Beyond trochaic shortening
a survey of Central Pacific languages

Kie Zuraw
kie@ucla.edu

UCLA Department of Linguistics
3125 Campbell Hall, Box 951543
Los Angeles, CA 90095-1543

Beyond Trochaic Shortening
a survey of Central Pacific languages

An underlying form like /ma:li/ is problematic for a stress system requiring word-final, bimoraic trochees. The grammar must sacrifice word-finality or bimoraicity, [(má:)li] or [(má:li)] (tolerating HL#); place stress on the second half of the long vowel, [ma(áli)] (breaking); or shorten the vowel, [(máli)] (trochaic shortening). This article surveys the Central Pacific language family, which hosts the most famous cases of breaking (Tongan) and trochaic shortening (Fijian), and finds that while trochaic shortening is poorly attested, breaking and tolerance are common. There are three findings of theoretical interest. First, length alternations suggest it is difficult to learn contrastive information that is absent in the core member of the morphological paradigm. Second, lexicalization of whole words is a possible response to this difficulty. Third, there is divergence between a language's root phonotactics and its alternations.[*]

Keywords: typology, stress, vowel length, trochaic shortening, breaking, Central Pacific, Polynesian

Beyond Trochaic Shortening: a survey of Central Pacific languages

1. INTRODUCTION: THE FACTORIAL TYPOLOGY OF TROCHAIC SHORTENING. Consider an underlying form ending in a heavy-light sequence, like /ma:li/. Footing and stressing /ma:li/ will be problematic in a language that requires a trochaic (stressed-unstressed) foot at the ends of words, as in *pa(táka)*. The table in 1 lists the compromises available, using the constraints defined in 2. Footing the entire word, as *(má:li)*, produces a foot with three moras. A homophonous alternative leaves the /li/ out of the foot, *(má:)li*, but now the foot is not aligned to the end of the word. The option of BREAKING (Poser 1985; Mester 1992), *ma(áli)*, yields a right-aligned, bimoraic foot, but is guilty of splitting underlying /a:/ over two syllables. There is also the option of TROCHAIC SHORTENING (Prince 1990; Hayes 1985; Hayes 1995), *(máli*). In terms of Optimality Theory (Prince and Smolensky 1993/2004), it is not possible to simultaneously satisfy FOOTBINARITY-mora, ALIGN(PWord, R; Foot, R), NOBREAKING, and faithfulness to length (as Kager 1999:175-177 discusses). The first two candidates can be said to tolerate the structure HL#, a heavy-light sequence at the end of the word.

(1) Options for underlying /…HL/ (i.e. ends with a heavy-light sequence)
<INSERT Tableau (1) HERE>

(2) Constraint definitions

- FOOTBINARITY-mora: a foot must have exactly two moras (in this case, that means one long vowel or two short vowels) (based on Prince & Smolensky 2004:47)
- ALIGN(PWord, R; Foot, R): the end of the phonological word must coincide with the end of a foot (McCarthy & Prince 1993)
- NOBREAKING: a single underlying vowel should not have surface correspondents in different syllables
- IDENT(long): an underlying vowel and its corresponding surface vowel should have the same length

This simple constraint system produces a factorial typology with three observably different languages, schematized in Table 1 (the column for $V_iÝ_i$, which refers to sequences like [aá], anticipates discussion below in section 2). All three languages treat an input like /pataka/ as [pa(táka)], but for /ma:li/ we can observe shortening, breaking, or tolerance of surface HL#.

Table 1 <INSERT Table 1 ABOUT HERE>

A language that opts for trochaic shortening should be free of HL# on the surface. Within a morpheme, there will also be no 'broken' sequence of the form $V_iÝ_i$ (e.g. [aá]; see section 2). The language should also display two types of alternation. First, shortening under suffixation: an unsuffixed form like /kuli:/ → [ku(lí:)] can form a right-aligned bimoraic foot unproblematically, but suffixed /kuli:-ŋa / → [ku(líŋa)] requires shortening to do so. Second, there should be 'de-shortening' under suffixation: underlying long penult vowels shorten, as in /ma:li/ → [(máli)], but if we add a suffix, /ma:li-ŋa/, the long vowel is pushed out of the penult and can now surface faithfully, as in /ma:li-ŋa/ → [(mà:)(líŋa)]. I use the term de-shortening atheoretically, to highlight how the derived form *(mà:)(líŋa)* differs from the unsuffixed form *(máli)*; the term is not meant to imply that the suffixed form is actually derived from the unsuffixed surface form (though section 4.1 considers that possibility)—by comparison with the underlying form, de-shortening simply represents no change.

A language with breaking should likewise lack surface HL#, but it will have $V_iÝ_i$, because of suffixation: a long final vowel surfaces intact,[1] as in /kuli:/ → [ku(lí:)], but adding a monomoraic suffix forces the stem-final vowel to be split over two syllables, as in /kuli:-ŋa / → [(kùli)(íŋa)]. Suffixation should also cause 'de-breaking': a penult vowel that is broken in the unsuffixed form, as in /ma:li/ → [ma(áli)], does not need to break if suffixation pushes it out of the penult, as in /ma:li-ŋa/ → [(ma:)(líŋa)]. Again, the term de-breaking is meant to highlight, atheoretically, the difference between the suffixed and unsuffixed forms.

A language that tolerates HL# should be free of shortening or breaking alternations, unless they are required for some other reason, and free of $V_iÝ_i$.

The Central Pacific language family hosts the most famous case of trochaic shortening, Fijian, and the most famous case of breaking, Tongan. In order to get a better idea of the typology of how underlying HL# surfaces, this article surveys the treatment, across Central Pacific languages, of words that underlyingly end heavy-light, either monomorphemically

(/...HL/) or through suffixation (/...H+L/). This language family gives good scope for comparison, because most of its members have right-aligned trochaic feet, and because long vowels have been repeatedly introduced into the family and its daughter branches, as discussed below. This survey includes all the Central Pacific languages where the needed data and generalizations were readily available in published descriptions. The discussions of Samoan and Tongan also draw on consultation of native speakers and exhaustive dictionary searches. For other languages, dictionaries were merely mined for supplemental examples, not systematically processed. Vowel-vowel sequences like *ai* sometimes have special behavior (e.g. Samoan [máile] 'dog', with stress on the antepenultimate vowel), and have therefore sometimes been analyzed as diphthongs—that is, as belonging to a single, heavy syllable ([mai]). But for most languages here, sources give little or no information on these sequences, and they will therefore be largely ignored below; the focus is on long vowels.

After some necessary background on syllabification and diachrony, subsequent sections examine first languages with trochaic shortening, then those with breaking, those that tolerate HL#, those with innovative stress patterns, and those that have lost contrastive length.

The overall picture that emerges from the survey does not support the clean factorial typology above. First, trochaic shortening does not seem to be impressively productive. The crucial data for Fijian are rather sparse (few crucial words' suffixed forms are given in the literature), and in Samoan, where more data are available, it appears that de-shortening alternations have been vulnerable to reanalysis as an arbitrary, morphologically conditioned length alternation. This suggests that alternations that neutralize the citation form can be fragile, in line with Albright 2002's theory that one member of the surface paradigm serves as a base from which the rest are derived—in this case, the citation form would be that base. When such an alternation degenerates, whole-word listing can occur (see Samoan in section 4.1, Tuvaluan in 5.2, and Tokelauan in 6.1 for cases of likely lexicalization). Breaking seems to be the best-attested system, in the sense that there are languages that closely match the set of behaviors predicted above. Simple tolerance is also well-attested, though languages that tolerate HL# may also retain de-shortening alternations, perhaps as lexicalized or morphologized phenomena rather than productive phonology.

Second, looking at the productivity and associatedness of phonotactics and alternations, and their diachronic stability, the data suggest, echoing Paster 2013, that alternations and

phonotactics do not go hand in hand as we expect in a classic conspiracy (Kisseberth 1970). Instead, phonotactics and alternations can match when initially created by the same diachronic change, but then drift apart (see especially Samoan in 4.1 and Rennell-Bellona in 5.2).

2. THE ANALYSIS AND TRANSCRIPTION OF LONG VOWELS. Researchers have varied in the structures they assume for long vowels in this language family—see especially Tongan, in section 5.1. In 1 above, a long vowel was shown as a single segment. But it could also be treated as a sequence of two identical short vowels, either in a single syllable (as in 3d) or in two separate syllables (as in 3e).

(3) Different possible representations for a long vowel
<INSERT Example (3) HERE>

If an underlying long vowel is treated as a sequence of two vowels that happen to be identical, then we can no longer rely on the faithfulness constraint NOBREAKING to explain why some languages prohibit [ma(áli)]. Instead, as illustrated in 4, we need a markedness constraint *$V_iÝ_i$ (and IDENT(long) is replaced by MAX-V); 4 is agnostic as to whether surface syllabification is as in 3d or as in 3e.

(4) Options for underlying /…HL/, sequences-of-short-vowels analysis[2]
<INSERT Tableau (4) HERE>

Even under the analysis where all long surface vowels are a single segment ([a:]), we still need *$V_iÝ_i$ to handle rich-base inputs like /maali/: if a language has no surface forms like [maáli], then faithfulness (NOBREAKING) is insufficient to explain the gap, and we need a markedness constraint.

For maximal generality, therefore, the rest of this article will transcribe long vowels as VV rather than V:, and tableaux will use *$V_iÝ_i$ (even though NOBREAKING would still be applicable if the input does contain a long vowel).

None of the languages here appear to have a contrast between underlying /V:/ and /V_iV_i/: that is, the distribution of [$Ý_iV_i$] versus [$V_iÝ_i$] (if it occurs) is always predictable. In terms of ranking, this means that *$V_iÝ_i$ is high-ranked in all but the breaking languages.

The markedness constraint *$V_iÝ_i$ is akin to two constraints proposed by Zuraw, Yu and Orfitelli (2014): *AÍ, '[a]n unstressed non-high vowel should not be followed by a stressed high vowel' (Zuraw, Yu & Orfitelli 2014:306), and *VALLEYASPEAK, '[a] stressed vowel must not be flanked by lower, unstressed vowels (i.e. a peak of stress should not be a valley of sonority)' (Zuraw, Yu & Orfitelli 2014:313). All of these constraints enforce a match between stress prominence and sonority prominence. *$V_iÝ_i$ penalizes sequences like [aá], where there is a stress difference but no sonority difference; *AÍ penalizes sequences like [aí], where the first vowel is more prominent in terms of sonority but less prominent in terms of stress; and *VALLEYASPEAK penalizes sequences like [aéa], where the middle vowel is more stress-prominent and less sonority-prominent than the vowels on either side. Other proposed constraints that connect sonority prominence and metrical or structural prominence include *PEAK/i,u (Kenstowicz 1997), *Δ_{Foot}/{ə,i/u} (de Lacy 2002a), and SONORITYPEAK (Clements 1997, Dell and Elmedlaoui 2002).

The terms 'break' or 'rearticulation' used by many descriptions to describe sequences like [aá], or, to use a real example, the [eé] in [(pèl**e)(é**-ŋa)] ('card game', Samoan), suggest something like a medial dip in amplitude, a change in formants, or a glottal stop. In the primary data used here from Samoan and Tongan, there was no such phonetic discontinuity. In both languages, stressed vowels bear a pitch rise (Zuraw, Yu & Orfitelli 2014 for Samoan, Garellek & White 2010 for Tongan). If that pitch rise, and therefore the stress, is on the second half of a long [e:]/[ee], an analysis of Samoan or Tongan as trochaic requires that the second half be foot-initial, which in turn requires us to posit a syllable (and foot) break: [e.é]. By contrast, in [pe(l**ée**)] 'play', the pitch rise, and therefore the stress, is on the first half of the long [ee], so the two halves are in the same foot, and could plausibly be either in the same syllable or in different syllables. In other words, there is no PHONETIC sense in which [ée] is monosyllabic/monosegmental but [eé] is disyllabic—that is, there is nothing between the two halves of [eé] like a dip in amplitude, change in formants, or glottal stop. Rather, it is a question of analysis: most analyses require [eé] to be disyllabic and allow [ée] to be monosyllabic. The same may well be true for the other languages with breaking surveyed here.

As much as possible, analyses in the case studies below will remain neutral as to how long vowels are treated. The abbreviation HL (heavy-light) will refer to both *má:li* and *máali*,

even though strictly speaking *máali* might be syllabified as *má.a.li*, which would be LLL (light-light-light).

Sources for the languages below vary in how they transcribe long vowels: *aa*, *a:*, or *ā*. Transcriptions have been standardized to *aa* here as much as possible.

Primary stress is marked with an acute accent (*á*), and secondary stress with a grave accent (*à*), whenever the source either marks it or states clearly where it should fall. Sources' segmental transcriptions were generally adjusted where drastically different from IPA.

3. DIACHRONIC BACKGROUND: WHERE LONG VOWELS COME FROM. Ethnologue (Ethnologue & Gordon 2005) classifies Central Pacific as a sub-family of Austronesian, containing 44 languages. The Central Pacific grouping includes Polynesian, East Fijian, West Fijian, and Rotuman languages, but excludes Loyalty Islands, Micronesian, New Caledonian, and North and Central Vanuatu languages. The appendix gives a table of the languages surveyed here and their affiliations.

The long vowels of Central Pacific arose mainly through multiple diachronic incidents of consonant deletion. Blevins (1994), using data from Geraghty (1990), gives the following examples in which *R (possibly a uvular fricative: Geraghty 1990:51-52) deletes between two identical vowels, yielding a long vowel. Additional examples of non-identical surrounding vowels are provided from Geraghty.

