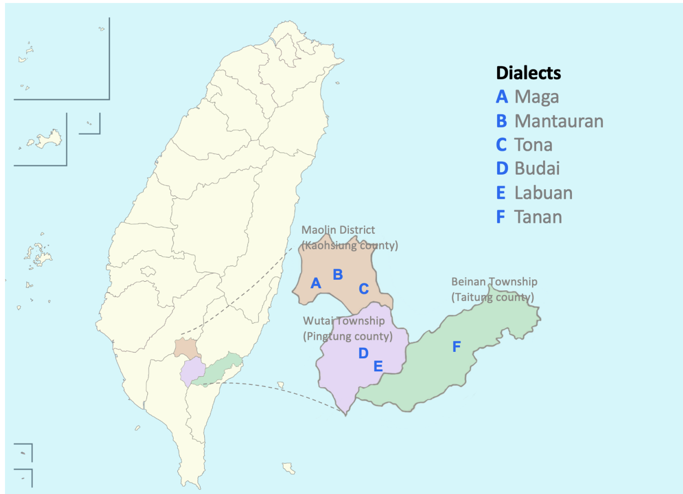

Figure 1: Dialects of Rukai
Adapted from Lin (2012): `https://ihc.cip.gov.tw/EJournal/EJournalCat/115`

(2000, 2003) provides a thorough description and analysis of Maga phonology and morphophonemic alternations; subsequent work by Chen (2008) focuses on the rhythmic syncope patterns. Additionally, Li (1977a) has done extensive work comparing the Rukai dialects and working out their internal subgrouping. Li (1977a), in particular, reconstructs the phonological system of Proto-Rukai and provides a list of around 500 cognates from dialects of Rukai. Work comparing the Rukai dialects includes Zeitoun (1995, 2024), which are cross-dialectal overviews of Rukai phonology and morphosyntax.

In the rest of this paper, unless otherwise stated, modern Maga forms are taken from Hsin (2000), supplemented with data from Li (1975). The resulting corpus has 1049 stems; of these, a subset of 790 words are given in both their stem and prefixed forms, allowing us to observe the syncope pattern. Proto-Rukai forms are taken from Li (1977a), who provides a list of 532 Proto-Rukai forms with their cognates in the six Rukai dialects.

2.1 Basics of Maga phonology

The Maga consonant inventory is given in Table 1. In addition, the language has seven vowels, /i e u o ɨ ə a/. The two central vowels /ɨ/ and /ə/ are marginally contrastive (and actually descended from the same phoneme in Proto-Rukai); [ə] is sometimes an allophone of /ɨ/, there are numerous transcription inconsistencies for the two vowels across authors, and the perceptual distinction between the two vowels is unclear, even for native speakers (Hsin, 2000, p. 37). However, there are minimal pairs between the two, such as [bɨkɨ] 'pig' vs. [bəkɨ] 'nose running'.

Stress in Maga is argued to be predictable, falling on either the penultimate syllable (e.g. [ábu] 'ash') or on final heavy (C)V: syllables (e.g. [adóː] 'mud'). Maga syllables have a (C)(C)V(C) structure with some additional restrictions. First, no final consonants are observed. Additionally, consonant clusters are restricted to two segments, regardless of position (initial vs. medial).

	Labial	Dental	Alveolar	Retroflex	Dorsal
Stop	p b		t d	ɖ	k g
Affricate			c [t͡s]		
Fricative	v	θ (ð)	s (z)		
Nasal	m	n		ŋ	
Lateral			l		
Trill			r∼ɾ		

Table 1: Consonant inventory of Maga Rukai

Note that in addition to transcription inconsistencies for the central vowels [ɨ] and [ə], there are some inconsistencies between Li (1975, 1997) and Hsin (2000) in the transcription of stop voicing. For example, 'to run' is transcribed as [u-pɖ̱akɨ] in Li (1997), and as [u-pt̠ákɨ] in Hsin (2000). For consistency, I use the transcription provided in Hsin (2000), but consonant voicing should not directly concern the syncope pattern that the present study is focused on.

2.2 Basic syncope pattern

In Maga, odd medial vowels counting from the left are deleted; as will be further discussed below, initial (onsetless) vowels and final vowels do not delete because they are non-medial. This type of rhythmic syncope, or the deletion of vowels in metrically weak positions, has been observed in many languages, including Macushi (Hawkins, 1950; Kager, 1997), Nishnaabemwin (Valentine, 2001; Bowers, 2012), and Southern Pomo (Kaplan, 2022).

In the case of Maga, affixation can shift the position of the deleted vowel, resulting in ∅∼V alternations between forms of the same paradigm. This is demonstrated in (2) for the stem and negative allomorphs of 'shrimp'; negation is marked with both the prefix /i-/ and lengthening of the final syllable. Nouns and stative verbs take an additional prefix /k-/. More examples of alternations between the stem and negative forms are given in (3) below.[1] Additionally, for underlyingly disyllabic stems like (3a-b), syncope is accompanied by lengthening of the remaining surface vowel (e.g. [smáː], cf. *[smá]). This can be be analyzed as the result of a typologically well-attested word-minimality requirement, where content words are minimally bimoraic (Hayes, 1995).

(2) *Derivation for 'shrimp'*

UR	/ta$_1$ ka$_2$ su$_3$ lu$_4$ ɖu$_5$/	/i$_1$-k- ta$_2$ ka$_3$ su$_4$ lu$_5$ ɖu$_6$-V/
Syncope	t_kas_luɖu	ik-tak_sul_ɖuː
Stress	tkaslúɖu	iktaksulɖúː
SR	[tkaslúɖu]	[iktaksulɖúː]

[1] Because stress is assigned from the right edge, but syncope is counted from the left edge, the two can conflict. These inconsistencies are discussed in Section 2.5.

(3) *Examples of rhythmic syncope in Maga*

	UR	CITATION	NEGATIVE	GLOSS
a.	/sima/	smáː	i-k-simáː	‘fat’
b.	/rana/	rnáː	i-k-ranáː	‘creek’
c.	/ɖamarɨ/	ɖmárɨ	i-k-ɖamrɨ́ː	‘moon, month’
d.	/timusu/	tmúsu	i-k-timśuː	‘salt’
e.	/giŋigiŋi/	gŋígŋi	i-k-giŋgiŋíː	‘longan’
f.	/takasuluɖu/	tkaslúɖu	i-k-taksulɖúː	‘shrimp’

In this paper, I focus on comparisons between the citation form and negative prefix, because these forms are the most well-attested and found across all lexical categories. Additionally, the stem and negative forms are enough to represent the possible syncope patterns. However, note that not all affixes in Maga are in the domain of syncope (i.e. they do not trigger syncope). Affixes that trigger syncope include the negative /i-/, plural /l-/, and imperative suffix /-a/. On the other hand, Rukai also has a set of verb marking prefixes which obligatorily appear in the citation form but are not in the domain for syncope; these include /ma-/ ‘stative verb marker’, /u-/ ‘dynamic verb marker’, and a few other less productive verbalizing prefixes. This contrast in affix types is demonstrated in (4). For example, in the citation form of ‘surprised’, [ma-rkɨ́mɨ], the stem still behaves as if it is at the left edge of the word, with syncope happening on the stem-initial vowel.

(4) *Verbal prefixes are not in the domain of syncope*

UR	citation	negative	gloss
/rɨkɨmɨ/	ma-rkɨ́mɨ	i-k-rɨkmɨ́ː	‘surprised’
/busuku/	ma-bsúku	i-k-buskuː	‘get drunk’
/kɨre/	u-kréː	i-kɨréː	‘hand’

2.3 Blocking of syncope

There are several regular (phonologically-motivated) exceptions to the syncope process. First, there is a near-exceptionless tendency for syncope to be blocked in onsetless syllables. Following Chen (2008), this can be characterized as the effects of a positional faithfulness constraint MAX-V/[$_{\sigma}$, which privileges syllable-initial positions (Beckman, 1998).

Examples of onsetless syllables being blocked from syncope are given in (5) for vowels in word-initial position (5a), and medial position in the negative form (5b). The vowel that otherwise would have been deleted (i.e. is underlyingly in an odd-numbered syllable) is underlined. Overall, syncope is blocked in 97% of onsetless syllables that should otherwise undergo syncope.

(5) *Onsetless vowels do not delete (137/141; 97%)*

a. *Initial vowels (N = 120/124; 97%)*

STEM	UR	GLOSS
ábu (cf. *bú)	/abu/	'ash'
alápi (cf. *lápi)	/alapi/	'slate'
uvácɨ (cf. *vácɨ)	/uvacɨ/	'vein'

b. *Medial onsetless syllables in negative forms (N = 17/17; 100%)*

NETATIVE	UR	GLOSS
i-k-nu**ə**ŋɨ́ː (cf. *i-k-nuŋɨ́ː)	/nuəŋɨ/	'cattle, cow'
i-ka**u**cúː (cf. *i-kacúː)	/kaucu/	'mean (temper)'
i-kul**u**ə́ŋəː (cf. *i-kulə́ŋə́ː)	/kuluəŋə/	'sparrow'

Additionally, syncope is usually blocked if it would result in a geminate consonant; examples of this are given in (6). This type of blocking, termed antigemination by Note also that there is a contrast between initial and medial geminates; syncope is always blocked if it would result in an initial geminate (6a), but only variably blocked in the case of medial geminates (6b,c). The examples in (6d) show cases where syncope applies, even if this results in medial geminates in the surface form. The difference in initial and medial geminates can be attributed to a typologically common and phonetically-motivated dispreference for initial geminate consonants (Thurgood, 1993; Muller, 2001).

cite McCarthy on antigemination raint, called antigemination, which prohibits syncope r from creating clusters of identical conso

(6) *Deletion is blocked by geminates*

a. *Initial geminates (N = 25/25; 100%)*

STEM	UR	GLOSS
t**u**túku (*ttúku)	/tutuku/	'rabbit'
m**u**muə́lɨ (cf. *mmuə́lɨ)	/mumuəlɨ/	'snail'
r**a**rámɨ (cf. *rrámɨ)	/raramɨ/	'bird'

b. *Medial geminates in stems (N = 33/51; 65%)*

STEM	UR	GLOSS
θlev**á**va (cf. *θlévva)	/θelevava/	'rainbow'
u-m**u**múɖu (cf. *u-mmúɖu)	/mumuɖu/	'to bud/flower'
vlir**ú**ru (cf. *vlírru)	/valiruru/	'glutinous rice'

c. *Medial geminates in negative forms (N = 28/32; 88%)*

NEGATIVE	UR	GLOSS
i-k-porirí: (cf. *i-k-porrɨ́ː)	/poriri/	'shade'
i-k-tagagáː(cf. *i-k-taggáː)	/tagaga/	'crow'
i-pasɨsɨ́ː (cf. *i-passɨ́ː)	/pasɨsɨ/	'to sift'

d. *Examples where syncope is not blocked by medial geminates*

STEM	UR	GLOSS
brúccu (cf. *bruc**ú**cu)	/barucucu/	‘egg’
u-ssípi (cf. *u-**si**sípi)	/sisipi/	‘count’
ma-nnɨ́kɨ (cf. *ma-n**a**nɨ́kɨ)	/nanɨkɨ/	‘breathe’

2.4 Other phonological alternations

In this section, I detail two morphophonemic alternations that interact with rhythmic syncope, but are not of direct interest to the current paper. First, high vowels preceding another vowel (i.e. in hiatus) become glides rather than undergoing syncope. Examples of glide formation are given in (7). Notably, although Hsin describes glide formation as a productive process, her data contains many forms transcribed with high vowels in positions where they should have become glides. Examples of this are given in (8) for stems (8a) and negative allomorphs (8b). It is unclear if these are a result of transcription inconsistencies, or other more systematic factors.

(7) *Glide formation in VV sequences*

STEM	UR	GLOSS
kmus**y**áː	/kamusiaː/	‘sugar’
ma-l**y**ácɨ	/liacɨ/	‘blind’
arib**w**áː	/aribuaː/	‘sweat’
u-k**w**áŋɨ	/kuaŋɨ/	‘shoot (with firearm)’

(8) *High vowels in hiatus that do not delete or become glides (n = 60)*

a. *Hiatus in stem forms (N = 43/45; 96%)*

STEM	UR	GLOSS
n**u**ə́ŋɨ (cf. *nə́ŋɨ)	/nuəŋɨ/	‘cattle, cow’
ŋ**í**u (cf. *ŋú)	/ŋiu/	‘owl’
tlap**ú**i (cf. *tlápi)	/talapui/	‘firefly’

b. *Hiatus in negative forms (N = 17)*

STEM	UR	GLOSS
n**u**ə́ŋɨ (cf. *nə́ŋɨ)	/nuəŋɨ/	‘cattle, cow’
ŋ**í**u (cf. *ŋú)	/ŋiu/	‘owl’
tlap**ú**i (cf. *tlápi)	/talapui/	‘firefly’

Regardless of whether or not they undergo glide formation, high vowels in hiatus almost exceptionlessly do not undergo syncope. There are two exceptions to this, where high vowels do delete, but both cases have variants where the vowel in hiatus does not undergo syncope. For example, ‘broom’ /suəpɨ/ has stem variants [sə́pɨ] (exceptionally undergoing syncope) and [suə́pɨ] (syncope blocked by hiatus).

The second process involves lowering of the high vowels /i, u, ɨ/ to mid vowels [e, o, ə]. Examples of vowel lowering are given in (9). Following Hsin’s (2003) description, all occurrences of /i, u, ɨ/ following an /a/ are lowered to [e, o, ə] respectively, but *only* if the /a/ is deleted. In words like (9a-b), this causes lowering of /i/ in the stem form, while in words like (9c), lowering happens in the negative form. Words like [amici] (*[amece]) ‘tree root’ and [pagu] (*[pago]) ‘gallbladder’ demonstrate that lowering does not happen if the preceding /a/ is not deleted.

(9) *Vowel lowering (alternating vowels are underlined)*

	UR	STEM	NEGATIVE	GLOSS
a.	/damili/	dm**e**l**e**	i-k-daml**iː**	‘hemp’
b.	/caki/	ck**eː**	i-k-cak**iː**	‘excrement’
c.	/alapi/	alap**i**	i-k-alp**eː**	‘slate’
d.	/valu/	vl**oː**	i-k-val**uː**	‘bee’
e.	/taluɖu/	tl**o**ɖ**o**	i-k-talɖ**uː**	‘bridge’
f.	/tesavuŋu/	pa-tsavŋ**u**	i-p-tesv**o**ŋ**oː**	‘bridge’
g.	/θadɨ/	u-θdə́ː	i-θad**ɨ́ː**	‘headdress for women’
h.	/θadɨ/	u-θdə́ː	i-θad**ɨ́ː**	‘to raise (animal, children)’
i.	/akaŋɨ/	akáŋ**ɨ**	i-k-akŋ**ə́ː**	‘fish’

2.5 Opacity in the interaction of syncope and stress assignment

Aside from the issue of UR abstractness, rhythmic syncope is also potentially challenging to learn because it is inherently opaque. In particular, syncope must happen after building the metrical structure which conditions it (Kager, 1997; McCarthy, 2008), making metrical syncope difficult to account for in parallel theories of phonology like OT.

In the case of Maga Rukai, there is additional opacity because the location of surface stress does not always match the location of syncope. Maga syncope is assigned on odd syllables counting from the left, but stress is counted from the right edge, surfacing on the penultimate syllable. As a result, for words of even syllable parity, syncope is inconsistent with surface stress. This inconsistency between stress and syncope is schematized in (10). Under the UR column, the position that should have been stressed (i.e. is in penultimate position) is underlined. For stems that have underlying even parity, syncope targets the position that should have been stressed, had deletion not occurred.

This is a type of *metrical incoherence*, where a metrically conditioned process (in this case syncope) does not appear to refer to surface stress (Dresher and Lahiri, 1991; Kaplan, 2022). As a result, there is no straightforward surface-oriented solution that can derive both stress and syncope. Prior analyses of Maga syncope have therefore either opted from derivational solutions (Hsin, 2000), or include phonemic stress in underlying representations (Chen, 2008).

(10) *Interaction of syncope and stress assignment by syllable parity*

Parity	UR	SR	Example		
Even	/$\underline{\boldsymbol{\sigma}_1}$ σ_2/	[$\varnothing_1$ ˈ$\underline{\boldsymbol{\sigma}_2}$]	/t**ɨ**ba/	tb**á:**	'wild yam'
	/σ_1 σ_2 $\underline{\boldsymbol{\sigma}_3}$ σ_4/	[$\varnothing_1$ ˈ$\underline{\boldsymbol{\sigma}_2}$ $\varnothing_3$ σ_4]	/tavaŋ**u**lu/	tv**á**ŋlu	'chili pepper'
Odd	/σ_1 $\underline{\boldsymbol{\sigma}_2}$ σ_3/	[$\varnothing_1$ ˈ$\boldsymbol{\sigma}_2$ σ_3]	/ɖamarɨ/	ɖm**á**rɨ	'moon'
	/$\sigma_1\sigma_2$ σ_3 $\underline{\boldsymbol{\sigma}_4}$ σ_5/	[$\varnothing_1$ σ_2 $\varnothing_3$ ˈ$\underline{\boldsymbol{\sigma}_4}$ σ_5]	/takasul**u**ɖu/	tkasl**ú**ɖu	'shrimp'

Overall, the opacity presented by rhythmic syncope makes it arguably less learnable, and therefore diachronically unstable. For example, Bowers (2012) shows that in Nishnaabemwin, extreme vowel reduction resulted in a rhythmic syncope pattern, but this pattern broke down within just a few generations of speakers. Bowers attributes the quick breakdown of the Nishnaabemwin pattern to its opacity. For the present study, this means that restructuring of the Maga syncope pattern could be attributed to the opacity presented by rhythmic syncope, rather than to the abstractness of composite underlying representations. In Section 4, however, I present cases of restructuring that have preserved the rhythmic syncope alternations, and therefore cannot be attributed to factors like opacity-avoidance.

2.6 Historical origins of the rhythmic syncope pattern

The Maga pattern of rhythmic syncope is likely to have developed from severe phonetic reduction of unstressed vowels (followed by reassignment of stress to be penultimate). Similar patterns of diachronic change have been reported from languages like Nishnaabemwin (Valentine, 2001; Bowers, 2012). In general, when Maga stem-negative paradigms are compared against cognates in other Rukai dialects, the vowels which surface match that of cognate words.

(11) *Cross-dialectal comparison of Rukai forms*

	Maga paradigms	Proto-Rukai	Tanan	Budai	Tona	Mantauran
'tongue'	rd**á**m**ə**∼ikr**i**dm**ə́:**	*ɭidamə	ɭidámə	ɭídamə	idámə	ɭidamə
'moon'	ɖm**á**r**ɨ**∼ikɖ**a**mr**ɨ́:**	*ɖamarə	ɖamárə	ɖámarə	ɖamáə	ɖamarə

The picture that emerges is that Maga Rukai, had a point earlier in history, paradigms with no ∅∼V alternations, and therefore did not require positing abstract URs. The subsequent loss of metrically weak vowels then resulted in a synchronic pattern which requires composite URs. The (in)stability of the syncope pattern over time can then give us insight into how well abstract representations are learned and tolerated.

3 Formalizing degrees of UR abstraction in Maga

When probing at the learnability of abstract URs, we should also make a distinction between levels of abstraction. At one end of the spectrum, learners might tolerate no abstractness at all, and only ever generalize from surface allomorphs of one particular slot in a paradigm. On the other end of the spectrum, learners may be able to learn phonemes that are never even observed in the surface forms of a language.

In this paper, I adopt Kenstowicz and Kisseberth's (1977, ch. 1) taxonomy of UR abstractness, which defines levels of abstractness based on how divergent a UR is from its associated SRs; following Wang and Hayes (under revision), I refer to this hierarchy as the KK-hierarchy. The KK-hierarchy not only allows for a more fine-grained evaluation of how abstract URs can be in Maga, but also allows for comparison with recent work by Wang and Hayes (under revision), who propose a model of UR learning that is situated using the same hierarchy. In this section, I first give an overview of the KK-hierarchy, characterizing the predicted types of restructuring, if learners are restricted to each level of the hierarchy. I then characterize the amount of abstraction required in Maga.

As it turns out, if syncope regularly applies in Maga, most of the lexicon would require positing relatively abstract composite URs. However, many stems fail to undergo the expected syncope pattern; in the following two sections (Sections 4,5), I describe the types of restructuring that have occurred, all of which suggest that Maga has changed in concretist directions.

3.1 The KK-hierarchy

KK-A: only invariant properties of allomorphs

Under KK-A, the UR of a morpheme "consists of all and only the invariant phonetic properties of that morpheme's various SRs" (Kenstowicz and Kisseberth, 1977, p. 8). This is highly restrictive, and disallows standard phonemicization. For example, English "pan" [pæ̃n], where the vowel is predictably nasalized in pre-nasal position, would have the UR /pæ̃n/. KK-A', which is a less restrictive variant of KK-A, states that the UR consists only of invariant features of corresponding SRs, but that predictable allophony may be treated as unspecified (rather than listed in the UR).

Both KK-A and KK-A', when applied to Maga Rukai, would not allow any of the alternating vowels to be present in the UR. A root like 'moon', given again in (12), would have the UR /ɖmrɨ/, whose only vowel is the non-alternating stem-final vowel. If KK-A is a hard restriction on UR learning, then URs should be reanalyzed in a way that renders the alternating vowels predictable from the non-alternating parts of the stem. Hypothetically, this could result in a grammar like (13a), where only the non-alternating parts of a root has been retained. Alternatively, Maga might be restructured so that the vowel(s) that surface are always predictable (e.g. epenthesize a default 'least marked' vowel), as in (13b).

(12) *Maga URs under KK-A*

stem	negative	gloss	UR-A
ɖmarɨ	i-k-ɖamrɨː	'moon'	/ɖmrɨ/
tmusu	i-k-timsuː	'salt'	/tmsu/
smáː	ik-simáː	'fat'	/smaː/

(13) a. Grammar A1: only retain non-alternating parts of the paradigm.

UR		STEM	NEGATIVE
/ɖmrɨ/	→	[ɖmrɨ]	[i-k-ɖmrɨː]
/tmsu/	→	[tmsu]	[i-k-tmsuː]

b. Grammar A2: always epenthesize the same vowel (e.g. [ɨ])

UR		STEM	NEGATIVE
/ɖmrɨ/	→	[ɖmɨ̲rɨ]	[i-k-ɖɨ̲mrɨː]
/tmsu/	→	[tmɨ̲su]	[i-k-tɨ̲msuː]

KK-B: single allomorph

Broadly, all variants of KK-B include idiosyncratic properties of a morpheme, but have the restriction that URs must be based on the same surface allomorph (i.e. the same slot in a paradigm). Under the basic version of KK-B, the UR must be based on the surface allomorph "that appears in isolation (or as close to isolation as the grammar of the language permits)" (Kenstowicz and Kisseberth, 1977, p. 11). The less restrictive KK-B' states that the learner should base the UR on the allomorph that appears in the most contexts, while KK-B", which has be extensively explored in work by Albright (2002a,b, 2010, et seq.), restricts URs to be based on the most informative slot in a paradigm.