(5) Long vowels and vowel sequences through C deletion

	Proto-Eastern Oceanic		*Proto-Central Pacific*	
a.	examples from Blevins (1994:509-510) with identical vowels			
	*tak**uRu**-	>	*tak**uu**-[3]	'back'
	*tumb**aRa**	>	*tumb**aa**	'kind of land-crab'
	*mb**aRa**	>	*mb**aa**	'fence'
b.	examples from Geraghty (1990:58-60) with non-identical vowels			
	*ðak**aRu**	>	*ðak**au**	'coral reef'
	*ŋ**iRa**	>	*ŋ**ia**	'k. hardwooded shore tree'
	*tʃ**iRu**	>	*s**iu**	'make hissing noise'

Blevins notes some trochaic shortening that seems to date to this early stage—that is, cases where *VCV becomes a short vowel rather than a long vowel, in penult position.

(6) Early trochaic shortening and vowel sequences

	Proto-E. Oceanic		*Proto-Central Pacific*		
a.	examples from Blevins (1994:509) with identical vowels				
	*mb**uRu**a	>	*mb**u**(ʔ)a	(rather than *mbu(ʔ)ua*)	‘k. mollusc, *Murex sp.*’
	*k**aRa**va	>	*k**a**va	(rather than *kaava*)	‘k. fish, *Liza vaigiensis*’
b.	additional examples from Geraghty (1990:61,64,73) with identical vowels				
	*k**aRa**ka	>	*k**a**ka		‘k. crab’
	*m**aRa**ko	>	*m**a**ko		‘k. light-wooded tree’
	*m**aRa**ʔa	>	*m**a**ʔa- m**a**ʔa		‘lightweight’
	*t**aRa**ma	>	*t**a**ma		‘answer call’
c.	examples from Geraghty (1990:64,72-73) with non-identical vowels—no shortening needed				
	*m**aRe**re	>	*m**ae**re		‘k. vine’
	*s**oRa**ŋga	>	*s**oa**ŋga		‘k. banana’
	*t**aRe**Re	>	*t**ae**e		‘rejoice’

De Chene (2014) gives additional examples in which *y and *h have deleted, as seen in modern Tongan in particular.

(7) More examples of long vowels through C deletion (de Chene 2014, ch. 1)

Proto-Austronesian		*Tongan*	
*l**aya**R$_1$	>	la-l**aa**	‘sail, canvas’
*b**aha**q	>	f**aa**	‘flood’

Similar, later changes have taken place in individual languages or subfamilies, with consonant deletion introducing a new long vowel.

(8) Later C deletions

Tongan ***maa*** cf. Mele-Fila ***mara*** ‘rotten; preserved food’ (Ross, Pawley & Osmond 1998)

Hawaiian ***ʔoo*** cf. Tokelauan ***koho*** ‘coconut-husking stick’ (Ross, Pawley & Osmond 1998)

Samoan *fet**uu*** cf. Tongan *fet**u?u*** ‘star’ (Milner 1993 vs. Churchward 1959)

As we will see, loanwords and morphological lengthenings have introduced additional long vowels in some languages.

4. TROCHAIC SHORTENING. The first row of Table 1 is repeated here as a reminder of what is expected in the simplest trochaic shortening language (converting from V: to VV notation), along with a tableau.

Table 2 <INSERT TABLE 2 HERE>

(9) Trochaic shortening tableau
<INSERT Tableau (9) HERE>

4.1 SAMOAN. Most of the synchronic material in this subsection is condensed from one section of Zuraw, Yu and Orfitelli 2014—see there for more information, phonetic data, and an OT analysis, within a broader account of Samoan word prosody—but some is new. Data presented with stress transcribed are the pronunciations of the primary consultant in that study; the other sources cited here mostly do not mark stress. The discussion here ignores violations of ALIGN that can arise from certain vowel sequences (e.g. [máile] ‘dog’), because of higher-ranking *AÍ and *VALLEYASPEAK; these data are discussed in depth in Zuraw, Yu and Orfitelli 2014.

We will see that Samoan presents a varied picture. In part, it fits well with the schematic in 9: *HL# (a cover term for the FOOTBINARITY and ALIGN constraints used above) and *$V_iÝ_i$ phonotactic constraints are strong, maybe even gaining in strength over the past several decades. But Samoan deviates from the simple picture in two ways. First, de-shortening looks unproductive. And second, there is a phonotactics/alternation mismatch: while the *$V_iÝ_i$ constraint is strong in monomorphemes, and can block certain morphology, an alternation in suffixed words variably produces $V_iÝ_i$ sequences instead of shortening.

PHONOTACTIC BANS ON HL# AND V_iV_i. Milner 1993, a dictionary, contains some words spelled as ...V̄CV#, such as *tōga* 'fine mat' (*g* spells [ŋ]). Milner describes such words as 'heard [...] with a medial pulse of rearticulation' (Milner 1993:xvi), and with stress falling on the second half of the long vowel—[toóŋa]. Thus, these words would obey *HL#, and violate *$V_iV́_i$.

Our primary consultant pronounced all of these words, when known to him, with a short penult; examples are given in 10. This was true for monomorphemes (a), for a reduplicated word's first copy (which forms its own footing domain) (b), and for a non-final stem in a compound (which also forms its own footing domain) (c).

(10) Shortening as compared to dictionary

	Milner spelling	*consultant pronunciation*	
a.	monomorphemes[4]		
	ōi	ói	'groan'
	āfu	áfu	'sweat'
	nāmu	námu	'smell'
	tāne	táne	'husband'
	pāsi	pási	'pass' (English loan)
	pāga	páŋa	'partner' (English loan)
b.	reduplicated words: < > surround the first copy		
	pōlepole	<pòle>póle	'worried'
	vāivai	<vài>vái	'weak'
	vāevaeina	<vàe>vàe-ína	'divide'
c.	compounds		
	vālalua	vàla-lúa	'divided in twos (divide+two)'
	tāfafā	tàfa-fáa	'four-sided (side+four)'

Why are our consultant's pronunciations in conflict with Milner's spellings? Perhaps the language is in the midst of a change. The dictionary was compiled in the 1950s and 1960s (the 1993 edition used is a reissue of the 1966 original)—and even at that time, the pronunciations shown might have been conservative, given a dictionary's goal of preserving maximum

information. Comparing Pratt's dictionary from about a century earlier (Pratt 1878), we can see that some long penults of Pratt's are given as short by Milner: *ānu* (Pratt) vs. *anu* (Milner) 'to spit', *ālo* (Pratt) vs. *alo* (Milner) 'to paddle', *'ōa* (Pratt) vs. *'oā* (Milner) 'to lather'. So, by the 1960s some long penults had already shortened.

Mosel and Hovdhaugen (1992:30), writing three decades after Milner's dictionary was compiled, describe words with heavy penults and light ultimas as rare, and as varying in whether the stress is on the first half of the vowel ([táane], ordinary speech) or the second ([taáne], very careful speech). In OT terms, in this small set of words, there is variation in whether it is *HL# or *$V_iÝ_i$ that is violated. As Mosel and Hovdhaugen point out, the [taáne] variant agrees with Milner's description, and the [táane] variant with Pawley's (Pawley 1960). They note that Pawley focusses on the *tautala leaga* register of Samoan (also known as k-language), while Milner focusses on *tautala lelei* (t-language). *Tautala lelei* is used in school, church, prayer, song, and broadcasting, and for conveying social distance; *tautala leaga* is used in daily interactions outside of school or church, and in traditional oratory (see Mayer 2001, ch. 3). This accords with Mosel and Hovdhaugen's characterization of [taáne] as the more careful variant.

Around the same time, Condax (1990), in investigating final lengthening that marks locatives, measured vowel durations in Samoan words of various shapes (two speakers from the Apia area, born in 1942 and 1954). Condax reported difficulty finding suitable items of the shape CV:CV in Milner's dictionary, and recorded only *pāpa* 'barber' (English loan). For that one item, Condax found that the supposedly long penult vowel was 'much shorter […] than any of the other long vowels', 110 msec. for one speaker and 109 for the other (Condax 1990:39). This was closer to a short vowel than to a long vowel: stressed, short penult vowels (e.g. ***á**pa* 'plank') were on average 84 msec. for both speakers, and long penult vowels in words like *(**à**a)(áa)* 'high titles and dignities' (where each of the last two syllables can form a foot and thus trochaic shortening does not apply) were on average 151 msec. and 154 msec. for the two speakers. Condax does not report standard deviations or ranges, so we can't say whether the penult of *pāpa* is well within the range of ordinary short penult vowels or would be an outlier.

Hovdhaugen (1992) replies to Condax's paper with duration measurements from another speaker, but unfortunately reports only the category of long penult vowels (Hovdhaugen 1992:282), not specifying whether they come from words like *p**ā**pa*, where trochaic shortening predicts that they should be short, or from words like *p**ā**pā*, where they should remain long

(Hovdhaugen 1992:282), or a mix.[5] The long durations obtained (mean of 173 msec., ranging from 158 to 188; compared to 53-150 msec., mean of 108, for short, stressed penults) are therefore not informative for our purposes.

Our primary consultant, also from Apia, was born in approximately 1993, a generation or two later than Condax's speakers. (He had recently moved from Samoa to California, and thus increased his daily use of English; Condax's consultants were also bilingual, having lived in Hawaii for many years.) Another source of recent data is Billington's online Samoan vocabulary (Billington n.d.). It does contain four …HL# items (*āla* for plural of *ala* 'awake', *'ie tōga* 'fine mat', *mālōsi* 'strong' and derivatives, *mālōlō-ga* 'interval, break'). Unfortunately these are not among the items accompanied by audio recordings, and we shouldn't draw strong conclusions from the spellings, because Billington's 'How to contribute' page acknowledges that many errors remain in the materials, and in some of the items that do have audio recordings, the audio length does not match the spelled length (e.g. *lenei*, audio [leenei] 'this'; *lisiti*, audio [liisiti] 'receipt'; *mālie*, audio [malie] 'to agree').

As for reduplicated words (as in 10a), Hovdhaugen (1990:98) states that the length in Milner's entries such as *fānafana* 'is very hard to perceive and most Samoans seem not to find such vowel lengthening natural or acceptable', again suggesting an ongoing tendency away from tolerating HL#.[6]

*HL# CAN BLOCK REDUPLICATION. The *HL# restriction influences the choice of morphology to mark plural in verbs. Among verbs with a listed plural in Milner's dictionary (Milner 1993), the most common pattern is CV reduplication of the primary-stressed foot, as in [láfi] 'hide', [<la>láfi] 'hide-*plural*'. But if the primary-stressed foot is vowel-initial, CV reduplication fails to occur. For example, [óso] 'jump' does not have a plural *[<o>oso], presumably because the phonotactic constraints would be violated; Hovdhaugen (1990:102) makes a similar argument. Instead, 'jump' uses different plural morphology: *fe-oso-(f)i*.

Table 3, adapted from Zuraw, Yu and Orfitelli 2014, shows the range of options for marking plurality, with examples and counts from the Milner 1993 dictionary). As the table shows, 14 % of all verbs have a vowel-initial final foot (which would serve as the base of reduplication), but none[7] of the CV- reduplicated verbs have a vowel-initial final foot.

Table 3 <INSERT TABLE 3 ABOUT HERE>

Pre-Samoan deleted *h and *ʔ (and ʔ was reintroduced by *k > ʔ) (Elbert 1953). Assuming that Pre-Samoan had some reduplicated plurals like hypothetical *<ha>haki or *<ʔo>ʔofo, they must have fallen out of use after the glottal consonants dropped, because of the ill-formedness of *[aaki], *[oofo]. A hint that this did happen comes from the difference between Pratt 1878, which gives *āla* as the plural of *ala* 'to wake', and Milner 1993, which gives no plural. The Tongan cognate is *ʔaa*, suggesting a proto-form along the lines of *ʔara, which would have been unproblematically reduplicatable in pre-Samoan ([<ʔa>ʔara]), until it lost its glottal stop. Presumably there was an intermediate period in which reduplicated [<a>ala] was tolerated, and this is reflected in Pratt's dictionary, but then the reduplicated form disappeared. Similarly, Pratt gives *īti* as the plural of *itiiti* 'small' (this is one of a minority of words whose plural is formed by trading two-syllable for one-syllable reduplication), but Milner gives *iti* as the plural. Ross, Pawley, and Osmond (2007) reconstruct Proto-Oceanic *qitik, so again, presumably at some point reduplicated *<qi>qitik was unproblematic, until the *q deleted.

PRODUCTIVE SHORTENING UNDER SUFFIXATION. The prohibition on HL# is also enforced under suffixation. When a word ending in a long vowel receives a monomoraic suffix, such as the nominalizer /-ŋa/ or the ergative /-a/, our primary consultant always resolves the resulting /...H+L#/ sequence in some way. The most common outcome is shortening (as in 11a), but breaking also occurs (as in 11c), and for some words our consultant accepted both options (as in 11b).[8] Only primary-stress feet are shown here; see Zuraw, Yu and Orfitelli 2014 for data on and analysis of secondary stress.