The subtypes of KK-B, when applied to Maga, predict grammars like (14a), where paradigms have leveled towards the stem form, or grammars like (14b), where paradigms have leveled towards the negative allomorph.[2] Note that while grammars B1 and B2 have both removed vowel alternations throuhg leveling, this is not a requirement of KK-B. KK-B also allows for grammars like (14C), where there are ∅∼V alternations, but the alternating vowels are completely predictable from one surface allomorph; in this hypothetical grammar, the UR is based on the stem allomorph, and the vowel that surfaces in the negative form is a copy of the stem vowel.

(14) a. Grammar B1: leveling towards the stem allomorph.

UR		STEM	NEGATIVE
/ɖmarɨ/	→	[ɖmarɨ]	[i-k-ɖmarɨ]
/tmusu/	→	[tmusu]	[i-k-tmusuː]

b. Grammar B2: leveling towards negative allomorph.

[2] Both allomorphs also surface in other morphological contexts

UR		STEM	NEGATIVE
/ɖamrɨ/	→	[ɖamrɨ]	[i-k-ɖ**ɨ**mrɨː]
/timsu/	→	[timsu]	[i-k-t**ɨ**msuː]

c. Grammar B3: ∅∼V alternations are predictable from the stem allomorph

UR		STEM	NEGATIVE
/ɖmarɨ/	→	[ɖm**a**rɨ]	[i-k-ɖ**a**mrɨ]
/tmusu/	→	[tm**u**su]	[i-k-t**u**msuː]

KK-C: choosing among allomorphs

In the KK-C level of the hierarchy, the UR of a morpheme is always based on some surface allomorph, but this allomorph does not need to come from the same paradigm slot for every morpheme. For Maga, this allows for grammars like (15), where some words have leveled towards the stem allomorph while others have leveled towards the negative allomorph.

(15) Grammar C: leveling towards either the stem or negative allomorph.

UR		STEM	NEGATIVE	
/ɖmarɨ/	→	[ɖmarɨ]	[i-k-ɖmarɨ]	(base = stem)
/timsu/	→	[timsu]	[i-k-timsuː]	(base = negative)

KK-D: segmentally composite URs

On this level of the hierarchy, every segment of the UR must be based on some surface allomorph, but not necessarily the same one. A traditional analysis of the Maga syncope pattern would require this level of abstraction. This is because, at least for a subset of roots, no single surface allomorph has information about all the underlying vowels. This is schematized in (16) for the word 'salt' /timusu/; the negative allomorph shows that V1 is /i/, while the stem allomorph shows that the V2 is /u/. Crucially, however, no single allomorph can tell us about both V1 and V2. If the KK-D level of the hierarchy is learnable, then we would predict Maga syncope to be stable over time, with few to no cases of restructuring.

(16) *Composite UR of 'salt'*

UR - - - - - - - - - - - - / t **i** m **u** su /

SR - - - - - - - [ikt **i** msuː] [tm **u** su]

KK-E: featurally composite URs

Under KK-E, the only restriction on each underlying segment /S/ is that it must have feature values that occur in some surface allomorph, but not necessarily the same allomorph. Consequently, URs are able to contain segments never present in the surface allomorphs (as long as every feature occurs in some allomorph). Kenstowicz and Kisseberth (1977) give the example of English /ng/ as an analysis that requires KK-E levels of abstraction. Specifically, if we take the underlying form of *sing* to be /sɪng/, with surface [ŋ] derived from /ng/, the /ng/ sequence is abstract in the sense that neither segment surfaces in any allomorphs (cf. *sing* [sɪŋ], *singer* [siŋɚ]). This level of the hierarchy does not directly concern the current paper, since the Maga pattern does not require a higher level of abstraction than KK-D.

3.2 Predictions of opacity-avoidance vs. abstractness-avoidance

As discussed in Sections 2.2 and 2.5, rhythmic syncope is opaque, as deletion must occur after the building of the metrical structure which conditions it. In most cases, rhythmic syncope is not analyzable using parallel frameworks like Classical OT. This is especially so for a metrically incoherent language like Maga. The opaque nature of rhythmic syncope could make it unstable and prone to restructuring over time; this is in fact what Bowers (2012) proposes has happened in Nishnaabemwin. The consequence is that if restructuring has occurred in Maga, it could be attributed to either abstractness avoidance or the avoidance of opaque processes.

However, the syncope pattern could be restructured in a way which renders it transparent, but still abstract in the sense of requiring composite URs. The example in (17) shows a hypothetical Maga grammar, where vowel reduction to [ə] has replaced syncope. In this new grammar, vowel reduction is driven by a constraint (or rule) that is surface satisfied (e.g. "only [ə] is licensed in metrically weak positions"). This new grammar would be more faithful to the old Maga pattern, as it maintains all the underlying vowel qualities. However, because no surface allomorph tells us about all the underlying stem vowels, a morphophonemic analysis of this "new Maga" would still require composite URs.

(17) *Hypothetical Maga: composite URs without opaque syncope*

UR		STEM	NEGATIVE
/ɖamarɨ/	→	[ɖə̠márɨ]	[i-k-ɖamə̠rɨ́ː]
/timusu/	→	[tə̠músu]	[i-k-timə̠súː]

In sum, if reanalysis has occurred in a way that makes Maga less abstract, due to some combination of leveling towards surface allomorphs, this cannot be explained purely as opacity-avoidance.

3.3 Degree of abstraction required in the Maga lexicon

As a result of rhythmic syncope, a subset of words require positing segmentally composite URs (KK-E). For other words, however, the UR is still predictable from one or both surface allomorphs.

Table (18) provides examples of paradigms that vary in their degree of abstractness.

For example, roots that are underlying /V1 CV2/ (i.e. disyllabic with on initial vowel) can be accounted for using KK-A levels of representation; there are no ∅∼V alternations because syncope prevents V1 from deleting in the stem, while in the negative form, neither vowels are in targeted positions for syncope. Some roots are predictable from one surface allomorph, so would require positing URs of levels KK-B/C. For example, roots of the shape /CV1 CV2/ never undergo syncope in the negative allomorph, so are fully predictable from this allomorph. Likewise, roots of the shape /V1 CV2 CV3/ are fully predictable from the stem allomorph; V1 does not delete because it is onsetless, and the other vowels are not in deleting positions. In general, blocking of syncope from vowel hiatus and/or geminates can also result in surface allomorphs that contain information about all the underlying vowels.

(18) *Examples of UR shapes that vary in their required degree of abstractness*

UR shape	**Predictable allomorph**	**KK level**	**UR**	**stem**	**negative**	**gloss**
/V1 CV2/	both	A	/abu/	ábu	i-abúː	‘ash’
/CV1 CV2/	negative	B/C	/tupe/	tpeː	i-k-tupéː	‘gourd dipper’
/V1 CV2 CV3/	stem	B/C	/amici/	amíci	i-k-amcíː	‘tree root’
/CV1 CV2 CV3/	neither	D	/dukacɨ/	dkácɨ	i-k-dukcɨ́ː	‘mud’

Table (19) summarizes the degree of abstractness required in the Maga corpus, *if* the syncope pattern is completely regular (i.e. assuming no restructuring has occurred). If Hsin’s corpus is taken to reflect the general distributions of the Maga lexicon, we see that the majority of the lexicon requires positing abstract URs of some kind. Only 8% of the corpus can be accounted for under KK-A. On the other end of the spectrum, over 60% of words require a composite UR, and can only be accounted for under KK-D.

In the rest of the lexicon (∼31%), the UR is predictable from either the stem or negative allomorph. Under KK-B (isolation allomorph only), paradigms are expected to level towards the stem form. However, under KK-B” (most informative allomorph only), we predict leveling towards the negative allomorph, since it the predictable allomorph for a larger proportion of the lexicon. Recall that KK-C permits leveling in both directions; if this leveling is driven by informativeness, then we can predict leveling towards the stem allomorph for words of type (19ii), but leveling towards the negative allomorph for words of type (19iii).

(19) *Summary of UR types required in the lexicon (assuming regular application of syncope)*

	Predictable allomorph	**KK-level**	**N**	**P**
i.	both	A	84	0.08
ii.	stem	B/C	96	0.09
iii.	negative	B/C	230	0.22
iv.	neither	D	639	0.61

3.4 Zooming out: restructuring in Maga

Assuming regular application of syncope, over half of the Maga lexicon would require composite URs. However, it turns out that in total, about 24% (250/1049) of stem forms and 23% (185/790) of negative forms fail to undergo the regular syncope pattern. If we just look at the subset of 790 forms where the full stem/negative paradigm is available, 40% (314/790) paradigms have at least one member that does not undergo syncope in the expected positions. This irregularity suggests that many Maga paradigms have been restructured.

Moreover, in one subset of paradigms, although the regular syncope pattern has been maintained, URs have been reanalyzed in such a way that makes the negative allomorph more predictable from the stem allomorph; these cases are discussed in Section 4. In Section 5, I discuss cases where paradigms have leveled towards either the stem or negative allomorph, removing $\varnothing\sim$V alternations. The data I use is the same corpus of 1049 words above; note that for some of these words, the negative allomorph is not available. In total, a subset of 790 words are listed with the stem/negative paradigm; for these, we can infer the direction of restructuring.

Note that throughout the rest of this paper, unless otherwise noted, I use 'UR' to refer to the hypothetical composite UR that would be posited under a traditional morphophonemic analysis, where representations of the type KK-D are permitted. The actual representation that speakers learn, as inferred by directions of restructuring, may not exactly match these URs.

4 Vowel-matching alternations

In the Maga Rukai lexicon, vowels in alternating positions tend to match each other; examples of this are given in (20), where the vowels that undergo syncope (in one of the paradigm slots) are underlined. In fact, in Hsin's corpus, of the 322 paradigms that still require a composite UR, 61% (n = 196/322) exhibit this vowel-matching pattern. Crucially, if **vowel-matching** is a strong enough tendency in the Maga lexicon, it would actually render paradigms with vowel alternations predictable from just one surface allomorph.

(20) *Vowel matching examples*

UR	STEM	NEGATIVE	GLOSS
/si̱mi̱tu/	smi̱tu	i-k-si̱mtuː	‘lips’
/go̱ro̱masɨ/	gro̱masɨ	i-k-go̱rmasɨː	‘centipede’
/la̱ma̱ta̱kɨ/	lma̱tkɨ	i-k-la̱mta̱kɨː	‘leech’

For example, given a paradigm like [smi̱tu∼i-k-si̱mtuː], assuming vowel-matching is predictable, we could derive the negative allomorph from the stem, as schematized in (21). For illustrative purposes, I show a rule-based derivation involving a combination of vowel copying and deletion, but the mapping from the stem to negative allomorphs could happen in other ways. Additionally, though I show a derivation from the stem to negative allomorph, the input could just as well be the negative allomorph. What matters is that vowel-matching renders one allomorph predictable from the other, so that the paradigm no longer requires a composite UR. Instead, forms will be predictable from one surface allomorph, and require just KK-B/C levels of abstraction.

(21) *Deriving [i-k-si̱mtuː] from a stem base*

Input	smitu (=stem allomorph)
Morphology	i-k-smituː
Vowel copying	i-k-si̱mituː
Vowel deletion	i-k-simtuː
Stress Assignment	i-k-simtúː
Output	i-k-simtúː

In the rest of this section, I show that the proportion of vowel-matching forms in Maga Rukai has increased over time. In other words, while vowel-matching might have already been present as a tendency in an earlier stage of the language, speakers have restructured Maga paradigms in a way that has strengthened this tendency. As a result, a large proportion of Maga paradigms that would otherwise require composite URs are now predictable from surface allomorphs.

4.1 Data

To infer the direction of vowel changes over time, I compare the distribution of vowel-matching forms in Proto-Rukai and Maga. Because other dialects of Rukai have largely maintained vowel contrasts, undergoing much less vowel reduction or deletion than Maga, Proto-Rukai reconstructions can be reliably used to infer what Maga vowels should look like, if no restructuring has occurred.