(11) Shortening and breaking under suffixation (consultant data)

a. shortening

maaloo(l**óo**)	'rest [verb]'	maaloo(l**ó**-a)	'rest-ERG'
too(f**áa**)	'sleep (polite)'	too(f**á**-ŋa)	'bedding'
faʔa-o(t**áa**)	'ripen (bananas)'	faʔa-o(t**á**-ŋa)	'hung-up banana bunch'
tau-sini(**óo**)	'compete'	tàu-sini(**ó**-ŋa)	'competition'
(ʔ**ée**)	'yell'	fe-(ʔ**é**-i)	'shout-*pl*'
su(s**úu**)	'come/go (pol.)'	su(s**ú**-ŋa)	'Your Honour'

b. variation

pe(l**ée**)	‘cards’ (<*play*)	pe(l**é**-ŋa),	pel**e(é**-ŋa)	‘card game’
taali(**ée**)	‘laugh’	taali(**é**-ŋa),	taali**e(é**-ŋa)	‘laughing’
muu(m**úu**)	‘red’	muu(m**ú**-a),	muum**u(ú**-a)	‘red-ERG’
ʔoo(n**áa**)	‘drunk’	ʔoo(n**á**-ŋa),	ʔoon**a(á**-ŋa)	‘group of drunk people’
u(**óo**)	‘friend’	faʔa-u(**ó**-ŋa),	faʔa-u**o(ó**-ŋa)	‘making friends’

c. breaking

(p**óo**)	‘slap’	p**o(ó**-a)	‘slap-ERG’
ta-(p**ée**)	‘kill’	ta-p**e(é**-a)	‘kill-ERG’
pa(ʔ**úu**)	‘fall’	paʔ**u(ú**-ŋa)	‘falling

We analyze this variation as resulting from variable ranking of two constraints (or similar weighting, depending on the framework for variation),[9] as shown by the jagged line in 12: *V_iV_i, the markedness constraint that penalizes breaking, and a faithfulness constraint. The faithfulness constraint can’t be the MAX-μ or IDENT(long) used earlier (depending on whether the input contains /VV/ or /V:/), because they were outranked by *V_iV_i to ensure no breaking in monomorphemes. Instead we use DON’TSHORTEN-V́:-BaseAffixed: assign a violation if a short vowel in an affixed form (*pelé-ŋa*) corresponds to a long, primary-stressed vowel in the affixed form’s base (*pelée*). This assumes an approach as in Kenstowicz 1996, Benua 1997, and others, where an affixed word’s segments enter into a correspondence relation with the surface segments of the affixed word’s morphological base. It also assumes that this constraint can penalize shortening/deletion regardless of whether the base is represented with a single, long vowel ([pelé:]) or a sequence of identical vowels ([pelée]). In other words, it is the base-affixed version of MAX-V/IDENT(long).

(12) Options for underlying /…HL/, sequences-of-short-vowels analysis

<INSERT Tableau (12) HERE>

DE-SHORTENING UNDER SUFFIXATION: POSSIBLY UNPRODUCTIVE. There are many cases of de-shortening under suffixation. Some examples are shown in 13b, with their presumed underlying

forms; non-alternating examples are included in 13a for contrast. Again, the term de-shortening is used atheoretically to emphasize that the shortening that occurs in the unsuffixed surface form is absent from the suffixed surface form.

(13) De-shortening under suffixation (consultant data)

a. underlying short vowel—no alternation

/fusi/	(f**ú**si)	‘hug’	f**u**(sí-a)	‘hug-ERG’
/ʔini/	(ʔ**í**ni)	‘pinch’	ʔ**i**(ní-a)	‘pinch-ERG’
/lolo/	(l**ó**lo)	‘flood’	lo(l**ó**-fi)	‘surge’
/moe/	(m**ó**e)	‘sleep’	m**o**(é-ŋa)	‘bed’
/tao/	(t**á**o)	‘cover’	t**a**(ó-mi)	‘hold down’

b. underlying long vowel—de-shortening

/tuusi/	(t**ú**si)	‘write’	t**uu**(sí-a)	‘write-ERG’
/noofo/	(n**ó**fo)	‘stay’	n**oo**(fó-i)	‘colonize, settle’
/taaŋi/	(t**á**ŋi)	‘cry’	fe-t**aa**(ŋí-si)	‘cry-*pl*’
/laau/	(l**á**u)	‘say’	l**aa**(ú-ŋa)	‘speech’
/looi/	(l**ó**i)	‘ant’	l**oo**(í-a)	‘overrun w/ ants’
/saʔeeu/	sa(ʔ**é**u)	‘stir’	saʔ**ee**(ú-a)	‘stirring’

The de-shortened forms do not seem to reflect a long vowel in the proto-language, and their origin is uncertain. Corresponding to the words in 13b, Greenhill and Clark (2011) give Proto-Eastern-Malayo-Polynesian *tosi ‘draw’, Proto-Oceanic *nofo ‘stay’, Proto-Austronesian *Caŋis ‘cry’, Proto-Polynesian *lau ‘recite’, and Proto-Ellicean *roe ‘ant’. The long vowels in the suffixed forms could have begun as idiosyncratic lengthenings to accompany suffixation, and it is hard to know whether they ever were analyzed by speakers as reflecting underlying length (as in the underlying forms given in 13) with productive shortening, or have always been treated as idiosyncratic.

The contemporary synchronic analysis is also uncertain. On the face of it, the obvious analysis is contrastive underlying length, neutralized in unsuffixed forms, as shown in 12. But the existence of a number of doublets undermines this analysis: there are several stems with both a short-vowel suffixed form and a long-vowel (de-shortened) suffixed form. Mosel and

Hovdhaugen (1992:195-196) list 12 such examples for the suffix /-ŋa/ (and one where a different vowel lengthens). A selection follows.

(14) Length doublets from Mosel and Hovdhaugen 1992

unsuffixed	*suffixed*	
p**u**le ‘authorize, power, control’	p**u**leŋa	‘authority, power, control’
	p**uu**leŋa	‘London Missionary Society administrative unit’
t**i**pi ‘cut, slice’	t**i**piŋa	‘cutting’
	t**ii**piŋa	‘surgical operation’
t**o**so ‘pull, draw, drag’	t**o**soŋa	‘pulling’
	t**oo**soŋa	‘tug-of-war’
t**u**ʔu‘put, leave, give’	t**u**ʔuŋa	‘share (of food)’
	t**uu**ʔuŋa	‘race (e.g. of horses)’

Mosel and Hovdhaugen (1992:195) state that where there is a discernible meaning difference, the long-vowel (de-shortened) form ‘usually indicate[s] plurality or frequency’, but another way of characterizing the meaning differences is that the de-shortening form has a less-transparent meaning. In Zuraw, Yu and Orfitelli 2014 we speculate that words like [puuleŋa] have their own lexical entries, with long vowels and idiosyncratic meanings: /puuleŋa/ ‘London Missionary Society administrative unit’. The lexical entry for the stem may at one time have had a long vowel but no longer does (/pule/ ‘authorize’), and the productive, transparent suffixed form is built from that underlying form with a short vowel: /pule/+/ŋa/.

Zuraw, Yu and Orfitelli 2014 concludes that existence of an alternating pair such as [tóso], [toosó-ŋa] is not sufficient to guarantee that speakers learn to derive both from an underlying form /tooso/—even if phonotactics would support that analysis, by ensuring shortening in the unsuffixed form. Albright (2002) proposes that learners choose one surface allomorph to use as an underlying form, and that this must be the same member of the paradigm across all stems. In Samoan, the unsuffixed form would be the best choice, because it exists for nearly all stems, whereas many stems lack suffixed forms and thus would have no available underlying form. Albright’s model predicts that learners would then have to treat the length in [toosó-ŋa] as exceptional (through a diacritic, listing, very specific rule, etc.), making it vulnerable to diachronic change. A weaker position would be that it is possible to construct an

underlying form /tooso/, but it comes at some cost, and therefore is not achieved in all cases, again leaving some forms like [toosó-ŋa] as exceptional. In Samoan, there are unpredictable morphological or lexical length alternations, shown in 15, and these may have given learners a precedent for treating de-shortened words as having somewhat-arbitrary lengthening rather than reflecting underlying stem length (see also Hovdhaugen 1990).

(15) Examples of morphological or lexical lengthening from Mosel and Hovdhaugen (1992:78, 221, 239)

vaʔai	'sit'	vaaʔai	'sit-*plural*'	length marking plural
alofa	'love'	aa<lo>lofa	'love-*plural*'	length accompanying reduplication
ʔumi	'long'	ʔumii	'very long'	length marking emphasis

A further factor undermining the productivity of de-shortening could be that in the *tautala leaga* register of Samoan, '[l]ong vowels in antepenultimate syllables are frequently shortened' (Mosel and Hovdhaugen 1992:9). If this optional shortening applies equally to suffixed words, opportunities to learn that a certain a stem has an underlying long vowel (or is an idiosyncratic lengthener) are further reduced.

*$V_iV́_i$ AND BREAKING. The *$V_iV́_i$ phonotactic constraint is strong in that Samoan largely bans $V_iV́_i$ in monomorphemes. As reviewed above, our primary consultant produced no words like *[aáfu] or *[paási], with the pitch rise characteristic of Samoan stress happening on the second half of the long vowel, although previous descriptions noted such pronunciations, at least as one option, in a small number of words. ($V_iV́_i$ sequences are allowed across a prosodic-word boundary, as in compounds, some affixation, and two-syllable reduplication; see Zuraw, Yu & Orfitelli 2014.)[10] But, the phonotactic constraint is not strong enough to prevent breaking as an option under suffixation, as we saw above.

This particular mismatch between phonotactics and alternations poses no great challenge for an OT analysis: a markedness constraint can have an intermediate ranking. In this case, as was shown in 12, *$V_iV́_i$ ranks above MAX-V, so that in general $V_iV́_i$ is avoided, but *$V_iV́_i$ (variably) ranks below DON'TSHORTEN-V́:-BA, so that $V_iV́_i$ is tolerated in stem-final vowels.[11] But, the mismatch is not what we expect in a classic conspiracy (Kisseberth 1970), where surface-true phonotactic constraints drive alternations. The mismatch also makes it difficult for

child learners to use the strategy of first learning what is phonotactically legal and then applying the resulting grammar to learning alternations (Hayes 2004, Prince & Tesar 2004): a child should learn that *$V_iÝ_i$ is high-ranked (since it is violated in few words), but then, as she begins to discover morphology, she must learn that breaking is a good option in suffixation and construct the even higher-ranked constraint DON'TSHORTEN-Ý:-BA in response (consistent with Hayes's 2004 proposal that learners assume output-output correspondence constraints to be highly ranked).

To sum up, we have seen that Samoan observes fairly strong bans on HL# and $V_iÝ_i$, but that de-shortening nevertheless may be unproductive, and that the phonotactics and the alternations don't quite go hand in hand.

4.2 FIJIAN. Fijian has the best-known case of trochaic shortening/de-shortening. The pattern is similar to Samoan's, however, in that while the phonotactic restriction against HL# is robust, the evidence is weak that related alternations are productive.

Fijian words appear to obey both *HL# and (except across certain morpheme boundaries) *$V_iÝ_i$. Schütz (1985, ch. 36 & 37) describes Fijian words as ending in two light syllables or a heavy syllable.

As Dixon (1988:26), Schütz (1985:528), and Hayes (1995) discuss, verb-final long vowels shorten when a monomoraic transitive suffix is added, as in [(ð**áa**)] 'bad', [(ð**á**-ta)] 'consider bad'. Dixon lists only four verbs like this, all monosyllabic, and he discusses a possible analysis in which the underlying forms actually have a short vowel (/ða/), with lengthening to satisfy a bimoraic word minimum.[12]

Shortening in nouns is more widespread. Dixon's mini-dictionary lists 20 nouns with a final long vowel. According to Dixon's description, these should all shorten when a monomoraic pronomimal suffix is added (though the suffixed forms are not explicitly given). Of these, six clearly involve a root of more than one syllable, such as [ʔolíi] 'dog', so that length must be underlying and not just a repair for a subminimal root.

As for de-shortening under suffixation, Dixon's mini-dictionary lists only one example (shown in 16a); Scott (1948:743, fn. 1) identifies 'three unusual cases' of an optional or obligatory long vowel appearing in a suffixed form (shown in 16b). As shown by the column *Capell's transitive* in 16, none of these de-shortenings occur in Capell's dictionary (Capell's

1957), which does generally mark length. (Scott does not give glosses, so there is a possibility that the words found in Capell are different from the words intended by Scott.)

(16) De-shortening

a. from Dixon (1988:357-375)

intransitive	*transitive*	*Capell's transitive*	*gloss*
one example of de-shortening			
s**í**βi	s**ii**βí-ta	s**i**βí-ta	'pass'
vs. many examples of non-alternation			
s**í**βi	s**i**βí-ta	s**i**βí-ta, s**i**βí-a	'carve'
^{m}b**á**le	^{m}b**a**lé-ta	^{m}b**a**lé-ta	'fall'
ð**é**le	ð**e**lé-a	ð**e**lé-ka	'uncover'
l**ú**a	l**u**á-ða	l**u**á-ða	'vomit'
p**ó**no	p**o**nó-ʔa	*(no entry for* pono*)*	'catch'
and many more			

b. from Scott (1948:743)

'three unusual cases' of length alternation			
^{n}d**ó**nu	^{n}d**oo**nú-ja ~ ^{n}d**o**nú-ja[13]	^{n}d**o**nú-ja	'coincide'
l**é**βe	l**ee**βé-a ~ l**e**βé-a	l**e**vé-a	'dodge'
^{n}dr**á**βu	^{n}dr**aa**βú-ja	*(no suffixed form*)	'rub with ashes'
vs. typical example of non-alternating short vowel			
^{n}d**ó**la	^{n}d**ó**lá-βa	^{n}d**ó**lá-βa	'open'

In summary, Fijian obeys the *HL# and *$V_iÝ_i$ phonotactic constraints, placing it in the trochaic-shortening part of the typology, but there is only limited evidence for productive alternation, with the literature citing only two obligatory and two optional cases of alternation (besides those that could be explained away as minimality effects), and a lack of agreement among sources, which could reflect dialect differences and/or diachronic change.