As with the above sections, Maga data comes Hsin’s corpus of 1049 items (of which 790 are complete paradigms). Data on Proto-Rukai comes from Li’s (1977a) list of 532 protoforms and their cognates in the six Rukai dialects. Table (22) lists Proto-Rukai vowels and their regular reflexes in Maga; in the rest of the paper, for comparability with the Maga data, Proto-Rukai vowels are presented and organized by their reflexes in Maga.

(22) *Vowel correspondences, Proto-Rukai and Maga*

PROTO-RUKAI	MAGA	EXAMPLE
*a	a	*kav**a** > kv**a**: 'hoe'
*i	i	*rig**i** > rg**i**: 'horse'
*o, oa	u	*timos**o** > tm**ú**s**u** 'salt'; ***oa**lopo > **u**lúpu 'hunt'
*ə	ɨ	*pələŋ**ə** > plɨ́ŋ**ɨ** 'ghost'
*(a...)i	e	*cak**i** > ck**é**: 'excrement'
*(a...)o	o	*val**o** > vl**ú** 'bee'
*(a...)ə	ə	*bal**ə**bal**ə** > bl**ə́**bl**ə** 'bamboo'
*ay, *ai	e, eː	*acil**ay** > acíle 'water'; *pag**ay** > pg**éː** 'rice plant'
*aw, *ao	o, oː	*maram**aw** > marámo 'same'; *ko**aw** > ku**oː** 'eagle'

In general, there is a straightforward correspondence between Proto-Rukai and Maga, with the exception that mid vowels *o and *ə have raised to Maga [u] and [ɨ]. Proto-Rukai diphthongs like *aw and *ay are reflected as monophthongs in Maga; Maga reflexes of Proto-Rukai diphthongs can be short or long, but notably resist syncope.

The one wrinkle is that *i, *o, *ə are reflected as mid vowels [e], [o], and [ə] if they follow a *deleted* low vowel (*a). This lowering process generally does not apply if [a] has not been deleted (e.g. *aŋat**o** > aŋat**u** 'tree'), though there are a few exceptions where vowels adjacent to *a exceptionally lower (e.g. *aθ**o** > aθ**ó**ː 'dog'). The result is that in Maga, we observe vowel-lowering alternations that interact with syncope (see Section 2.4 for more details).

To approximate the degree of vowel matching in the Maga Rukai lexicon, I focus on just underlying V1 and V2 (e.g. /u/ and /a/ in /dukacɨ/ 'mud'). This dataset, though slightly pared down, should still be a good reflection of vowel-matching patterns across Maga, since most stems have three or fewer underlying vowels (70%, N = 735/1049), and for these words, only V1 and V2 are alternating.

Additionally, forms are categorized as having either **composite** or **non-composite** URs. Some forms in Maga should require composite URs, but have leveled towards one of the surface allomorphs. For example, the verb 'to tattoo' has the paradigm [u-ckícki]∼[i-ckickíː]; the negative allomorph does not show the expected syncope pattern (cf. *[i-cikcikíː]), suggesting that it has leveled towards the stem allomorph. Leveling is discussed more in Section 5; for now, we categorize the leveled paradigms as having non-composite URs. For the Proto-Rukai forms, where we do not have direct access to the negative allomorphs, the nature of the UR is inferred from word shape. For example, a protoform with the shape CVCVCV, such as *timoso 'salt', should correspond to a Maga paradigm that requires composite URs (and in fact it does, cf. [tmúsu]∼[i-k-timsúː]).

4.2 Results: degree of vowel-matching in Proto-Rukai and Maga

Going into the results, Fig. 2 shows the proportion of vowel-matching forms in Proto-Rukai vs. Maga Rukai. We see that there has generally been an increase in the proportion of vowel-matching

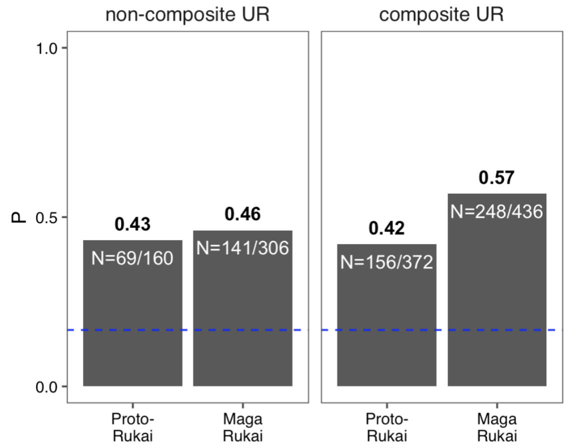

Figure 2: Proportion of vowel-matching (between V1 and V2) in Proto-Rukai vs. Maga Rukai. Dashed line indicates the chance-level rate of vowel-matching alternations.

forms between Proto-Rukai and Maga. Notably, this increase is much larger for the composite UR forms. This suggests that restructuring of UR vowels has occurred specifically in the paradigms where vowel-matching would remove composite URs, resulting in a less abstract lexicon.

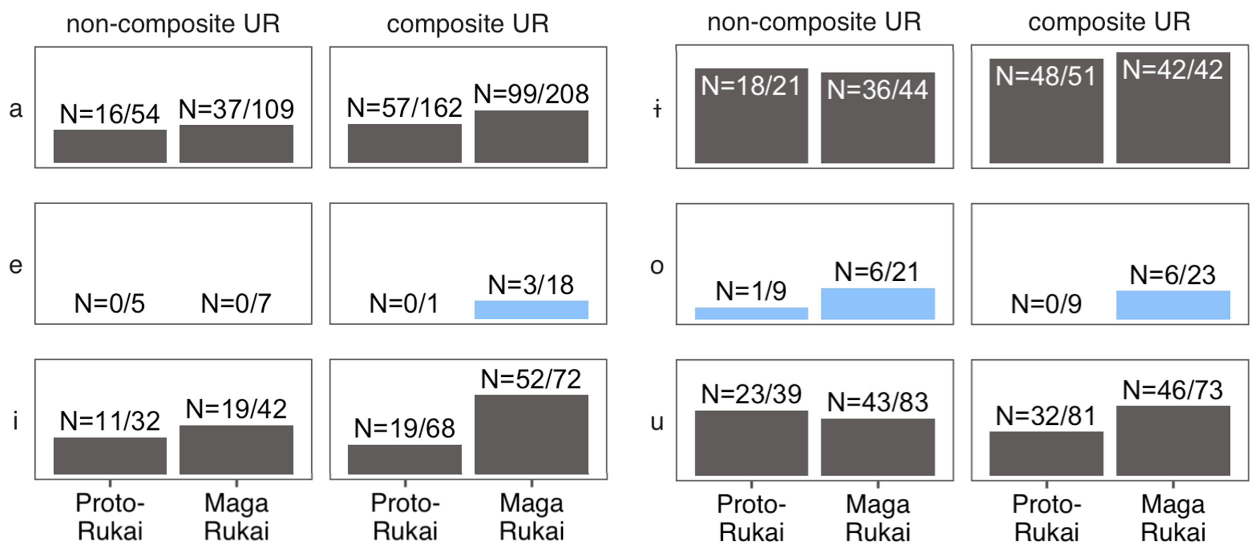

Figure 3: Proportion of vowel-matching in Proto-Rukai vs. Maga Rukai, by /V1/. Blue bars show cases where vowel-matching was historically below-chance.

Fig. 3 shows the same results, but separated by identity of underlying V1. For example, the subfigure for /u/ represents forms like /dukacɨ/ 'mud' ([dkácɨ]∼[i-k-dukcɨ́ː], non-vowel-matching) and /surugu/ 'liver' ([srúgu]∼[i-k-surgúː], vowel-matching). Forms where V1 is /ə/ are omitted because there just two of them. Interestingly, based on Fig. 3, rates of vowel-matching has increased even for vowels that were not historically matching. For example, in Proto-Rukai, there was no vowel-matching with /e, o/[3], but vowel-matching involving the mid vowels is observed in

[3]This is expected given that /e, o/ developed either from diphthongs or lowering of high vowels after deleted *a.

Maga. These results suggest that speakers learned a general vowel-matching pattern, extending it past environments observed in the lexicon. Similar trends are observed if we group forms by identity of underlying V2; this figure is given in the Appendix.

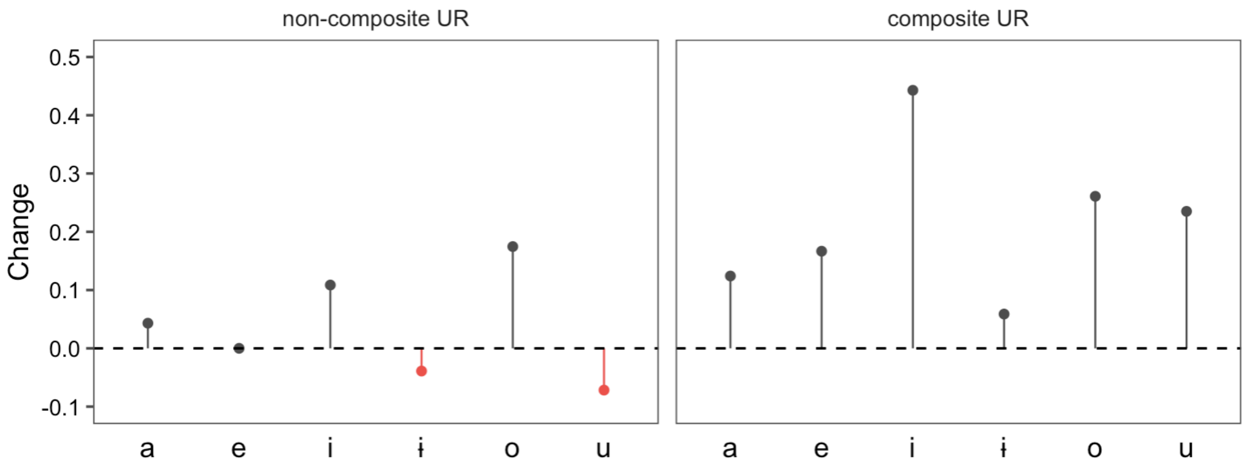

Figure 4: Change in proportion of vowel matching (Maga – Proto-Rukai)

Fig. 4 shows the same data as Fig. 3 above, but visualized in terms of the *change* in proportion of vowel-matching between Proto-Rukai and Maga Rukai. A positive value on the y-axis indicates that vowel-matching increased over time, while a negative value indicates a decrease in vowel-matching. This figure makes it evident that across almost every vowel and UR type, there has been increased rates of vowel-matching. There are just two exceptions (V1 = /u,ɨ/ in non-composite URs), both involving non-composite URs. For both, the decrease in vowel-matching is very small.

Finally, it should also be noted that vowel-matching appears to already have been present as a statistical tendency in Proto-Rukai. Maga has 6-7 phonemic vowels (depending on whether /ə/ is treated as phonemic), meaning that if vowel-matching occurred at chance, it should occur about 16.7% (1/6) of the time; this value is shown for reference in Fig. 2. Even in Proto-Rukai, vowel-matching happened at well above this rate. This is corroborated by a Monte Carlo test of significance, where I find that vowel-matching was already overrepresented for most vowels in Proto-Rukai (all vowels excluding /e/ and /o/), and that this tendency was further exaggerated in Maga Rukai. Details of the Monte Carlo test are given in Appendix 2. These results suggest that when faced with resolving the ambiguity caused by syncope, speakers utilized a tendency that was already present in the lexicon, causing it to become stronger over time.

4.3 Results: direct examples of vowel-matching changes

In this section, I look at the subset of Maga Rukai paradigms that have known Proto-Rukai forms. These forms can provide more direct evidence for vowel changes, and also give us insight into the direction of restructuring. The examples in (23) show the types of changes we are interested in, where the Maga paradigm does not show the expected vowel alternation, given the Proto-Rukai form. In these examples, the changed vowel is underlined.

(23) *Examples of vowel-matching changes*

Proto-Rukai	expected UR	actual UR	paradigm	gloss
*samito	/s**a**mitu/	/s**i**mitu/	smítu∼i-k-s**i**mtú:	'lips'
*tobakə	/t**u**bakɨ/	/t**a**bakɨ/	tbákɨ∼i-k-t**a**bkɨ́:	'shell'
*sinaw	/sin**o**sino/	/sin**i**sino/	u-sn**i**snó:∼i-sinsinó:	'wash (clothes)'
*bicoka	/b**i**cuka/	/b**u**cuka/	bcúka∼i-k-b**u**cká:	'stomach'

Overall, there were 56 forms where Proto-Rukai and Maga Rukai vowels mismatched. Table (24) summarizes the types of changes observed. The majority of changes (∼70%) have resulted in increased vowel-matching, consistent with the findings of the previous section. Of the remaining changes, most neither increase nor decrease the rate of vowel-matching; for example, Proto-Rukai **ma-riday* should correspond to Maga [ma-rdé:]∼[i-k-r**i**dé:], but the observed paradigm is instead [ma-rdé:]∼[i-k-r**u**dé:].

(24) *Summary of vowel changes*

CHANGE RESULTS IN...	N	P	EXAMPLE
MORE vowel-matching	39	0.70	*s**a**mito → /s**i**mitu/ 'lips'
LESS vowel-matching	5	0.09	*v**i**li → /v**u**li/ 'leech'
Neither	12	0.21	*ma-r**i**day → /r**u**de:/ 'fast'

For a few words (N = 5), vowel-matching forms were changed to be *non*-matching. Of these cases, two could potentially be explained as the result of speakers extending a vowel-lowering alternation, where high vowels lower following a deleted /a/; this pattern is discussed in Section 2.4, and its historical basis is discussed in Section 4.1. For example, Proto-Rukai **ma-pilay* 'tired' should correspond to Maga [ma-plé:]∼[i-k-pilé:] (UR = /pile/), but instead we observe [ma-plé:]∼[i-k-p**a**l**í:**] (UR = /pali/). This could be because speakers, when given the stem allomorph [ma-ple:], interpreted the final [e] as the result of /i/-lowering after a deleted /a/.

Direct comparison of Proto-Rukai and Maga forms can also give us insight into the direction of vowel-matching changes (i.e. whether they were based on the stem or negative allomorph). Given a protoform like **samitu*, where V1 and V2 mismatch, there are two ways it could be reanalyzed as vowel-matching, schematized in (25). One option would result in the paradigm [smitu]∼[i-k-simtú:], with the UR /simitu/ (in this example, and examples below, the restructured allomorph is shaded in). This is a case of restructuring based on the stem allomorph, as speakers are extending the vowel that is present in the stem allomorph to the negative allomorph. The opposite scenario, where speakers extend the vowel present in the negative allomorph to the stem allomorph, would result in the paradigm [smatu]∼[ik-samtu:].

(25) *Hypothetical directions of vowel-matching changes for *samito*

		UR	stem	negative
a.	Expected Maga form	/samitu/	smitu	i-k-samtuː
b.	Restructured from stem base	/s**i**mitu/	smítu	i-k-s**i**mtuː
c.	Restructured from negative base	/sam**a**tu/	sm**a**tu	i-k-samtuː

Table (26) below summarizes the observed directions of vowel changes, with representative examples for each direction. Overall, restructuring has largely been based on the stem allomorph; this accounts for 45/56 (≈80%) of all vowel changes observed. A much smaller number of changes (N = 6, 0.11%) were based on the negative allomorph. Finally, for a few paradigms, both V1 and V2 have changed, making it unclear whether restructuring was from the stem or negative allomorph.

(26) *Directions of vowel-matching reanalysis*

Base	N (P)	Proto-Rukai	Expected UR	Actual UR
Stem	45 (0.80)	*s**a**mito	/s**a**mitu/	/s**i**mitu/ (smitu∼ik-s**i**mtuː)
Negative	6 (0.11)	*sin**aw**	/sin**o**sino/	/sin**i**sino/ (u-sn**i**sno∼i-sinsinoː)
Unclear	5 (0.09)	*t**o**l**a**rə	/t**u**l**a**rɨ/	/t**a**l**u**r**a**/ (tl**ura**∼i-k-t**a**lr**aː**)

Note that in many dialects of Rukai (Mantauran, Tanan, Labuan, Budai), there is a process described by Zeitoun (2007, 2024) as anticipatory copying triggered by reduplication, where CiCo sequences are reduplicated sa CiCi-CiCo. For example, Mantauran *la-maʔ-ivoko* '(male) friends' has reduplicated form *la-maʔ-ivi-ʔivoko* (Zeitoun, 2007, p. 29). The vowel-matching alternations in Maga are unlikely to be a result of this i-copying process, as only one case of vowel-matching fits the structural description for i-copying: /sin**i**sino/ [u-snísno]∼[i-sinsinóː].

4.4 Interim summary: vowel-matching

In comparing Proto-Rukai and Maga Rukai, we find that an earlier stage of Rukai (represented by Proto-Rukai) already had a statistical tendency towards vowel-matching. In Maga Rukai, specifically for stems that require composite URs, forms were restructured in a way that increased the degree of vowel-matching. In other words, speakers appear to have restructured forms in a way that reduced the need for composite URs. This vowel-matching pattern is particularly interesting in that while it does not remove the opaque syncope alternations, it does remove the need for composite URs. This suggests that vowel-matching changes were really motivated by the avoidance of abstract URs, rather than by other pressures such as paradigm uniformity or the avoidance of opaque patterns.

In comparing the directions of restructuring, I also find that most vowel changes have been based on the stem allomorph; in other words, the vowel quality present in the stem allomorph was extended to the negative allomorph. Restructuring in the other direction (negative→stem) was attested, but much less frequent. Taken together, these results suggest that speakers can learn URs of type KK-C (choosing among allomorphs), there is a bias to learn URs that match the unsuffixed stem allomorph, in line with predictions of KK-B. Note that these results actually contradict a well-supported variant of KK-B, Albright's single surface-base hypothesis, which predicts restructuring towards the overall most informative allomorph, which for Maga is the negative allomorph. In Section 7 below, I discuss potential reasons for why my results do not line up with the predictions of the single surface-base hypothesis.

5 Paradigm leveling

In this section, I discuss cases where paradigms appear to have leveled towards either the stem or negative allomorph, removing ∅∼V alternations. Rather than direct comparison with Proto-Rukai (as was done in Section 4 above), direction of leveling is inferred from the syncope pattern found in modern Maga. For example, given the word /simitu/ 'lips', there are three potential syncope patterns (not leveled, leveled to stem, leveled to negative) given in (27); where leveling has occurred, the vowel that should have deleted is underlined, and vowels that were exceptionally deleted are shown as '∅'. If we observe a paradigm like (27a), where the negative allomorph does not show the regular syncope pattern, and instead has the same syncope pattern as the stem allomorph, this implies that leveling towards the stem allomorph has occurred.

(27) *Types of leveling for 'lips' /simitu/*

	Pattern	Paradigm
a.	Not leveled (regular syncope)	[smítu] ∼ [i-k-símtuː]
b.	Leveled to stem	[smitu] ∼ [i-k-s∅mituː]
c.	Leveled to negative	[sim∅tu] ∼ [i-k-smituː]

Recall from Section 3.3 that Maga paradigms can be grouped in terms of whether they are predictable from one (or both) surface allomorphs. For example, /tupe/ [tpéː]∼[i-k-tupéː] 'gourd dipper' is predictable from the negative allomorph, which contains both underlying vowels. With this in mind, we can categorize paradigms along two dimensions: the original base (i.e. predictable allomorph prior to leveling), and the direction of leveling. Table (28) provides examples of each type of leveling; the UR listed in each example is the hypothetical UR that we would expect, if the syncope pattern were regular.

For example, /udalɨ/ 'rain' represents a case where prior to leveling, the paradigm was predictable form the stem, and leveling towards the stem has occurred. The stem [událɨ] is a predictable base in that all three vowels surface. The 'regular' negative form should be [i-udlɨ́ː], but instead, it has leveled towards the stem and is now [i-udalɨ́ː]. On the other hand, /ɨdɨpɨ/ 'quench' is also predictable from the stem allomorph, but has instead leveled towards the negative allomorph. As a result, the isolation stem surfaces as [ɨ́ɖpɨ], with the syncope pattern of the negative form.

(28) *Examples of leveling in Maga*

Base	**Level to**	**UR**	**Expected SRs**	**Actual SRs**	**Gloss**
stem	stem	/ud**a̲**lɨ/	událɨ∼i-udlɨ́ː	událɨ∼i-udalɨ́ː	‘rain’
	negative	/ɨd**ɨ̲**pɨ/	ɨɖɨ́pɨ∼i-ɨɖpɨ́ː	ɨ́ɖpɨ∼i-ɨɖpɨ́ː	‘quench’
negative	stem	/k**u̲**cia/	u-kcyáː∼i-kucyáː	u-kcyáː∼i-kcyáː[a]	‘scissor’
	negative	/t**e̲**θo/	tθóː∼ik-teθóː	téθo∼i-k-teθóː	‘turnip’
neither	stem	/b**u̲**lavanɨ/[b]	blávnɨ∼ik-bulvanɨ	blávnɨ∼ik-blavnɨː	‘silver’
	negative	/θil**i̲**bi/	u-θlíbi∼i-θilbíː	u-θílbi∼ i-θilbíː	‘fly’

[a]The negative of ‘scissor’ does actually have a regular variant [i-kucyáː]

[b]The first vowel of this (hypothetical, pre-leveling) UR is inferred from Proto-Rukai *bolavanə.

Additionally, for a subset of forms that would otherwise require composite URs (i.e. neither allomorph is a fully informative base), one or both allomorphs fail to undergo syncope at all. As a result, the the UR is predictable from at least one allomorph. Examples of this are given in (29), and the vowel(s) that should have deleted are underlined.

(29) *Forms where syncope exceptionally fails to apply*

Form w/o syncope	**Paradigm**	**Gloss**
Stem	ma-r**i̲**múru∼i-k-rimruː	‘forget’
Negative	plɨ́ŋɨ∼ik-pɨl**ɨ̲**ŋɨ́ː	‘ghost’
Both	l**u̲**bɨlɨ∼i-k-lub**ɨ̲**lɨː	‘mouth harp’

Table (30) summarizes the proportion of forms that have been leveled; the row titled ‘no syncope’ shows cases where syncope exceptionally failed to apply in either one or both allomorphs. Examples from (28) are given again; where leveling has occurred, the vowel that should have deleted is underlined, and vowels that were exceptionally deleted are shown as ‘∅’.

First, as expected, non-alternating forms (i.e. where the original base is both allomorphs) have not undergone any change. For forms that were predictable from one allomorph, there has generally been leveling towards the predictable allomorph. For example, paradigms that were predictable from the negative allomorph almost always leveled towards the negative allomorph, if they were leveled (30f; n = 69). In contrast, there is only one example of leveling towards the stem allomorph (30e).

For the composite UR forms (i.e. ones predictable from neither allomorph), there has generally been less leveling (N = 61, 13%).[4] Where leveling has occurred, it has happened towards both stem and negative allomorph, with a bias towards the stem allomorphs.

[4]Note that this is also the subset of forms that underwent vowel-matching restructuring.