4.3 TROCHAIC SHORTENING SUMMARY AND CONCLUSIONS. The two trochaic-shortening languages found in this survey largely adhere to the *HL# and *$V_iV́_i$ prohibitions, with (mostly) shortening under suffixation. But, the evidence for productive de-shortening is weakened by the existence of doublets in Samoan, and the paucity of data in Fijian.

A trochaic-shortening language with the full range of productive phenomena predicted remains elusive. The other cases typically cited are Tongan, Hawaiian, Middle English, Chamorro, some diachronic Italian, and Abruzzese (Prince 1990, Hayes 1995:148-149, Mellander 2005).

In Tongan, trochaic shortening is limited to a very few words, and could actually be lengthening rather than shortening (see 20 below). In Hawaiian, there was diachronic shortening in a specific consonantal environment (see 44 below).

The English shortening (which ignores an extrametrical final light syllable; Myers 1987) produces pairs like [sə(ɹí:n)] ‘serene’ vs. [sə(ɹɛ́nɨ)<ti>] ‘serenity’, where < > enclose the extrametrical material (the analysis as trochaic shortening is due to Prince 1990). Both the phonotactic constraint and the alternation are now of limited productivity, and so difficult to probe further.

Chamorro and diachronic Italian, in Prince’s analysis (Prince 1990), both involve failing to lengthen, rather than shortening (Chamorro facts from Chung 1983, Italian from Calabrese 1986). Stress is contrastive. As shown in 17a, when stress is on the penult, the penult must lengthen to provide a bimoraic foot, because the final syllable is extrametrical and cannot belong to the stress foot. But, as shown in 17b, when stress is on the antepenult lengthening is blocked: rather than supplying a bimoraic foot, lengthening would prevent achieving a binary foot that is aligned to the end of the word (not including the extrametrical final syllable).

(17) Chamorro and diachronic Italian have lengthening in penults but not antepenults

	Chamorro		*diachronic Italian*				
a.	a(lí:)<tus>	‘earrings’	Latin	pedem	>	Italian (pié)<de>	‘foot’
b.	(íga)<du>	‘liver’	Latin	medicus	>	Italian (médi)<ko>	‘physician’
	*(í:ga)<du>, *(í:)ga<du>					*(miédi)<ko>, *(mié)di<ko>	

Similarly, in modern Abruzzese dialects of Italian, stressed open penults have a long or diphthongized vowel (e.g. Agnone [(kéw)<rə>] ‘heart’), but other syllables, including stressed,

open antepenults, have a short, undiphthongized vowel (Fong 1979). Little data or discussion is available, though.

In summary, robust, productive trochaic shortening remains elusive.

5. BREAKING. As a reminder, a breaking language is expected to have the characteristics below.

Table 4 <INSERT TABLE 4 HERE>

(18) Breaking tableau
<INSERT Tableau (18) HERE>

5.1 TONGAN. Tongan is the best-known case of breaking. As illustrated in 19a, Tongan has long vowels (examples from Churchward 1953:3-4 and Churchward 1959). As in other languages, authors have varied in whether they treat a sequence like the [úu] in [patúu] as a long vowel belonging to a single syllable, or as two short vowels that happen to be identical. Taumoefolau (2002) argues for a disyllabic analysis; Garellek and White (2010) present phonetic evidence for a monosyllabic analysis. In the typology given in sections 1 and 2, it doesn't matter whether an underlying long vowel is one segment or two, nor whether a surface long vowel is one segment, two segments in the same syllable, or two segments in different syllables. All analyses can be made consistent with the fact that the pitch rise typical of Tongan stress (Kuo & Vicenik 2012; see Garellek & White 2015 for further phonetic correlates of stress) occurs on the first half of the long vowel in [patúu].

Unlike in Samoan, there are a fair number of 'broken' long vowels even in monomorphemes (as in 19b), so transcribed because the pitch rises on the first half of the long vowel, and falls on the second half. This violates *$V_iÝ_i$. Broken vowels occur especially often in English loans; among words with this structure, loans probably outnumber native words.

Tongan lacks words ending HL# (as in 19c), [14] with the possible exception of rising vowel sequences: authors disagree about where stress falls in words like ***tauhi*** 'look after', and the three speakers I consulted varied.[15]

(19) Tongan phonotactics

a. long vowels exist and contrast with short vowels

(káka)	‘to climb’	ka(káa)	‘parrot’
		(kàa)(káa)	‘to cheat’
(áfa)	‘to resemble; part of pandanus branch’		
		af(áa)	‘hurricane’
(pátu)	‘callus’	pa(túu)	‘to make slapping sound’
ta(kéle)	‘base’	(tàa)(kéle)	‘to have a bath (hon.)’
ka(táki)	‘to eat vegetables together’	(kàa)(táki)	‘to endure’

b. $V_iÝ_i$ is possible

ma(áma)	‘lamp’
ve(ési)	‘verse’ (loan)
me(ési)	‘mercy’ (loan)
ho(ósi)	‘horse’ (loan)

c. *HL#

*máama

Because breaking is allowed within monomorphemes, it’s not surprising that breaking, rather than shortening, usually occurs when a monomoraic suffix is added to a stem that ends in a long vowel, as illustrated in 20. There are a very few items that shorten instead (or perhaps the synchronic underlying form, unlike the proto-form, is short, and requires lengthening when unsuffixed to meet the bimoraic word minimum).

(20) Breaking under suffixation

a. usual pattern: breaking

(húu) ‘to go in’ hu(ú-fi) ‘to open officially’ (hùu)-(fía) ‘to sneak in’

b. very few words: shortening (Churchward 1953:11)

(óo)	‘to go’	(ó-mi)	‘to come’
(úu)	‘sheltered’	(ú-ŋa)	‘to take shelter’

Breaking always occurs when a word that ends in a long vowel bears definite accent, which can be described as adding another mora to a word (Poser 1985, Schütz 2001, Taumoefolau 2002, Anderson & Otsuka 2006).

(21) Breaking under definite accent

a. definite accent on a short vowel

(fále) 'house' fa(lé-e) 'house-*definite*'

b. definite accent on a long vowel: breaking

(póo) 'night' po(ó-o) 'night-*definite*'

De-breaking also occurs: a broken vowel reverts to a regular long vowel—that is, with stress on the first half—under definite-accent suffixation.

(22) De-breaking (Churchward 1953:12)

plain	*definite*	
ma(áma)	(màa)(má-a)	'lamp'
ta(áu)	(tàa)(ú-u)	'befitting'
ve(ési)	(vèe)(sí-i)	'verse'
me(ési)	(mèe)(sí-i)	'mercy'
ho(ósi)	(hòo)(sí-i)	'horse'

Again, the term DE-BREAKING is used atheoretically, to highlight that the broken V_iV_i sequence in the unsuffixed form corresponds to unbroken $\acute{V}_iV_i$ in the suffixed form.

Tongan has one-mora reduplication (as well as two-mora reduplication), and unlike in Samoan, it can apply to a vowel-initial foot, producing a $V_i\acute{V}_i$ sequence (though there are not many cases, since Tongan, having retained *ʔ and *h, has many fewer vowel-initial roots than Samoan).

(23) Reduplication can produce $V_iV́_i$ (from Churchward 1959); < > marks the reduplicant

plain		*reduplicated*	
áfu	'to spray finely'	<a>áfu	'to mist finely'
áke	'to swell'	<a>áke	'to have a swell'
áno	'lake'	<a>áno	'to be neap (tide)'
ápe	'to flow'	<a>ápe	'to be slimy'

It seems plausible that long vowels in Tongan arose in two diachronically separate instances. Greenhill and Clark (2011) reconstruct of modern [óo] 'to come' as Proto-Oceanic *oRo (as in 24a), which, speculatively, could have already had a suffixed form like *oRo-mi. At a later stage, *R deleted, with trochaic shortening where appropriate (as in 24b); this would be the source of the small number of synchronic shortening alternations. At a later stage, consonants such as the *r in *ma-ráma 'light' deleted (as in 24c). The *r is retained in nearly all other Polynesian languages (Tokelauan *malama*, East Futuna *malama*, Tuvaluan *malama*, Niuafo'ou *malama*, Maori *maarama*, Hawaiian *malama*; Greenhill & Clark 2011), and missing in Tongan and fellow Tongic language Niuean, suggesting deletion somewhere around Proto-Tongic. For some reason, the language at that point tolerated the $V_iV́_i$ sequence that resulted from *ma-rama > maama and similar consonant losses. This influx of $V_iV́_i$ words could have been the reason why suffixation began to cause breaking rather than shortening (except in the small number of lexicalized suffixed forms). Or, the change to breaking under suffixation could have occurred earlier, for an unknown reason, and paved the way for *maáma* and its ilk to be accepted without shortening.

(24) Two diachronic waves? Speculative sequence of events

a. Proto-Oceanic

*óRo 'to go, to come' *oRó-mi

b. post-Proto-Oceanic: *R deletes, with trochaic shortening

*óRo > *óo *oRó-mi > *ó-mi, rather than oó-mi

c. Proto-Tongic, or thereabouts: *r deletes, but no adjustment to length

*ma-ráma > maáma 'light'

Whereas trochaic shortening is neutralizing—for example, a Fijian learner, on hearing [síβi], must decide whether the underlying form is /siβi/ (yielding suffixed [siβí-ta]) or /siiβi/ ([siiβí-ta])—breaking preserves all underlying contrasts in both unsuffixed and suffixed forms. Tongan [maáma] must be from underlying /maama/, [màamá-a] from /maama-a/; [máma] 'ring' from /mama/, and [mamá-a] from /mama-a/. There is therefore no reason for breaking to lose productivity and become lexicalized, as seems to have happened for Samoan shortening. There do not seem to be Tongan doublets along the lines of [huú-fi] beside hypothetical [hú-fi], which would also be derived from [húu] but with a more idiosyncratic meaning.

Except for a very few cases of trochaic shortening, Tongan appears to be a good example of everything the typology predicts for a breaking language.

5.2 OTHER BREAKING SYSTEMS: NIUAFO'OU, TUVALUAN, NIUEAN, RENNELL-BELLONA

NIUAFO'OU. Niuafo'ou, also spoken in Tonga but not belonging to the Tongic family, has breaking in monomorphemes, and is claimed to stress the penultimate vowel even when it is a high vowel preceded by /a/ (compare Samoan, where stress generally falls on the /a/ in such a case: [máile] 'dog'). This is illustrated in 25. Also shown there is the possibility of reduplicating a vowel-initial syllable, as in Tongan.

(25) Niuafo'ou (Tsukamoto 1988:26, 39, 43, 47)

a. breaking in monomorphemes

pe(ési) 'page' (loan)

ŋa(áhi) 'make'

b. no special behavior for /ai/, /au/—stress is still on penultimate vowel

pa(íto) 'kitchen'

pa(úʔu) 'naughty'

c. breaking under suffixation

(táa) 'hit' ta(á-ʔi) 'hit+*Ci*'

d. reduplication of V is possible

(ínu) ‘drink’ <i>(ínu) ‘drink-*collective*’

Sequences of identical vowels, like *aa*, or low-high sequences, like *ai*, are still special, however—they behave differently from other VV sequences. Tsukamoto states that ‘secondary stress never occurs on the second vowel of a sequence of two identical vowels or the second vowel of a rising sequence of two non-identical vowels’ (as in 44). That is, there are no words like *[to(òke)(táa)], though there is [(tòo)ke(táa)] ‘doctor’—this example could merely reflect a preference for initial secondary stress though (an initial-dactyl effect, in Prince 1983’s terms, or ALIGN(PWord, L; Foot, L)). More convincing is the absence of words like *[(tàa)(ùpo)(óu)] (instead, [(tàa)upo(óu)] ‘virgin’).[16] The tableaux in 26 illustrate the ranking that produces breaking (a), and splitting of low-high sequences (b), but also avoidance of $V_iÝ_i$ or AÍ sequences when the only cost is unfooted syllables (violating PARSE-σ, Prince & Smolensky 2004).

(26) Niuafo’ou analysis

a. breaking

<INSERT Tableau (26)a HERE>

b. allow AÍ in antepenult-penult

<INSERT Tableau (26)b HERE>

c. avoid AÍ earlier in word

<INSERT Tableau (26)c HERE>

TUVALUAN. Tuvaluan also allows breaking in monomorphemes, and reduplication of single vowels.

(27) Tuvaluan

a. breaking in monomorphemes (Besnier 1995:xvii)

ni(ísi)	'some'	
pa(ála)	'wahoo'	(Proto-Polynesian *paqala, Greenhill & Clark 2011)

b. reduplication of single V (Besnier 2000:620); < > marks the reduplicant

plain	*reduplicated*		
(óla)	<o>(óla)	'alive'	
laoa	laa<o>oa	'choke'	(stress not given)
gaofe	gaa<o>ofe	'crooked'	(stress not given)

According to Proto-Polynesian reconstructions in Greenhill and Clark 2011, one of these reduplicated words had a consonant that would have provided an onset for the reduplicant (*laqoa), but the others did not (*ola, *gaofe).

Unlike in Niuafo'ou, certain vowel sequences attract stress onto the underlyingly antepenultimate vowel—as in Samoan, a misaligned or trimoraic foot is better than a sonority/prominence mismatch. Besnier analyses the underlying penultimate vowel as a surface glide in these cases (Besnier 2000:614).