(30) *Patterns of leveling*

	Orig. base	Leveled...	N	P	Example	
a.	both	not leveled	62	1	ábu∼i-abúː	'ash'
b.	stem	towards stem	32	0.37	událɨ∼i-ud**a̲**lɨ́ː	'rain'
c.		towards negative	2	0.02	ɨ́ɖ∅pɨ∼i-ɨɖpɨ́ː	'quench'
d.		not leveled	51	0.61	amíci∼i-k-amcíː	'tree root'
e.	negative	towards stem	1	0.01	u-kcyáː∼i-k∅cyáː	'scissor'
f.		towards negative	69	0.42	t**é̲**θo∼i-k-teθóː	'turnip'
g.		not leveled	94	0.57	tpeː∼i-k-tupéː	'gourd dipper'
h.	neither	towards stem	34	0.07	blávnɨ∼i-k-b∅l**a̲**vnɨː	'silver'
i.		towards negative	27	0.06	u-θ**í̲**l∅bi∼i-θilbíː	'fly'
j.		no syncope	97	0.20	l**u̲**bɨlɨ∼i-k-lub**ɨ̲**lɨː	'mouth harp'
k.		not leveled	322	0.67	ɖmarɨ∼i-k-ɖamrɨː	'moon'

In sum, forms show leveling towards both the stem and negative allomorphs, supporting KK-C levels of representation (i.e. choosing among allomorphs). Additionally, leveling appears to be driven in part by informativeness; forms that are predictable from one of the surface allomorphs have almost exceptionlessly leveled towards that allomorph. This is in line with the existing literature (see review in Section 3) on base-driven leveling. For the composite UR cases, there is a slight preference for leveling towards the stem allomorph. Notably, the composite UR forms have also undergone less leveling ($P = 0.13$, or 0.33 if non-syncope are counted) than other forms ($P = 0.39$ for stem-base paradigms; $P = 0.33$ for negative-base paradigms).

5.1 Alternative accounts to leveling

5.1.1 Syncope as the innovative variant

Chen (2008) puts forth the idea that syncope is a relatively innovative and recent development, and that paradigms which fail to undergo syncope actually reflect an earlier stage of Maga. Under this account, a case like (30b) could be treated as failure of the negative allomorph to undergo syncope, rather than leveling of the paradigm towards the stem allomorph. Similarly, cases like (30f) can be treated as failure of the stem allomorph to undergo syncope. More generally, if we adopt this idea, then around 75% of the leveling cases in (30) can be re-interpreted forms that reflect a conservative stage of Maga, and have not undergone syncope.

This account is less satisfactory for several reasons. First, it does not explain all cases of leveling, as around 25% of cases discussed in (30) cannot be explained as the failure of stems to undergo syncope. In particular, in cases of leveling where the original paradigm was predictable from neither allomorph (30h,i), the surface forms still undergo syncope and cannot reflect an earlier pre-syncope stage of Maga.

Additionally, under this account, there is no straightforward explanation for why syncope fails to apply at a much higher rate specifically where there is a predictable base in the paradigm. That is, why should syncope fail to apply for negative allomorphs specifically when the stem happens to be a predictable base (30a), and for the stem allomorph specifically when the negative allomorph is a predictable base (30f)? Under a leveling analysis, these patterns straightforwardly follow if leveling is driven by informativeness.

Finally, comparison of more recent data and Hsin's (2000) corpus suggests that leveling (or failure to undergo syncope) is the innovative variant. The *Austronesian indigenous words and narrations*, offered by the Indigenous Languages Research and Development Center (ILRDC), is an online corpus made for educational purposes (2022). Henceforth I will refer to this source as the ILRDC corpus. The data were recorded from 2020-2022, almost 20 years after Hsin's corpus, in consultation with a younger generation of speakers. In other words, the ILRDC corpus should reflect a relatively more recent stage of Maga. If we compare the two data sources, there are many examples where syncope applied in the Hsin corpus, but fails to apply to the ILRDC dictionary.

Examples of such mismatches are given in (31). The ILRDC corpus does not present words in morphological paradigms, and does not contain the negative allomorph of most words. As such, I focus on comparing the stem forms, but the negative allomorph from Hsin (2000) is also given in parentheses for reference. In examples (31a-c), the ILRDC stems don't undergo the expected vowel syncope, and also have the same vowels as the negative allomorph given in Hsin (2000). In (31d-e), the ILRDC stems actually show the syncope pattern of the negative allomorph, suggesting that there has been leveling towards the negative allomorph.

(31) *Comparisons between Hsin (2000) and the ILRDC web corpus*

	ILRDC CORPUS	HSIN (2000)	GLOSS
a.	s**u**roː	sróː (cf. i-k-suroː)	'grass'
b.	u-l**u**ŋéː	u-lŋéː (cf. i-luŋéː)	'siwm'
c.	u-t**ú**ta	u-ttáː (cf. i-tutaː)	'vomit'
d.	c**ɨ́**ŋ∅lɨ	u-cŋɨlɨ (cf. i-cɨŋlɨː)	'look'
e.	ɖ**í**k∅si	u-ɖkísi (cf. i-ɖiksiː)	'squeeze'

5.1.2 Historical explanations for failure to undergo syncope

Vowels that were historically long in Proto-Rukai have resisted syncope in Maga. As a result, apparent exceptions to the syncope pattern, as described in (30), may actually be an artifact of historic long vowels, rather than the result of base-driven leveling. In this section, I show that historic long vowels only account for a small proportion of cases where syncope fails to apply, and cannot fully explain my results.

There are two sources of long vowels that resisted syncope in Maga. First, Proto-Rukai had diphthongs which monophthongized in Maga (e.g. *ɭikol**aw** → rkúl**o**); more examples are given above in Section 4.1. In addition, Proto-Rukai *ʔ was uniformly deleted in Maga, and as a result *VʔV sequences became either a short monophthong V or bimoraic Vː (e.g. *aboŋ**oʔo** → abúŋ**u** ‘ant’); presumably, deletion of *ʔ resulted in an intermediate stage of Maga with VV vowels. Representative examples are given in (32); words like [d**a**nɨ] ‘house’ don’t undergo the expected syncope, but this falls out from the fact that [a] is derived from *aʔa.

Of the Maga Rukai words where we have available protoforms, historic long vowels block syncope in just 39/109 stems, and 4/6 negative forms.[5] In other words, most cases of exceptional non-syncope *cannot* be traced back to historic long vowels; examples of forms like this, where a non-deleted vowel corresponds to a historic short vowel, are given in (33). In total, only 39% (N = 43/115, P = 0.37) of non-syncope have a historic explanation.

(32) *Maga reflexes of Proto-Rukai diphthongs and *VʔV*

Proto-Rukai	Maga	gloss
*d**aʔa**nə	d**a**nɨ (dnɨː)	‘house’
*p**ay**so	p**e**su (psuː)	‘money’
*b**əʔə**kə	b**ɨ**kɨ (bkɨː)	‘pig’

(33) *Examples of Proto-Rukai short vowels that fail to undergo regular syncope*

Proto-Rukai	**Maga**	**Expected Maga**	**Gloss**
*d**o**koʔo	d**u**ku	d∅kuː	‘grow (plant)’
*p**ə**kə	p**ɨ**kɨ	p∅kɨː	‘house lizard’
*ma-r**i**moro	ma-r**i**múru	ma-r∅muru	‘forget’

5.2 Picture so far

Examination of Maga paradigms (and comparison of these paradigms with Proto-Rukai) suggests that two types of restructuring have occurred, both of which have reduced the need for composite URs. First, leveling has removed ∅∼V alternations in many paradigms, making them amenable to KK-A levels of analysis. Table (34) compares the degree of abstractness required if we assume regular application of syncope (as discussed in Section 3.3) against the degree of abstractness in Hsin’s corpus. If we take the regular syncope pattern to represent Maga pre-restructuring, then there has been an overall shift towards non-alternation, with a 24% increase in forms that are predictable from both allomorphs. Composite UR forms have seen the largest decrease (by 20%), supporting a view where learners are biased against learning composite URs (i.e. KK-D representations).

[5]This count does not include Proto-Rukai forms where the long vowel is not in a non-syncope position, e.g. *capəʔə→cpɨ́ː.

(34) *Comparison of UR abstractness before vs. after restructuring*

Predictable allomorph	KK-level	Before	After	Change
both	A	0.08	0.30	+0.24
stem	B/C	0.09	0.07	−0.02
negative	B/C	0.22	0.22	0
neither	D	0.61	0.41	−0.20

Aside from leveling, forms have also been restructured in a way that increased the rate of vowel-matching alternations. If we take vowel-matching forms to be predictable from just one allomorph, then an even smaller degree of abstractness is required in the lexicon. Table (35) shows the degree of abstraction required in the Maga lexicon if we account for vowel-matching; the row labeled 'either (vow matching)' represents paradigms that should need composite URs, but have alternating vowels that match. Now, only 16% of forms still require a composite UR.

(35) *Degree of abstractness required in Maga lexicon (assuming that vowel-matching forms do not require composite URs*

Predictable allomorph	KK-level	N	P
both	A	238	0.30
stem	B/C	57	0.07
negative	B/C	173	0.22
either (vow matching)	B/C	196	0.25
neither (no matching)	D	126	0.16

Overall, Hsin's corpus suggests that over time, paradigms have been restructured to result in less abstract URs. Additionally, the direction of restructuring suggests that when the two allomorphs are equally informative, there is a strong tendency to level towards the stem allomorph. When one allomorph is more informative than the other, then leveling tends to occur towards the more informative allomorph. This direction of leveling suggests that while URs are amenable to KK-C levels of representation (choosing from allomorphs), there is a bias towards learning URs that match the unaffixed allomorph (KK-B level of representation).

One issue that has not yet been addressed is the fact that the lexicon reflects two restructuring strategies: vowel-matching and leveling. Why should both strategies exist, and did they occur in parallel? It is possible that they reflect different stages of diachronic change in Maga, or that they reflect different pressures on paradigm structure. In the following section, I present data from a newer generation of speakers, whose strategies for paradigm restructuring can provide insight into these questions.

6 Restructuring in a newer generation of speakers

This section reports results from fieldwork with three Maga Rukai speakers. To preview the results, I find that these speakers have leveled paradigms to a greater degree than the speakers represented in Hsin's corpus. This suggests that leveling is the more innovative strategy; in the following section, I tentatively propose a frequency-based account for the two diverging restructuring strategies.

6.1 Speaker background

Data collection was carried out by the author in Maolin District, Kaohsiung, Taiwan over the course of one week in June 2024. Three native speaker consultants, ages 54-61 at time of data collection, were consulted. They represent a younger generation of speakers compared to the two consultants that Hsin (2000) worked with, who at the time of data collection (1999-2000) were both 74 years old.

Consultant details are given in (36). S1 reports speaking Maga regularly with elders (because of his involvement with the District Office, and with church duties). S2 and S3 speak Maga less regularly, and S3 is married to a Tona speaker; while Tona and Maga are mostly mutually intelligible, the two dialects have some vocabulary differences. Unlike Maga, Tona has not undergone vowel reduction/deletion or monophthongization.

(36) *Consultant details*

SPEAKER	GENDER	AGE	NOTES
S1	M	61	speaks Maga regularly with elders
S2	M	56	
S3	F	54	is married to a Tona speaker

The Maga spoken by younger speakers is different in several ways from what is reported in earlier work. One relevant difference is the negative prefix is no longer used productively with most nouns. Instead, speakers use lexical negation strategies. For example, while Hsin (2000) reports the negative form for 'mud' to be [i-k-dukcɨ], my consultants uniformly judged this as unacceptable, and instead say [ikmani dkacɨ] (where [ikmani] is a negation word). This means that overall, speakers see more words in just their unaffixed stem form, and consequently get less input for the rhythmic syncope pattern.

6.2 Data collection

Consultants were asked to provide the citation form (unaffixed stem), negative form (/i-/ prefix), and imperative form of (/-a/) of items from a wordlist. The imperative form has the same syncope pattern as the negative, and was included to check if both forms show a consistent syncope pattern.

Items were were presented in Chinese orthography using Microsoft PowerPoint; each word was given in its own slide, and accompanied by an illustrative photo.

The wordlist consisted of 75 verbs reported to have a regular syncope pattern in Hsin's corpus. Words were restricted to verbs because, as stated above, the negative prefix is only productive for verbs. To the extent possible, items were selected to represent the following types of paradigms: ones predictable from the stem (n = 20), from the negative allomorph (n = 25), and from neither allomorph (n = 30). Data were limited because items had to be both verbs and regularly alternating. Because of this, there are less items in the first two categories.

6.3 Results: patterns of restructuring

Overall, when speakers restructured a form, it was always through leveling, not vowel-matching alternations. Table (37) summarizes the number of leveling responses by speaker and base type. Note that for some items, speakers would give a null response; because these were omitted from Table 937), the TOTAL column varies by speaker. The number of null responses per speaker is as follows: S1 = 4, S2 = 12, S3 = 11.

The first trend we can observe is that there has been the most leveling in paradigms predictable from the stem allomorph, and conversely the least leveling in paradigms that are predictable from the negative allomorph. There is also a strong preference for leveling towards the stem. For example, consider row (37c); for S1, for the forms that were predictable from neither base (i.e. composite UR), 15/18 (83%) of leveled forms leveled towards the stem allomorph. In fact, in contrast to the corpus results in Table (30), even forms with a negative base have sometimes leveled towards the stem allomorph.

The stronger tendency towards leveling towards stem allomorphs could be because there is a general bias for leveling towards isolation allomorphs (i.e. KK-B). However, it could also be a consequence of usage frequency; if Maga speakers are using lexical negation more overall (vs. prefixal negation), then the stem allomorphs would be attested at higher rate than the negative allomorph, and therefore preferred as a base for leveling.

(37) *Proportion of leveled paradigms by speaker and base type*

	SPEAKER	BASE	N LEVELED.	TOTAL	P	LEVELED TO STEM
a.	S1	stem	9	16	0.56	8 (0.89)
b.		negative	9	24	0.38	3 (0.33)
c.		neither	18	31	0.58	15 (0.83)
d.	S2	stem	13	17	0.77	11 (0.85)
e.		negative	8	21	0.38	4 (0.50)
f.		neither	16	25	0.64	12 (0.75)
g.	S3	stem	14	15	0.93	14 (1.00)
h.		negative	9	23	0.39	5 (0.56)
i.		neither	18	26	0.69	17 (0.94)

Finally, the three speakers also differed in terms of their baseline rate of leveling. S1 leveled the least (P = 0.51), while S2 and S3 leveled more forms ($P_{S2} = 0.59$; $P_{S3} = 0.64$). This difference could also be related to usage frequency; S1 self-reports as someone who speaks Maga regularly, because of his work which involves speaking with older community members who don't understand Mandarin. S2 and S3, on the other hand, speak Maga less; this is especially true for S3, who is married to a Tona speaker. While drawing conclusions based on just three speakers is difficult, future work could potentially look more closely at the relationship between speakers' self-reported language use and their rates of paradigm leveling.

7 Discussion

7.1 Directions of restructuring

The direction in which paradigms have restructured in Maga can give us insight into which level of the KK-hierarchy speakers are able to entertain. The results suggest that when one allomorph is more informative than the other, then leveling tends to be informativeness driven. However, leveling can occur towards both the stem and negative allomorphs, not just towards the globally most informative allomorph. This suggests that learners are amenable to learning KK-C (choosing from allomorphs) levels of representations, rather than more restrictive levels like KK-B (fixed allomorph).

However, when the two allomorphs are equally (un)informative, there is a tendency to level towards the stem allomorph. This bias for leveling towards the stem allomorph actually contradicts a well-substantiated variant of KK-B, which posits that restructuring is based on the overall most informative allomorph (Albright, 2002b). In Section 3.3, we saw that the negative allomorph is overall most informative, and can predict about 22% (n = 230) of paradigms in Hsin's corpus. On the other hand, the stem allomorph only predicts 9% (n = 96) of paradigms.

There are some potential reason for why results do not line up with the predictions of the single surface-base hypothesis. First, although the negative allomorph is overall more informative than the stem allomorph, neither allomorph is very informative, and the majority of stems (over 60%) are predictable from neither allomorph. In contrast, most evidence for the single surface-base hypothesis comes from morphophonological paradigms that do not require positing composite URs.

Additionally, it could be that the stem allomorph is used more frequently than the negative allomorph. In fact, Albright (2008) also suggests that when one allomorph is used much more than other allomorphs, this can outweigh the effects of informativeness. This explanation is especially plausible for the newer generation of speakers, who use the negative prefix (and potentially other prefixes) less productively, and therefore see the isolation stem allomorph much more frequently.

7.2 Explaining divergent restructuring of abstract URs

The Maga lexicon, as represented by Hsin's corpus, reflects two restructuring strategies that have both moved Maga towards less abstract URs. Vowel-matching preserves the rhythmic syncope alternations, which are opaque. Leveling, on the other hand, removes a marked syncope alternation and results in non-alternating paradigms, which has been shown to be preferred by learners in various work (e.g. Benua 2004 on output-output correspondence; Kenstowicz 1996 on Uniform Exponence; Steriade 2000 on Paradigm Uniformity).

I tentatively propose that the strength of rhythmic syncope alternations is directly tied to its frequency in the lexicon, and this has resulted in the diverging restructuring strategies. That is, in an earlier generation of speakers, who spoke and heard Maga much more frequently, the rhythmic syncope generalization was learned more robustly. As a result, speakers restructured paradigms in a way that rendered them more concrete, while preserving the syncope pattern. The newer generation of speakers, on the other hand, leveled paradigms in a way that not just made them less representationally abstract, but also less marked in other ways.

This idea, that morphophonological representations are tied to frequency, is substantiated by a growing body of work (e.g. Bybee, 2001; Ellis, 2002; Albright and Hayes, 2003; Czaplicki, 2021). In particular, work such as Bybee (2001) has proposed that morphophonological generalizations emerge from the lexicon, and that the strength of these generalizations is directly tied to their token frequency. High-frequency patterns resist pressures from paradigm uniformity and opacity-avoidance, due to their robust representations in the grammar. On the other hand, low-frequency patterns are prone to leveling (Mańczak, 1980; Bybee, 2001).

In the case of Maga, it could be that as prefixes became less productive, speakers received less input for the rhythmic syncope pattern, resulting in a stronger tendency towards leveling. The same pressures could also apply at the level of individual lexical items; that is, items that are used often in both their prefixed and isolation forms maintain the syncope pattern, while those that are observed mostly just one allomorph will be more prone to leveling. This prediction isn't currently testable because there is little to no token frequency data available for Maga, but shoudl be tested in future work.

8 Conclusion

In Maga, a historical process of syncope in prosodically weak positions resulted in a synchronic rhythmic syncope pattern. The resulting Maga lexicon requires positing highly abstract composite URs. Paradigms were restructured in two ways – vowel matching and leveling of alternations – which have shifted the lexicon in concretist directions, reducing the need for positing composite URs. In particular, while leveling can be attributed to pressures like paradigm uniformity and opacity-avoidance, the vowel-matching pattern cannot. This suggests that avoidance of abstract representations, rather than other factors, is at least in part the cause of the restructuring that has happened in Maga.

The results of this paper suggest that while speakers can learn KK-C levels of representation, there is a bias towards leveling towards the stem allomorph (in line with predictions of KK-B). Future work should further quantify how strong this bias is, and whether it is attributable to usage frequency effects or a mroe fundamental preference for isolation forms. One way to do this would be with computational modeling; in recent years, there growing body of work on the task of UR learning (e.g. Tesar et al., 2003; Cotterell et al., 2015; Jarosz, 2015; O'Hara, 2017; Hua et al., 2020). More recently, Wang and Hayes (in prep) have utilzied such models to directly test the learnability of abstract underlying representations. Their approach could be applied to more directly quantify the degree of abstractness needed to explain the directions of restructuring that have happened in Maga Rukai.

Finally, the field data presented in Section 6 provides support for the idea that the robustness of morphophonological generalizations directly stems from their token frequency. This data is preliminary, and future work more carefully test if lexical frequency correlates with degree/type of restructuring. Obtaining frequency estimates for Maga is difficult given the state of language attrition that the language is facing. One workaround would be to obtain subjective frequency estimates from native speakers. Such estimates have been shown to predict behavioral measures just as well as corpus-based estimates (Balota et al., 2001).

References

Albright, Adam. A restricted model of ur discovery: Evidence from Lakhota. Paper given at the GLOW Phonology Workshop, Utrecht. `http://web.mit.edu/albright/www/papers/Albright-Lakhota.pdf`, (2002)a.

Albright, Adam. Explaining universal tendencies and language particulars in analogical change. In Good, Jeff, editor, *Linguistic universals and language change*, 144–184. Oxford University Press, (2008).

Albright, Adam (2010). Base-driven leveling in Yiddish verb paradigms. *NLLT* 28(3). 475–537. doi: 10.1007/s11049-010-9107-z.

Albright, Adam and Bruce Hayes (2003). Rules vs. analogy in English past tenses: A computational/ experimental study. *Cognition* 90(2). 119–161.

Albright, Adam C. (2002)b. *The identification of bases in morphological paradigms*. PhD dissertation, University of California, Los Angeles.

Balota, David A, Maura Pilotti, and Michael J Cortese (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & cognition* 29. 639–647. doi: 10.3758/BF03200465.

Beckman, Jill N (1998). *Positional faithfulness*. PhD Thesis, University of Massachusetts Amherst.

Benua, Laura. Transderivational identity: Phonological relations between words. In McCarthy, John J, editor, *Optimality theory in phonology: a reader*, 417–437. Wiley Online Library, (2004). doi: 10.1002/9780470756171.

Bowers, Dustin (2012). *Phonological restructuring in Odawa*. MA thesis, University of California, Los Angeles.

Bybee, Joan (2001). *Phonology and Language Use*. Cambridge University Press. ISBN 9780511612886. doi: https://doi.org/10.1017/CBO9780511612886.

Chen, Yin-Ling Christina (2008). On the Rhythmic Vowel Deletion in Maga Rukai. *Concentric: Studies in Linguistics* 34. 47–84.

Chomsky, Noam and Morris Halle (1968). *The sound pattern of English*. Harper & Row.

Cotterell, Ryan, Nanyun Peng, and Jason Eisner (December 2015). Modeling Word Forms Using Latent Underlying Morphs and Phonology. *Transactions of the Association for Computational Linguistics* 3. 433–447. ISSN 2307-387X. doi: 10.1162/tacl_a_00149. URL `https://direct.mit.edu/tacl/article/43278`.

Czaplicki, Bartłomiej (February 2021). The strength of morphophonological schemas: Consonant mutations in Polish. *Glossa: a journal of general linguistics* 6(1). ISSN 2397-1835. doi: 10.5334/gjgl.1255. URL `https://www.glossa-journal.org/article/id/5405/`.

Derwing, Bruce L. (1974). *Transformational grammar as a theory of language acquisition*. Number 10 in Cambridge studies in linguistics. Cambridge University Press, Cambridge, MA. ISBN 978-0-521-08737-7 978-0-521-09798-7.

Dresher, B Elan and Aditi Lahiri (1991). The germanic foot: metrical coherence in Old English. *Linguistic inquiry* 22(2). 251–286.

Ellis, Nick C. (June 2002). Frequency Effects in Language Processing: A Review with Implications for Theories of Implicit and Explicit Language Acquisition. *Studies in Second Language Acquisition* 24(2). 143–188. ISSN 0272-2631, 1470-1545. doi: 10.1017/S0272263102002024. URL `https://www.cambridge.org/core/product/identifier/S0272263102002024/type/journal_article`.