(28) Tuvaluan nonhigh-high sequences

			possible footing(s)
/faiva/	fájva	'fishing method'	(fáj)va,(fájva)
/taua/	táwa	'fight'	(táwa)
/fe-taui/	fetáwi	'meet'	fe(táwi)
/avaifo/	avájfo	'send down'	a(váj)fo, a(vájfo)
/peleue/	peléwe	'coat'	pe(léwe)

Tuvaluan has a productive nominalizing suffix *–Vŋa*, illustrated in 29. The *V* in *–Vŋa* stands for a copy of the preceding vowel, underlyingly perhaps an empty mora. The stem-final

vowel is lengthened, presumably producing $V_iÝ_i$ (though stress is not transcribed). If the stem-final vowel is already long, it remains so (Besnier 2000:615-616).

(29) Tuvaluan nominalization

plain		*nominalized*
vvolu	'red-faced'	vvolu-uŋa
mate	'die'	mate-eŋa
sii	'fly-fish'	sii-ŋa
tuu	'stand'	tuu-ŋa
ffoo	'massage'	ffoo-ŋa

Alongside this regular pattern, many words have a 'relexicalized' (Besnier 2000:616) form with length and meaning differences from the basic form, as shown in 30. The stem-final vowel may fail to lengthen, or the stem's penultimate vowel rather than final vowel may lengthen (the preceding consonant, if long, shortens in that case). The relexicalized form often either has a somewhat opaque meaning, or can carry both the transparent meaning and an opaque meaning. (This is similar to the Samoan doublets in 14.)

(30) Tuvaluan 'relexicalized' suffixed forms (Besnier 2000:616-617)

plain		*regular suffixed*		*relexicalized suffixed*	
fai	'do'	fai-iŋa	'do, make'	fai-ŋa	'deed'
kai	'eat'	kai-iŋa	'eat'	kkai-ŋa	'festive meal'
moe	'sleep'	moe-eŋa	'sleep'	moe-ŋa	'bedding'
saina	'sign'	saina-aŋa	'sign'	saina-ŋa	'contract'
ssali	'flow'	ssali-iŋa	'flowing'	saali-ŋa	'flowing' *or* 'crack in reef through which water flows at ebbing tide'
pele	'favorite'	pele-eŋa	'beloved'	peele-ŋa	'beloved' *or* 'CARE package'
saka	'dance'	saka-aŋa	'dance'	saaka-ŋa	'dance' *or* 'choreography'

Tuvaluan, then, seems like another solid breaking language, freely tolerating violations of $*V_iÝ_i$, within a morpheme or across a morpheme boundary. The cases that might look like trochaic shortening or avoidance of trochaic lengthening (as in 30) seem to be lexicalized.

NIUEAN. In Niuean (Sperlich 1997), which is closely related to Tongan, monomorphemes can contain breaking, as illustrated in 30. Sperlich gives cognates and proto-forms to illustrate the origins of some of these broken vowels: 'in many cases there is good historical evidence which supports such rearticulated vowel sequences being the result of intervocalic consonant loss' (Sperlich 1997:6). Unlike Tongan, Niuean lost proto-Polynesian glottal stop, generating additional broken vowels. Sperling notes that 'rearticulation' does not occur in examples like proto-Polynesian **fara* 'pandanus' > Niuean *fā*; this is to be expected, since even if the representation is [(fáa)], with two separate vowel segments, stress is on the first one; 'rearticulation' tends to mean 'occurrence of pitch rise or other correlates of stress on second half of long vowel', and this does not apply to [fáa], which many authors therefore transcribe or analyze as a single long vowel.

Reduplication of a single vowel is also possible in Niuean, as in 31b. Greenhill and Clark's proto-Polynesian form for one word has a consonant (*qene), but the other does not (*afe) (Greenhill and Clark 2011).

(31) Niuean (Sperlich 1997:5-7)

a. monomorphemes with breaking

ha(áu)	(cf.Tongan *ha'au*)	'your'
ha(áku)	(cf. Tongan *ha'aku*)	'mine'
ha(ána)	(cf. Tongan *ha'ana*)	'their'
mo(óli)	(Proto-Polynesian *ma(a)qoli)	'true'
ma(áma)	(Proto-Polynesian *ma(a)rama)	'bright'
fo(óu)	(Proto-Polynesian *foqou)	'new

b. reduplication of V

<e>ene	'to poke'
<a>afe	'to branch off'

RENNELL-BELLONA. Rennell-Bellona (Elbert 1988), a Polynesian language in contact with non-Polynesian Oceanic languages, may be a breaking language. Elbert does not give an explicit stress rule, but he transcribes stress on a few words (Elbert 1988:14-16). The examples *hágе*

'house' and *manáha* 'settlement' suggest a system with stress on the penultimate mora, and the behavior of final long vowels is consistent with this: *táa* 'hit' and *totóo* 'to fall'. Elbert states that there are two types of long vowel, one transcribed as $\acute{V}_iV_i$, and the other, 'much less common' (Elbert 1988:14), described as 'rearticulated' and transcribed $V_i\acute{V}_i$. The $V_i\acute{V}_i$ cases usually involve the low vowel *aá*, according to Elbert.

Elbert states that 'none of the words with the rearticulated *aá* are of Polynesian origin', and gives a list of 12 words like *ghaághe* 'to hum', plus one English loan, *ghaási* 'glasses' (IPA [ɣaási], Elbert 1988:15). It is unclear whether broken vowels of other qualities are also always loans, or the generalization is meant to apply only to *aá* specifically. It is also unclear whether the only alternative to a broken vowel is a short penult vowel, or if unbroken long penult vowels are also possible, as in hypothetical *gháato*. Elbert does give one example that appears to be of this type, *ebéebe* 'to scatter' (Elbert 1988:14), but as this example comes right after an example whose stress mark was omitted, presumably by accident (*ibiibi* 'bony'), one wonders if the position of the stress mark in *ebéebe* is a typesetting error.

Examples of monomorphemic breaking, converted into phonetic transcription, are in 32, along with items illustrating that certain vowel sequences—/ai, ae, au, ao ei, eu, oi, ou, iu, ui/—can pull stress onto the antepenultimate vowel.

(32) Rennell-Bellona (Elbert 1988:15-16)

a. broken [aa] in loan monomorphemes

β**aá**to	'canoe connective'
m**aá**lu	'pandanus grove'
ɣ**aá**si	'glasses' (English loan)

b. [ao], [ai] require stress on first vowel

ᵑg**áo**i	'good'
t**ái**na	'younger sibling'

In terms of the constraints used here, this would mean that *AÍ (or a broader constraint to also cover the low+mid, mid+low, and high+high cases) outrank footing constraints, which in turn outrank *$V_i\acute{V}_i$.

Unlike in the other breaking languages discussed here, under suffixation Rennell-Bellona seems to display length alternations rather than mere shifting of stress. The examples in 33a and 33b are consistent with an underlying length difference, neutralized in unsuffixed forms through trochaic shortening. In 33c we see apparent shortening of final long vowels under suffixation, although the half-dozen shortening examples are all of the form CVV, and thus could be underlying /CV/ with lengthening in the unsuffixed form to meet a size minimum. There do not seem to be examples of stem-final long vowels that fail to shorten under monomoraic suffixation, but there is one example given of a de-shortening that affects the 'wrong' vowel (as in 22d).

(33) Rennell-Bellona suffixation (Elbert 1988:226, 228, 233, 238)

a. examples of non-alternation

ᵑgaŋa	'group'	ᵑgaŋa-hi	'to do together'
huᵑgu	'hair'	huᵑgu-ki	'hairy'

b. examples of de-shortening under suffixation

ᵑgohi	ᵑgoohi-a	'cold'
ᵑguku	ᵑguuku-a	'to bring food from the bush'
soᵑgo	he-sooᵑgo-i	'to play'

c. examples of shortening under suffixation

taa	he-ta-ʔi	(no gloss)
ᵑguu	he-ᵑgu-ti	(no gloss)

d. one example in which stem-final V lengthens

taᵑga	taᵑgaa-ki	'to alter'

Although we should be cautious about assuming productivity, it appears that Rennell-Bellona allows $V_iV́_i$ in some monomorphemes, but adopts shortening rather than breaking as an alternation. This is the inverse of the Samoan pattern (section 4.1), where breaking occurred under suffixation but not in monomorphemes. Like Samoan, Rennell-Bellona falls under Paster's

'loss of a static generalization' category of diachronic change (Paster 2013), where the OT analysis can be rescued by adding an extra constraint.

The proposed solution for Samoan was to add DON'TSHORTEN-V́:-BA (breaking is forbidden in general, but shortening is, variably, even worse under suffixation). For Rennell-Bellona, if we treat the loans with [aá] as fully legal, then we need the ranking MAX-V >> *$V_iV́_i$. But to ensure shortening under suffixation, we need a higher-ranked constraint to rule out broken *[he-taá-ʔi] but not broken [maálu], such as NOBREAKING-BA.

(34) Constraint rankings for Rennell-Bellona

- NOBREAKING-BA: a long vowel or $V́_iV_i$ in the unsuffixed base must not correspond to $V_iV́_i$ in the affixed form

a. simple word

<INSERT Tableau (34)a HERE>

b. affixed word

<INSERT Tableau (34)b HERE>

A speculative diachronic explanation for the current state is that Rennell-Bellona at one time had length alternations (whether productive or not), and banned $V_iV́_i$ absolutely—explaining the dearth of native words with broken vowels. New long penults introduced by loans underwent breaking, introducing a novel phonotactic pattern. What is curious about this scenario is that at least for English loans, it is surprising that if a new, previously illegal, pattern was to be introduced, it was breaking rather than HL tolerance. That is, if [ɣáasi] and [ɣaási] were both illegal when English loans came in, why was British English [glá:s]/[glás] (or American [glæs]) adapted as [ɣaási] rather than the phonetically closer [ɣáasi]?

5.3 SUMMARY OF BREAKING SYSTEMS. If the diachronic sketch in section 3 is correct, then Central Pacific languages inherited a system with no $V_iV́_i$ sequences. The languages in this section innovated them, through consonant loss and through borrowed words with long-vowel penults. Assuming that loan adaptation is driven by native-language-shaped perception (e.g. Peperkamp 2005), this means that loan adapters perceived foreign $V́_iV_i$ or V́: (e.g. English [glɑ:s(əz)] 'glass(es)') as closer to $V_iV́_i$ (Rennell-Bellona [ɣaási]) than to short V́ (hypothetical [ɣási]), even if the short vowel conformed better to the phonotactics of the language.

The tight coupling that we might expect between phonotactics and alternation in a classic conspiracy thus breaks down, as Paster (2013) argues that we should expect. Diachronic shortening in both monomorphemes and affixed words initially produced a system in which phonotactics and alternations matched, but learners and speakers then allowed them to diverge, resulting in systems with $V_iV́_i$ sequences, but also some length alternations that appear to be driven by $*V_iV́_i$. If the alternations are unproductive, the system is easy to analyze in Optimality Theory (rank $*V_iV́_i$ low and list the forms that alternate), though it is still surprising that Rennell-Bellona speakers did not infer from the absence of $V_iV́_i$ words that $*V_iV́_i$ should be ranked high. If the alternations are productive, then the grammar requires constraints specific to morpheme boundaries ($*V_i$-V́$_i$ and $*V_iV́_i$-), or even specific to particular morphemes ($*V_iV́_i$-a).

6. (PARTIAL) TOLERANCE FOR HL#. The typology in sections 1 and 2 included languages that simply tolerate a long vowel in the penult, and should have no breaking or trochaic shortening, as schematized in Table 5 and 35. Depending on the ranking of the bottom two constraints, there is either a right-aligned, trimoraic foot or a non-aligned, bimoraic foot. The two options would sound the same.

Table 5 <INSERT TABLE 5 HERE>

(35) Tolerance tableau
<INSERT Tableau (35) HERE>

Several languages were found that approximate this state, though none match it perfectly.

6.1 TOKELAUAN: TOLERANCE WITH ALTERNATIONS. Tokelauan is closely related to Samoan. According to Hooper (1996:2; glosses from dictionary cited below), the basic word stress pattern is penultimate (when all the vowels in a word are short: [manátu] 'opinion'), but a long vowel attracts stress. In Hooper's examples, the long vowel is final ([pakúu] 'fall') or antepenultimate ([máahina] 'moon'). In [máahina], the long vowel is clearly different from an ordinary sequence of vowels, which would presumably not attract stress so far to the left.

What about a long vowel in the penult? Simona, Huntsman, and Hooper's dictionary (Simona, Huntsman & Hooper 1986) lists 83 words that end HL (not including prefixed forms of words already counted). Of these, 19 appear to be monomorphemic native words (e.g. *tāne* 'man, husband'), and 12 are loans from English (*māpu* 'marble').[17] The dictionary's preface does not discuss the pronunciation of long vowels in this position, or stress in words of this shape, but we can tentatively assume that stress falls on the first half of the long vowel, since that is what a vowel with a macron usually indicates in transcriptions of Polynesian languages. This means that Tokelauan tolerates either a non-aligned final foot, [(táa)ne]—and, for that matter, [(máa)hina]—or a trimoraic final foot, [(táane)].

Hooper (1996:34) gives several examples of what look like de-shortening alternations, shown in 36. As in other cases seen above, the diachronic origin of the long vowel in the suffixed form is unknown (proto-Polynesian forms from Greenhill & Clark 2011 given when available).

(36) a. de-shortening

k**a**ve	'to carry'	k**aa**ve-ŋa	'load'	*qawe
		cf. k**a**ve-ŋa	'action of carrying' (dictionary)	
t**e**u	'to decorate'	t**ee**u-ŋa	'decorations'	*teu
t**i**pi	'to cut'	t**ii**pi-ŋa	'surgical operation'	*tipi
		cf. t**i**pi-ŋa	'action of cutting' (dictionary)	
faka-heet**o**nu	'to be doubtful'	faka-heet**oo**nu-ŋa	'thoughtlessness'	

b. non-alternating short

m**o**e	'to sleep'	m**o**e-ŋa	'bed'	*mohe
taaf**a**o	'to play'	taaf**a**o-ŋa	'toy, game'	
inu	'to drink'	**i**nu-ŋa	'drinking party'	*inu

c. non-alternating long (from dictionary)

p**ee**hi ‘(epidemic) to rage’ p**ee**hi-a ‘to be affected (by epidemic)’ *peqe-si

Are the de-shortening alternations productive? If they are, the language’s tolerance of HL# is not across-the-board. One factor undermining an account of the alternations as productive is the existence of both non-alternating short and non-alternating long stem penults (as in 36b and 36c). A plausible analysis for learners to adopt would be that always-short vowels are underlyingly short, and the few always-long vowels underlyingly long, which leaves the alternating vowels to be marked as exceptions or memorized as separate words. The two doublets in 36a, where the de-shortened option carries an idiosyncratic meaning, support this analysis.

As in Samoan, Tokelauan has sporadic morphological vowel lengthening outside the penult, as in *hāvali* ‘messenger’, claimed to be derived from *havali* ‘walk’, and *mā<li>lie* ‘agree-*plural*’ from *malie* ‘agree’ (Hooper 1996:33-34). So whatever the diachronic origin of the long vowels in 36a, it is possible that speakers now interpret them as lexically idiosyncratic lengthenings rather than as reflexes of an underlying long vowel.

Unfortunately, the sources do not give examples of adding a suffix to a stem with a final long vowel (e.g. hypothetical [pakúu-ŋa], [pakuú-ŋa] or [pakú-ŋa]).

To summarize, Tokelauan is a language that tolerates HL#. It has some de-shortening alternations, but these likely are lexicalized.

6.2 TUAMOTUAN: VARIATION BETWEEN BREAKING AND TOLERANCE. Tuamotuan (Kuki 1970) has variation between breaking a long-vowel penult (i.e. stressing its second half) and stressing it all the way through.

(37) Variation in stress for long-vowel penult (Kuki 1970:69-71)

ʔoóna ~ ʔóóna ‘he’

taáku ~ tááku ‘my (alienable)’

Kuki states, ‘primary stress occurs on the second member of each long vowel. In this case, primary stress usually begins on the first member of each long vowel [...]. Thus, the most

frequent stress patterns are /=’óóna=/ instead of /=’oóna=/, and /=tááku=/ instead of /=taáku=/, respectively, even in deliberate speech.’ (as in 71). (Skimming Stimson & Marshall 1964, a dictionary, shows that unlike in many Central Pacific languages, long-vowel penults are plentiful in Tuamotuan.)

This variation suggests that in Tuamotuan, two constraints are active and in competition. The more-frequent variant, [ʔóóna], satisfies *V_iV_i but presumably has either a misaligned foot, *(ʔóó)na*, or a trimoraic foot, *(ʔóóna)*. The less-frequent variant is well footed, but at the expense of violating *V_iV_i: *ʔo(óna)*.

Like long vowels, sequences such as *ei, ai, ae* have two pronunciations in the penult: the second (higher) vowel can be stressed, or it can be demoted to a glide, which Kuki transcribes as a superscript vowel, with stress on the preceding (lower) vowel.

(38) Variation in rising vowel sequences (Kuki 1970:74-75)

keíga	~	ké^{i}ga	‘bone’
paími	~	pá^{i}mi	‘if’
haére	~	há^{e}re	‘to go’

Tuamotuan has optional devoicing of final vowels. Unusually, this devoicing behaves as ‘earlier’ than stress, as illustrated in 39a: if final devoicing applies, the final vowel is no longer counted for stress, and stress falls on the underlying antepenult instead. Kuki also transcribes some final vowels, including low vowels, as demoted to glides when preceded by another vowel; in these cases, the final vowel again does not count for stress, and stress shifts (as in 39b). This means that glides (and devoiced vowels) do not contribute a mora, and the forms in 38 are footed as [(ké^{i}ga)] rather than [(ké^{i})ga].

(39) Devoicing or demotion to glide of final vowel (Kuki 1970:72-75)

a. final vowel devoiced

tagáta	~	tágat̥a	‘man’
pakóti	~	pákot̥i	‘scissors’

b. final vowel demoted to glide

ràakáu	~	raáka^{u}	~	rááka^{u}	tree’
pàuróa	~	paúro^{a}	~	pá^{u}ro^{a}	‘all’

In sum, Tuamotuan is a language that varies between tolerating HL# and tolerating $V_iV́_i$. It also has the unusual feature that stress interacts transparently with final devoicing or gliding, such that stress counts surface moras, not underlying vowels.

6.3 OTHER (PARTIAL-)TOLERANCE SYSTEMS: NUKUORO, TAHITIAN, KAPINGAMARI, HAWAIIAN, TAKUU, ONTONG JAVA, NORTH MARQUESAN, EAST FUTUNA

NUKUORO. In Nukuoro (Carroll 1965, Carroll & Soulik 1973), long vowels in the penult are possible, both in monomorphemes and under suffixation (as in 40). Proto-Polynesian forms from Greenhill and Clark 2011 are given in 40 where available; at least two of the long penults derive from consonant loss (*hoou, ttaane*), and one is inherited, but from a word that was formerly HH (*maalo*).

(40) Nukuoro long penults

a. monomorphemes (Carroll 1965:5, 9)

maalo	‘wide awake’	< *maaloo	vs.	malo	‘cloth’	< *malo
nuui	‘green’	< *qui	vs.	nui	‘coconut tree’	< *niu
hoou	‘new’	< *foqou	vs.	hou	‘drill’	< *fohu
siili	‘type of stick’		vs.	sili	‘to stop’	
ttaane	‘man’	< *taqane				

b. suffixation (no shortening) (Carroll & Soulik 1973)
pakuu ‘fall over’ pakuu-ŋa haka-pakuu-a

I read Carroll 1965’s description of stress as ambiguous, but suggesting that words like *maalo* are not broken—that is, not pronounced as [maálo]. He states that ‘[s]yllables take the shapes V, VV, VVV, CV, CVV and CVVV. All possible V and VV combinations occur. [...] The first member of a diphthong is always the syllabic peak when the syllable is stressed; elsewhere there is little difference between members, the peak of sonority tending to occur on the most naturally sonorous vowel’ (Carroll 1965:8), and that ‘[p]rimary stress (′) occurs at least once in each contour word [content word and associated function words], predictably on the penultimate syllable of each base, pronoun or other morpheme occupying the nuclear position. [...] Primary stress is phonetically defined by rising pitch.’ (Carroll 1965:8) These statements would be consistent with a syllabification [maa.lo], with stress falling somewhere on [maa]. They would also be consistent with [ma.á.lo], but I take Carroll’s statement that double vowels are realized ‘about twice as long as single vowels; not rearticulated’ (Carroll 1965:7) to mean that [maálo] is not what is intended, and rather it is [máalo] or [máálo].

There are many examples of de-shortening-like alternations, illustrated in 41, as well as non-alternation under suffixation and some more-arbitrary length alternations (not shown). As in other languages we have seen above, the length alternations do not derive from a long vowel in the proto-form, but must have been innovated at some point. For example, *holo* ‘swallow’ is from *folo, with a short vowel, but still has a long vowel under suffixation.

(41) Nukuoro length alternations (Carroll:30-31, with gaps filled in from Carroll & Soulik dictionary)

unsuffixed	*bimoraic suffix*	*monomoraic suffix*	
kk**a**di		k**aa**di-a	‘bite’
k**a**lo	kalo-hia	k**aa**lo-a, k**aa**lo-ŋa	‘stir’
kk**u**mi	kumi-dia	k**uu**mi-a	‘squeeze’
s**e**ŋa		s**ee**ŋa-ŋa	‘crazy’
h**o**lo	holo-maŋa	h**oo**lo-ŋa	‘swallow’
unu	unu-maŋa	**uu**nu-ŋa	‘drink’

TAHITIAN. Tahitian, which also has penultimate stress, tolerates HL# even more than Nukuoro, with no variation reported. Stress simply falls on a long penult vowel, including in loans, as illustrated in 42a. Sequences like *ae, ao, au* also draw stress onto the /a/ (as in 42b). There are sporadic examples of attaching a suffix to a stem-final long vowel, and these show shortening (as in 42c); I was not able to determine whether shortening under suffixation is obligatory.

(42) Tahitian heavy penults

a. stressed, long penult — short penult for contrast (Bickmore 1995:412)

váahi[18]	'place'	túpu	'happen'
máaha	'satisfied'	mána	'power'
péeni	'paint' (loan)	póto	'short'
faráani	'French' (loan)	ferúri	'reflect, think'

b. rising vowel sequences — other VV sequences for contrast (Bickmore 1995:413)

ʔáeto	'eagle'	teáta	'theater' (loan)
faráoa	'flour' (loan)	moána	'ocean'
táura	'rope'	huáre	'saliva'

c. shortening (Lazard & Peltzer 2000:234)

tuu	'être bord à bord'	tu-ʔi	'effleurer, toucher'

KAPINGAMARANGI. Kapingamarangi, again with penultimate stress, also allows stressed, long penults (as in 43a), including etymologically unexpected ones like *ráaŋi* 'sky' (Proto-Austronesian *laŋiC, Proto-Polynesian *lagi). There are three possibilities for stem penults under suffixation: always short, always long, and alternating (short when unsuffixed, long when suffixed), shown in 43b. (The items in 43b are all the disyllabic stems with a verbal *–a* suffix that could be found in the *h* sections of the two sources, taken as a hopefully representative sample. At least in this sample, we see that always-short is by far the biggest category. Elbert's *haki* 'pluck' and Lieber and Dikepa's *haki* 'to pick' are presumably the same root.)

(43) Kapingamarangi (Elbert 1948)

a. stressed, long penult

máanu	‘cramped; float’
ráaŋi	‘sky’
ráawa	‘choke’
táahi	‘hold, paddle slowly’

b. length under suffixation (rows with just one gloss: lexicon portion of Elbert; rows with two glosses: Lieber & Dikepa 1974)

always short:	haki		haki-a	‘tell, say’
	haki		haki-a	‘pluck’
	hati		hati-a	‘break in two’
	hihi		hihi-a	‘write’
	hina	‘grey hair’	hina-a	‘white haired’
	huri		huri-a	‘turn’
	hui		hui-a	‘dip in’
	hunu		hunu-a	‘paint’
	hoŋi		hoŋi-a	‘smell’
always long:	haahi		haahi-a	‘slice’
alternating:	haki	‘to pick’	haaki-a	‘twist it off!’

HAWAIIAN. Hawaiian, which likewise has penultimate stress, tolerates stressed, long penults (as in 44a). Diachronically, Elbert and Pukui (1979) state that long vowels were shortened before *k/current *ʔ*, as illustrated by the three pronouns in 44b, where shortening occurred before *k/*ʔ* only. The shortening examples given are all in the penult, and Elbert and Pukui state that the sequence CV:ʔV IS allowed in longer words (as in 44c, 44d), where the long vowel does not receive primary stress. Thus, although Hawaiian generally tolerates ...HL#, and generally tolerates V:ʔ, it does not tolerate both in the same place. This could be analyzed as constraint conjunction, (Smolensky 1995, Hewitt & Crowhurst 1996), or cumulativity in harmonic grammar (Legendre, Sorace & Smolensky 2006).

(44) Hawaiian

a. stressed, long penult possible (Elbert & Pukui 1979:15)

máala ‘garden’ < Proto-Polynesian *maqala

b. diachronic trochaic shortening before *k/ʔ (Elbert & Pukui 1979:14)

	*taaku	>	káʔu	‘my’	*naaku	>	náʔu	‘my-*focus*’
vs.	*taau	>	káau	‘your’	*naau	>	náau	‘your-*focus*’
	*taana	>	káana	‘his’	*naana	>	náana	‘his-*focus*’

c. V:ʔV not shortened in other positions (Pukui & Elbert 1986)

*maakona	>	màaʔóna	‘satisfied’
*feekii	>	hèeʔíi	‘papaya’
		màaʔóhu	‘misty’
*pi(i)kao	>	pìiʔáo	‘fold into cup’

TAKUU. Takuu (Moyle 2011), again a language with penultimate stress, allows stressed, long penults (as in 45a). There are some examples of alternations that look like de-shortening, some examples of shortening, and idiosyncratic length changes.

(45) Takuu (Moyle 2011, from dictionary section if no page number given)

a. stressed, long penult (Moyle 2011:7)

aróoha ‘compassion’

b. some de-shortening

	llomi	‘press down’	loomi-a	‘pressed down’ (Moyle 2011:9)
	llana	‘weave’	laana-a	*imperative*
cf.	hati	‘break into pieces’	hati-a	*imperative*[19]

c. shortening and optional shortening

hookii	‘hand over’	hooki-na	‘hand over- pass.’
too	‘take’	too-a ~ to-a	‘take-pass.’

d. idiosyncratic length changes

haanai ‘feed’ hanai-a *imperative*

As a side note, although Takuu has penultimate stress (supporting a bimoraic foot), and most of its basic vocabulary is disyllabic, it has the unusual feature of a minimum word size of three moras for nouns and four for verbs. This minimum is enforced by procliticization: one- and two-syllable nouns require a proclitic when given in citation form, as do one- through three-syllable verbs (Moyle 2011:7-8).[20] See Blust (2007) for further unusual aspects of Takuu prosody.