Hawkins, W Neil (1950). Patterns of vowel loss in macushi (carib). *International Journal of American Linguistics* 16(2). 87–90.

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

Hsin, Tien-Hsin (2000). *Aspects of Maga Rukai Phonology*. PhD Thesis, University of Connecticut.

Hsin, Tien-Hsin (2003). The mid vowels of Maga Rukai and their implications. *Journal of East Asian Linguistics* 12(1). 59–82. Publisher: Springer.

Hua, Wenyue, Adam Jardine, and Huteng Dai. Learning Underlying Representations and Input-Strictly-Local Functions. In *Proceedings of the 37th West Coast Conference on Formal Linguistics*, (2020).

Indigenous Languages Research and Development Center. Austronesian indigenous words and narrations. `https://web.klokah.tw`, (2022). [Online; accessed 9-July-2024].

Jarosz, Gaja. Expectation driven learning of phonology. Publisher: Citeseer, (2015).

Kager, René. Rhythmic vowel deletion in optimality theory. In Roca, Iggy, editor, *Derivations and constraints in phonology*, 463–500. Oxford University Press, (1997).

Kaohsiung City Government Civil Affairs Bureau. Population census. `https://cabu.kcg.gov.tw/web/HRStatReports.aspx?appname=HRStatReports`, (2023).

Kaplan, Max J. (2022). Stratal overgeneration is necessary: metrically incoherent syncope in Southern Pomo. *Phonology* 39(4). 597–639. doi: 10.1017/S0952675723000234.

Kenstowicz, Michael. Base-identity and uniform exponence: alternatives to cyclicity. In Durand, Jacques and Bernard Laks, editors, *Current Trends in Phonology: Models and Methods*, 363–394. Salford: University of Salford, (1996).

Kenstowicz, Michael J. and Charles W. Kisseberth (1977). *Topics in phonological theory*. Academic Press, New York. ISBN 978-0-12-405150-8.

Kiparsky, Paul. Linguistic universals and linguistic change. In Bach, Emmon W. and Robert Thomas Harms, editors, *Universals in Linguistic Theory*, 170–202. Holt, Rinehart, and Winston, (1968).

Kiparsky, Paul. Abstractness, opacity, and global rules. In Fujimura, Osamu, editor, *Three dimensions in linguistic theory*, Publications, 57–86. TEC, Tokyo, (1973). URL `https://books.google.com/books?id=YMaSOwAACAAJ`.

Kiparsky, Paul. Lexical morphology and phonology. In Yang, I.-S., editor, *Linguistics in the morning calm*, 3–91. Hansin, (1982).

Li, Paul Jen-Kuei (1975). Maga phonology: A preliminary report. *Bulletin of the Department of Archaeology and Anthropology, National Taiwan University* 37-38. 16–28.

Li, Paul Jen-kuei (1977)a. The internal relationships of Rukai. *Bulletin of the Institute of History and Philology* 48(1). 1–92.

Li, Paul Jen-kuei (1977)b. Morphophonemic alternations in Formosan languages. *Bulletin of the Institute of History and Philology, Academia Sinica* 48(3). 375–413.

Li, Paul Jen-kuei. Maga dialect. In Li, Paul Jen-kuei, editor, *The Formosan Languages in Kaohsiung County*, 8–118. Kaohsiung Prefecture Government, Kaohsiung, (1997).

Mańczak, Witold. Laws of analogy. In Fisiak, Jacek, editor, *Historical Morphology*, 283–288. De Gruyter Mouton, Berlin, New York, (1980). ISBN 9783110823127. doi: doi:10.1515/9783110823127.283.

McCarthy, John J (2008). The serial interaction of stress and syncope. *Natural Language & Linguistic Theory* 26. 499–546.

Muller, Jennifer Suzanne (2001). *The phonology and phonetics of word-initial geminates*. The Ohio State University.

O'Hara, Charlie (August 2017). How abstract is more abstract? Learning abstract underlying representations. *Phonology* 34(2). 325–345. ISSN 0952-6757, 1469-8188. doi: 10.1017/S0952675717000161. URL `https://www.cambridge.org/core/product/identifier/S0952675717000161/type/journal_article`.

Saillard, Claire. Is Maga accusative or ergative? evidence from case marking. In Tsao, Feng-fu and Mei-hui Tsai, editors, *Proceedings of the First International Symposium on Languages in Taiwan*. The Crane, (1995).

Steriade, Donca. Paradigm uniformity and the phonetics-phonology boundary. In Pierrehumbert, Janet and Michael van Broe, editors, *Papers in laboratory phonology V: acquisition and the lexicon*, 313–334. Cambridge University Press, Cambridge, (2000).

Tesar, Bruce, John Alderete, Graham Horwood, Nazarré Merchant, Koichi Nishitani, and Alan S Prince. Surgery in language learning. In *Proceedings of the Twenty-Second West Coast Conference on Formal Linguistics*, 477–490. Cascadilla Press, (2003).

Thurgood, Graham. Geminates: A cross-linguistic examination. In Nevis, Josh A., Gerald McMenamic, and Thurgood Graham, editors, *Papers in honor of Frederick H. Brengelman on the occasion of the twenty-fifth anniversary of the Department of Linguistics, CSU Fresno*, 129–139. Department of Linguistics, California State University Fresno, (1993).

Tranel, Bernard (1981). *Concreteness in generative phonology: evidence from French*. University of California press, Berkeley Los Angeles London. ISBN 978-0-520-04165-3.

Valentine, Randy (2001). *Nishnaabemwin reference grammar*. University of Toronto Press.

Zeitoun, Elizabeth (1995). *Issues on Formosan linguistics (Problèmes de linguistique dans les langues aborigènes de Taiwan)*. PhD Thesis, University Paris 7.

Zeitoun, Elizabeth (2007). *A grammar of Mantauran (Rukai)*. Institute of Linguistics, Academia Sinica Taipei.

Zeitoun, Elizabeth. Rukai. In Li, Paul Jen-kuei, Elizabeth Zeitoun, and Rik De Busser, editors, *Handbook of Formosan Languages (part 3): The Indigenous Languages of Taiwan*, 396–445. Brill, Leiden, The Netherlands, (2024). ISBN 978-90-04-52671-6. URL https://brill.com/view/title/63523.

A Section 4 supplementary figures

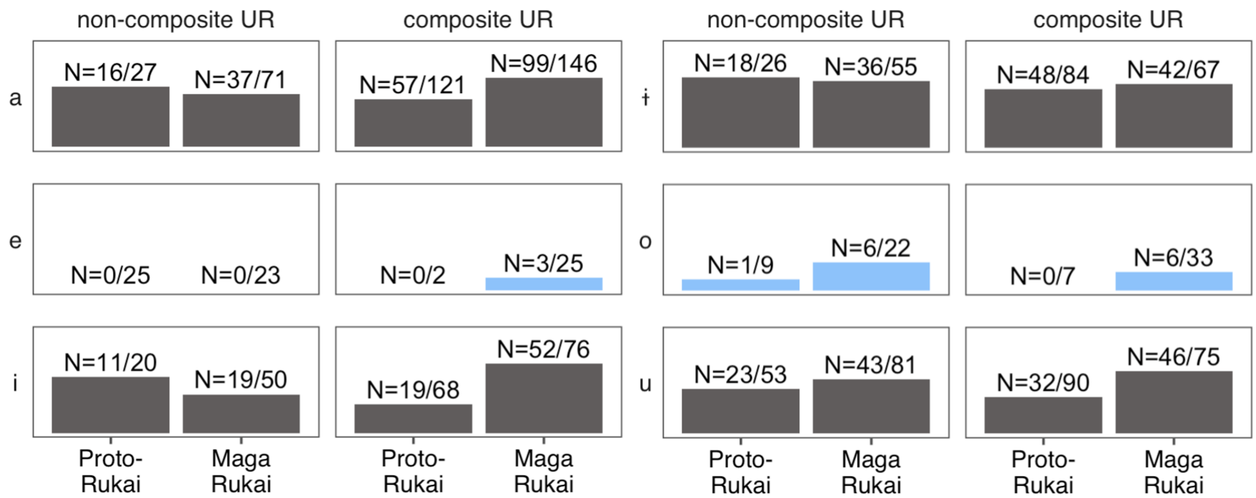

Figure 5: Proportion of vowel-matching in Proto-Rukai vs. Maga, by /V2/

B Monte Carlo test of significance: vowel-matching

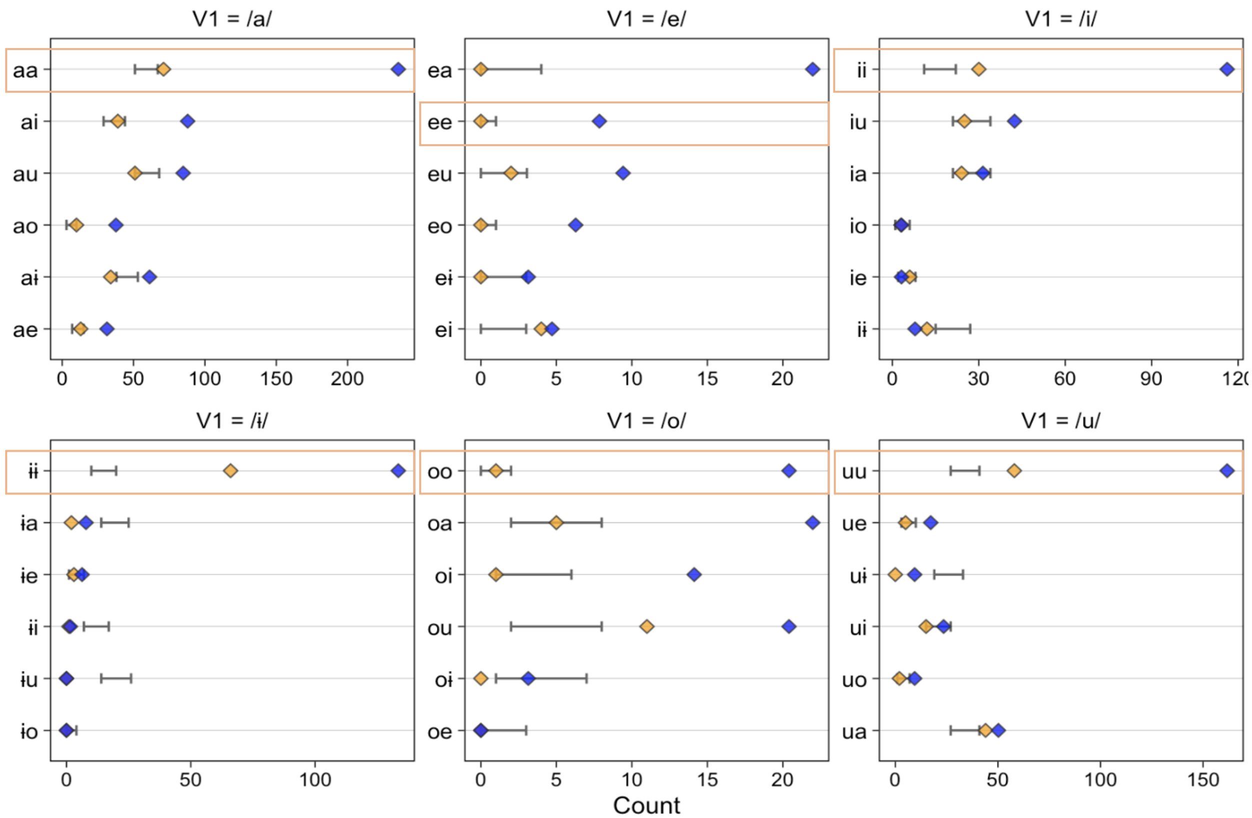

Figure 6: Monte Carlo test of significance: vowel-matching