ÙA POU. The Ùa Pou dialect of North Marquesan (Mutu & Teìkitutoua 2002) has an essentially penultimate stress system (Mutu & Teìkitutoua 2002:34-35), and does allow stress to fall on a long penult in an HL# word: *ʔéemi* ‘draw up’, *áaka* ‘root’ (Mutu & Teìkitutoua 2002:22-23).

Ùa Pou is an especially interesting case because there is one sense in which it actually PREFERS HL#: unusually for Polynesian languages, a stressed penult of a phrase is greatly lengthened. Mutu and Teikitutoua call this “penulatimate vowel extension” (Mutu & Teìkitutoua 2002:31-33), and give duration measurements for several examples. In 46, we see that the stressed penult of phrase-final *óke* is lengthened, but that of phrase-medial *máte* is not.

(46) Ùa Pou penultimate vowel extension (Mutu & Teìkitutoua 2002:31-33)

ʔua m**á**te ʔoutóu i te **ó**ke → [**ó**::ke]

PERFECTIVE die 2PLURAL PREPOSITION DEFINITE hunger

‘You are hungry.’

Mutu and Teikitutoua don’t report whether this phrasal lengthening neutralizes the underlying length distinction between words like /oke/ and words like /aaka/.

There is one other unusual aspect of Ùa Pou stress worth mentioning. Long vowels attract primary stress, even when non-final (similar to Tokealuan in 6.1 and Maori in 7.1). As shown in 47, when there are multiple long vowels in a word, whether it is the first or the last that gets stressed depends on the length of the word. (Certain vowel-vowel sequences such as [ai] also attract stress if there is no long vowel, as in [háika] ‘medicine’ or [kéitani] ‘jealous’.)

(47) Ùa Pou stress system (Mutu & Teìkitutoua 2002)

a. default when all light: stress penultimate mora (suggests right-aligned moraic trochee)
   máta ‘eye’
   vehíne ‘woman’
   puáʔa ‘pig’
   kaukáu ‘bathe’

b. two-syllable words: stress LAST long vowel, if any
   het**úu** ‘star’
   paot**úu** ‘all’
   t**áa**tou ‘we’
   k**oo**p**úu** ‘stomach’

c. three-syllable words: stress FIRST long vowel, if any
   m**áa**maʔi ‘egg’
   kaikai**áa** ‘demon’
   p**áa**ʔaiʔai ‘(fish species)’
   p**áa**k**oo**k**oo** ‘knock (at door)’

All of the examples in 47 are consistent with a right-aligned moraic-trochee footing, but with complex rules for which foot gets primary stress.

ONTONG JAVA. Ontong Java (Lanyon-Orgill 1944:9-13) is another penultimate-stress language that tolerates a stressed, long penult, including etymologically unexpected *liima* ‘arm’ (Proto-Austronesian *lima).

EAST FUTUNA. East Futuna has penultimate stress unless the final vowel is long (Grézel 1878:9; Rensch 1986:ix; Moyse-Faurie 1993:22). East Futuna tolerates a stressed, long penult, including in native words (*fugāo* ‘son-in-law, daughter-in-law’), loans (*pūsi* ‘cat’), and suffixed words (*tō* ‘punch’, *tō’i* ‘to punch’—all examples from Moyse-Faurie).

Moyse-Faurie (1993:21) states that there only two non-prefixed, non-compound, non-reduplicated cases of $V_iV́_i$, where two identical vowels in a row are ‘pronounced successively

(and without intervening glottal)':[21] *tuu'i* 'to be shaken; refuse' and *muu'i* 'to light'. I take Moyse-Faurie's description to indicate a pronunciation like [tuúʔi]. The verb *tuu'i* is presumably a suffixed form of *tū* 'to pour', indicating breaking under suffixation; there is no verb stem *mū* in the dictionary, but *muu'i* could well be a verb, bearing the *–Ci* ending that is often a verb suffix. (The claim is undermined by the entry for *pāki* 'to draw', which lists a variant pronunciation *paaki*.)

6.4 TOLERANCE OF HL#: SUMMARY. Quite a number of Central Pacific languages have reacted to long vowels by allowing HL#, rather than allowing V_iV_i—or they vary, as in Tuamotuan and perhaps East Futuna. They often retain length alternations consistent with trochaic shortening, but these may be lexicalized.

7. INNOVATIVE STRESS PATTERNS. Although the vast majority of Central Pacific languages have retained penultimate stress, there are a few whose stress pattern has changed, potentially making an underlying form that ends HL# unproblematic.

These languages form a heterogeneous group. Mele-Fila could be classified as a tolerance language like those above, except that final moras are extrametrical. Emae seems to be in flux between left and right alignment of stress, so that it is unclear whether ...(H)L# violates the basic alignment pattern. Maori seems to have changed stress orientation (left vs. right), and retains a priority for stressing long vowels above alignment; some de-shortening persists, though possibly with an entirely new motivation.

7.1 MAORI. There are many subtly differing descriptions of basic Maori stress (Schütz 1985, Bauer 1993, de Lacy 2002b, Harlow 2007), but they mostly agree that stress is aligned more to the beginning of the word than to the end. For example, a disyllabic, trisyllabic, or quadrisyllabic word made of all light syllables has initial stress.

(48) Maori basic stress

mána	'power' (Bauer 1993:556)
mánawa	'heart' (Bauer 1993:556)
ráŋatira	'chief' (Harlow 2007:82)

Descriptions also differ as to how long vowels and VV sequences are treated. Bauer 1993's synthesis states that the first long vowel in the word is stressed, if there is one; otherwise the beginning of the first VV sequence is stressed—but with some variation for final VV sequences, as in [fenúa ~ fénua], which could reflect a dialect difference.

(49) Maori non-light syllable stress

stress first long vowel

háaŋii	'earth oven'	(Bauer 1993:557)
kóofai	'type of tree'	(Bauer 1993:557)
kuríi	'dog'	(Bauer 1993:557)

if no long vowels, stress beginning of first VV sequence

káuwae	'jaw'	(Bauer 1993:557)
háere	'move'	(Bauer 1993:557)
fenúa ~ fénua	'land'	(Bauer 1993:557)
tamáiti	'child'	(de Lacy 2002:4)

We shouldn't particularly expect trochaic shortening in Maori, at least not for the usual reasons, if there is no requirement for feet to be right-aligned—that is, if ALIGN(PWord, R; Foot, R) is ranked too low to play a role. If /maali/ surfaces as [(máa)li], the only real cost is non-footing of one syllable, just as in [(mána)wa]. There are indeed words with long penults.

(50) Maori HL# words[22]

futupooro	'football' (loan)	(Harlow 2007:69)			
wuuru	'wool' (loan)	(Harlow 2007:69)			
feeke	'creak'	vs.	feke	'octopus'	(Bauer 1993:534)

What should be problematic in Maori, if the main-stress foot is left-aligned, is a /LH.../ word, like /manaaki/ 'support' or /mataa/ 'flint, bullet'. The pronunciation is [ma(náa)ki] (Lynch 1998:81) and [ma(táa)] (Harlow 2007:82), showing that maintaining faithful length and stressing long vowels[23] is more important than aligning the main-stress foot to the left.

Harlow (2007:117) lists about twelve verbs that undergo what looks like de-shortening.

(51) Maori length alternations

ako	aako-na	‘learn, teach’
huti	huuti-a	‘hoist, haul up’
kume	kuume-a	‘pull, drag

There are also many verbs that don’t alternate. De Lacy (1996) proposes that in those words that alternate, the suffixed form is lengthened in order to allow a footing like [(aa)(ko-na)], with both stem and suffix associated to bimoraic prosodic words, rather than [(ako)-na].

7.2 MELE-FILA. Mele-Fila is spoken in Vanuatu, and has contact with distantly related South Efate (Oceanic, but not Central Pacific), which has mostly initial stress (Thieberger 2006). Perhaps for this reason, Mele-Fila has developed antepenultimate stress (Capell 1942), illustrated in 52. Antepenultimate stress is enforced even on disyllabic words: citation forms add a proclitic if necessary to bring a word up to the three-syllable minimum.

(52) Mele-Fila antepenultimate stress

a. antepenultimate primary stress

táŋata	‘man’	(Capell 1942:155)
máua	‘find (transitive)’	(Clark 1998:x)[24]
samásama	‘happy’	(Biggs 1975:8)

b. trisyllabic minimum in citation form, enforced through proclisis (Biggs 1975:8 notes this too)

té-fare	‘the house’	(Clark 1998:x)
ée-fano	‘goes, went’	(Clark 1998:x)
ée-tawa	‘it flares up’	(Clark 1998:x)
ée-rua	‘two’	(Capell 1942:155)

The exceptions to the trisyllabic minimum are disyllabic loans with final consonants, and some words with word-internal codas (as in 53). Clark points out that these words could be analyzed as having underlying vowels that count for stress but then are deleted (e.g. /nooti/), but it would also be possible to analyze coda consonants as moraic.[25] If Mele-Fila's real requirement is not trisyllabicity but rather a bimoraic foot, with the final mora unfooted, then these words are easily understood as long enough to meet both requirements even if they are not strictly trisyllabic (speculative footings are shown in 53).

(53) Mele-Fila moraic codas (Clark 1998:x)

		speculative footing—final mora not footed	
noot	'north' (loan)	(nóo)t	
karso	'watercress'	(kár)so	
vunta	'numb'	(vún)ta	
	cf.	(táŋa)ta	'man'

There are also suffixes/enclitics that systematically trigger penultimate stress, perhaps through a morpheme-specific requirement that they be footed:

(54) Mele-Fila suffixes/enclitics that shift stress to penultimate

taŋatá-ra	'that man'	(Capell 1942:155)
maará-na	'garden-*possessive*'	(Capell 1942:155)

If Mele-Fila essentially has right-aligned moraic trochees, with the only difference from penultimate-stress languages being the ban on footing a word-final mora, then we might expect breaking or trochaic shortening to apply when an underlying ANTEpenult is long. For example, hypothetical /maalita/, if not shortened or broken, would surface as [(máali)ta], with a trimoraic foot, or [(máa)lita], with a foot less right-aligned than normal. Shortening would avoid these problems: [(máli)ta].

Long vowels are abundant in the antepenult, though. Examples are shown in 55. Looking through the dictionaries it appears that long antepenults may even be the majority. Clark (1998) uses a double-vowel spelling, so it is impossible to know whether a $V_iV́_i$ pronunciation is

intended (e.g. [a(áma)ta]). But Biggs (1975) uses macrons, as in *kāmoa*, suggesting that the pronunciation is $\acute{V}_iV_i$ ([(káa)moa] or [(káamo)a].

(55) Mele-Fila long antepenults

aamata	'new leaves'	(Clark 1998)
aaoa	'banyan'	(Clark 1998)
kaamoa	'take'	(Biggs 1975)

Long penults exist but are rare—perhaps that scarcity is a legacy of trochaic shortening before stress shifted to antepenultimate, even though the phonotactic near-gap is no longer motivated by the language's stress system.

(56) Mele-Fila long penults: few examples

auskeele	'house-girl'	(loan, Clark 1998)
kooli	'lame'	(Biggs 1975)
maara	'garden'	(Clark 1998)

In summary, Mele-Fila appears to fall abstractly in the HL#-tolerance category, except that final moras are ignored, so what it tolerates is really HLL#.

7.3 EMAE. Emae (Capell 1962), like Mele-Fila, is in contact with South Efate. Little information is available, but there are some words with antepenultimate/initial stress (*nánafi* 'yesterday', *táŋata* 'man', Capell 1962**:**6) amidst a general pattern of penultimate stress. The language may be in flux between a right-oriented and a left-oriented stress system: the enclitic *ni* draws stress to the right (unlike most suffixes), suggesting right-aligned stress, but the proclitic *a* draws stress to the left (*á kai* 'food'), suggesting left-aligned stress.

Long penults appear to be tolerated (*póoki* 'to beg', from dictionary section), perhaps taking advantage of the option for a left-aligned foot.

8. LOSS OR PARTIAL LOSS OF CONTRASTIVE LENGTH: ROTUMAN, RAPA NUI, FUTUNA-ANIWA, VAEAKAU-TAUMAKO. In most Central Pacific languages, the contrast between short and long vowels is weak: short vowels greatly outnumber long, and minimal pairs are few. But some of these languages have lost the contrast altogether, so that length is predictable. Such languages still mostly lack ...HL#, either because this is not an environment where vowel lengthening needs to apply (Rotuman, possibly Rapa Nui) or because of a lexical gap (Vaeakau-Taumako).

ROTUMAN. Rotuman appears at first glance to have contrastive length, but Blevins (1994) argues that long vowels in Rotuman are always the result of lengthening to satisfy a minimal word requirement (/CV/ → [CVV]), or to create binary feet in words with lexically marked final stresses (/ravá/ → [ra(váa)] 'to be defeated') or lexically marked secondary stresses in some positions (/kàré/ → (kàa)(rée) 'curry').

In other positions, long vowels do not occur. For example, there are no morphemes of the form *[paalemia], because even if the word is marked with initial secondary stress (/pàlemia/), lengthening is not needed to provide binary feet: [(pàle)(mía)] is the optimal outcome whether the first syllable is marked underlyingly as stressed or not.

Blevins notes that diachronically expected final long vowels have shortened, as in *faʔu* 'star' < *vetuʔu, suggesting that all or most long vowels were shortened at some point (except in monosyllabic words), with new long vowels introduced through loans and other means.

Under this analysis, apparent instances of shortening under suffixation are misleading. As illustrated in 57, a word marked with stem-final stress will have a long vowel when unsuffixed, but there is no need for lengthening when a suffix is present, because stress can still fall on the stem-final syllable without sacrificing foot binarity.

(57) Rotuman: shortening as an illusion

/maró/		/maró-si/
ma(róo)	'to be taut'	ma(ró-si) 'inflexible' (Churchward 1940:261)

Very little separates the Rotuman surface system from a system like Samoan's, despite the radically different analyses. The main difference is that Samoan does have words like *pàalemía* 'Premier (loan)' (Mosel & So'o 1997), with a long vowel that did not need to be long in order to bear stress—that is, *pàlemía* would have been well-formed too. If a learner were to

overlook this small set of words, he or she could plausibly develop a contrastive-stress analysis rather than a contrastive-length analysis.

RAPA NUI. In Rapa Nui, length seems to be essentially non-contrastive (Du Feu 2012). According to Du Feu, the only remnant of a contrast is in final syllables; there are two possible analyses of these. Under one analysis, final vowels can be underlyingly long, and if so attract stress (which is otherwise penultimate), with optional lengthening of the preceding vowel too, as in /mataa/ → [ma(a)táa] 'axe head' (Du Feu 2012:186). Under the other analysis, there is no underlying length, but some words are marked with final stress, causing lengthening: /matá/ → [ma(a)táa].

There would therefore be no question of trochaic shortening or breaking in monomorphemes; the only case where underlying HL# could arise would be, under the analysis where final length is contrastive, if a stem with a final long vowel, like /mataa/, were suffixed. The sources I consulted did not address such forms.

Some examples in older sources appear to indicate length in other positions. For example, Fuentes 1960[26] lists words like *auráa* 'because', *tuuría* 'shell', and examples that look like breaking, such as *huaái* 'family', *aámo ~ ámo* 'to lick one's lips', *aámu* 'tattler', *aái* 'who', and *abaái* 'to concede'. Englert 1978, in discussing the transcription of stress and length, contrasts transcriptions like *haúru* (*a* and *u* pronounced separately and stress on *u*), *maîka* 'banana' (*i* is longer than the preceding *a* and bears stress), and *máûa* 'we-*dual, exclusive*' (*u* is longer than *a* but *a* bears stress) (Englert 1978:10). Englert also uses circumflexes in other positions, presumably to indicate length: *hâpaki* 'to hit', *mâkona* 'to eat one's fill'. Perhaps loss of contrastive length has been completed only in recent generations.

FUTUNA-ANIWA. Futuna-Aniwa (Dougherty 1983, Capell 1984) seems to be similar to Rapa Nui. The only relic of length is some words with final stress (including monosyllables) in an otherwise mostly-penultimate system, such as [afá] 'storm, hurricane' (< *afaa), [fatú] 'star' (< *fetu'u), [t^{h}ú] 'upright' (< *tu'u), presumably through a process like *fetú'u > *fetúu > *fetú, where a final vowel shortened but remained stressed. A long vowel can result from optional consonant deletion, as in [fakaɾa] ~ [faːɾa] 'to sunbathe' (Dougherty 1983:12-13; 197 for *fakara*).

It seems plausible that a broader system of contrastive stress is developing. The default pattern (Dougherty 1983:13) is generally for penultimate stress ([fetakáro] 'idle'), but stress is

antepenultimate if the word ends VV ([sikófia] ‘to grasp’) or VVCV ([móeŋa] ‘clothing’, though the generalization appears to depend on whether the two vowels are considered tautosyllabic). But, in addition to the exceptional final stresses, there are exceptional preantepenultimate stresses ([wárusia] ‘scrape’) and antepenultimate stresses that don’t meet the conditions for predictable antepenultimate stress ([pákasi] ‘pig’) (Dougherty 1983:14).

VAEAKAU-TAUMAKO. Vaeakau-Taumako (Næss & Hovdhaugen 2011), with a basically penultimate stress system, has retained contrastive vowel length only in final and penultimate syllables, but with a gap for HL#. That is, words can end LL# (*papa* ‘plank’), LH# (*papaa* ‘district’), or HH# (*paapaa* ‘vulva’), but not HL# (Næss & Hovdhaugen 2011:26). Otherwise, length is subject to a fair amount of free variation, and some predictable lengthening of stressed vowels. Many words that in conservative pronunciation have final stress are shifting towards penultimate stress, or even antepenultimate if certain vowel sequences are involved (conservative *haláa* vs. innovative *hála* ‘if’; *kaiáa* vs. *káia* ‘steal’; *tauí* vs. *táui* ‘price’; *taumíi* vs. *táumi, taúmi* ‘fish with traps’), suggesting that the length contrast is also being lost from final position.

The languages in the section illustrate the diachronic vulnerability of the Central Pacific length contrast, particularly its propensity to be reanalyzed in terms of a stress contrast. They also illustrate the uncertainty that can exist, for the analyst and presumably for the learner, in deciding whether a contrast is one or length or one of stress. For languages like Rotuman and (one analysis of) Rapa Nui, long vowels occur only to provide a bimoraic foot, and will therefore never produce ...HL#.

9. SUMMARY AND CONCLUSIONS. Although the Central Pacific languages probably inherited a trochaic-shortening system, very few retained it. Samoan is a strong case (except for the low productivity of de-shortening, and the availability of breaking under suffixation), and Fijian is a weaker case. But the rest of the languages examined here have gone in other directions. Table 6 organizes the languages by subfamily and pattern. (Emae is shown in two cells, because its stress pattern has changed, and it also now tolerates HL—although because Emae has antepenultimate stress, the crucial pattern is HLL# rather than HL#; see 7.3.) There seems to be little correlation between genetic affiliation and phonological behavior, which would be consistent with some patterns’ having been multiply innovated.

Table 6 <INSERT TABLE 6 ABOUT HERE>

A few points of theoretical interest have emerged in the course of this survey. First is the fragility of neutralizing alternations. De-shortening seems to be unstable, presumably because it is hard to learn. On encountering a citation form [tóso], a learner can't decide whether the underlying form is /toso/ or /tooso/ without knowing a suffixed form and recognizing it as related. Albright (2002) proposes that the underlying form (or base, in his terms) has to be taken from a surface allomorph, with the same member of each paradigm being used for each word. In these languages, a suffixed member of the paradigm would not be suitable to use as the underlying form, because many stems lack any suffixed forms. If the unsuffixed form is used as the base, then de-shortening in a suffixed form is an exceptional behavior that must be memorized, and is vulnerable to being lost diachronically.

Second is the viability of listing whole words. In Samoan, and possibly in Tuvaluan and Tokelauan, some de-shortened words seem to have been given their own lexical entries, with idiosyncratic semantics—quite plausibly their vowel length is treated as idiosyncratic too. The existence of other morphological length alternations in a language may support this, or perhaps the original causal mechanism was the other way around: if the learner has already decided that some or all de-shortening must be memorized as exceptional, he or she may be predisposed to allow other morphology to bear unpredictable lengthening too. Samoan is yet another case in which we shouldn't assume that every pattern we observe is driven by a productive phonological alternation.

Third is the relationship between alternation and morpheme structure constraints. Paster (2013) argues that a morpheme structure constraint and a similar alternation are separate phenomena, not in need of a unified analysis. Morpheme structure constraints and alternations tend to arise in tandem diachronically. For example, the Proto-Central-Pacific loss of intervocalic *R and change from *...aRá... to *...á... results in a lack of morphemes of the form /...aaCV/, as words like *kaRáva change to *káva. It also produces length alternations under suffixation, as hypothetical suffixed *mbaRá-na changes to *mbána, but unsuffixed *mbáRa changes to*mbáa, with no shortening (because it contained the sequence *áRa, not *aRá). Synchronically, however, Paster argues that the morpheme structure constraint and the alternation need not be linked. One can lose productivity while the other remains robust, or they can drift apart in their

details. We see this in the Central Pacific case. For example, Rennell-Bellona inherited shortening and de-shortening alternations, but now enforces *HL# through breaking rather than through shortening. Samoan appears to be losing breaking in unaffixed words, but retains it as an option under suffixation.

Appendix: languages surveyed

Table 7 <INSERT TABLE 7 HERE>

REFERENCES

ALBRIGHT, ADAM. 2002. *The identification of bases in morphological paradigms*. Los Angeles, CA: UCLA dissertation.

ANDERSON, VICTORIA, and YUKO OTSUKA. 2006. The phonetics and phonology of "definitive accent" in Tongan. *Oceanic Linguistics* 45.21–42.

BAUER, WINIFRED. 1993. *Maori*. New York, London: Routledge.

BECKMAN, JILL. 1998. *Positional faithfulness*. Amherst, MA: University of Massachusetts, Amherst dissertation.

BENUA, LAURA. 1997. *Transderivational identity: phonological relations between words*. Amherst, MA: University of Massachusetts, Amherst dissertation.

BESNIER, NIKO. 1995. *Literacy, emotion and authority: reading and writing on a Polynesian atoll.* Cambridge: Cambridge University Press.

BESNIER, NIKO. 2000. *Tuvaluan: a Polynesian language of the central Pacific*. New York: Routledge.

BICKMORE, LEE. 1995. Refining and formalizing the Tahitian stress placement algorithm. *Oceanic Linguistics* 34.410–442.

BIGGS, BRUCE. 1975. *A Mele-Fila vocabulary*. Auckland: Linguistic Society of New Zealand.

BILLINGTON, MICHAEL. n.d. Samoan language resources. Online: http://mike.bitrevision.com/samoan/.

BLEVINS, JULIETTE. 1994. The bimoraic foot in Rotuman phonology and morphology. *Oceanic Linguistics* 33.491–516.

BLUST, ROBERT. 2007. Disyllabic attractors and anti-antigemination in Austronesian sound change. *Phonology* 24.1–36.

BOERSMA, PAUL. 1998. *Functional phonology: formalizing the interaction between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.

BOERSMA, PAUL and JOE PATER. 2016. Convergence properties of a Gradual Learning Algorithm for Harmonic Grammar. *Harmonic Grammar and Harmonic Serialism*, ed. by John J. McCarthy & Joe Pater, 389-434. London: Equinox Press.

BUCKLEY, EUGENE. 1994. Persistent and cumulative extrametricality in Kashaya. *Natural Language & Linguistic Theory* 12.423–464.

CALABRESE, ANDREA. 1986. Metaphony in Salentino. *Rivista di grammatica generativa* 9–10.1–141.

CAPELL, ARTHUR. 1942. Notes on the Fila language, New Hebrides. *Journal of the Polynesian Society* 51.153–180.

CAPELL, ARTHUR. 1957. *A new Fijian dictionary*. 2nd edition. Glasgow: Wilson Guthrie & Co.

CAPELL, ARTHUR. 1962. *The Polynesian language of Mae (Emwae), New Hebrides*. Auckland: Linguistic Society of New Zealand.

CAPELL, ARTHUR. 1984. *Futuna-Aniwa dictionary, with grammatical introduction*. Canberra, A.C.T., Australia: Dept. of Linguistics, Research School of Pacific Studies, Australian National University.

CARROLL, VERN. 1965. *An outline of the structure of the language of Nukuoro*. Welllington: Polynesian Society.

CARROLL, VERN and TOBIAS SOULIK. 1973. *Nukuoro lexicon.* Honolulu: University of Hawai'i Press.

CHURCHWARD, C. MAXWELL. 1940. *Rotuman grammar and dictionary*. Sydney: Australasia Medical Publishing Co. [Reproduced 1978, New York: AMS Press].

CHURCHWARD, C. MAXWELL. 1953. *Tongan grammar*. London: Oxford University Press.

CHURCHWARD, C. MAXWELL. 1959. *Tongan dictionary*. London: Oxford University Press.

CLARK, ROSS. 1998. *A dictionary of the Mele language.* Canberra: Research School of Pacific and Asian Studies, Australian National University.

CLEMENTS, G. N. 1997. Berber syllabification: derivations or constraints? *Derivations and constraints in phonology*, ed. by Iggy Roca, 289–330. Oxford: Oxford University Press.

CONDAX, IOVANNA. 1990. Locative accent in Samoan. *Oceanic Linguistics* 29.27–48.

DAVIDSON, JEREMY H. C. S. 1990. *Pacific island languages: essays in honour of G. B. Milner*. London & Honolulu: School of Oriental and African Studies, University of London, and University of Hawaii Press.

DAVLETSHIN, ALBERT. 2014. A seemingly on-going sound change in Takuu language of Papua New Guinea: historical and theoretical implications. *Journal of Language Relationship* 12.1–20.

DE CHENE, BRENT. 1979. *The historical phonology of vowel length*. Bloomington: Indiana University Linguistics Club.

DE LACY, PAUL. 1996. Circumscription revisited: an analysis of Māori reduplication. University of Auckland, MS.

DE LACY, PAUL. 2002a. The formal expression of markedness. Amherst, MA: University of Massachusetts, Amherst dissertation.

DE LACY, PAUL. 2002b. Maximal words and the Māori passive. In Rackowski & Richards (eds.), 20-30.

DELL, FRANÇOIS and MOHAMED ELMEDLAOUI. 2002. *Syllables in Tashlhiyt Berber and in Moroccan Arabic*. Dordrecht: Springer.

DIXON, R. M. W. 1988. *A Grammar of Boumaa Fijian*. Chicago: University of Chicago Press.

DOUGHERTY, JANET. 1983. *West Futuna-Aniwa: An Introduction to a Polynesian Outlier Language*. University of California Press.

DU FEU, VERONICA. 2012. *Rapanui: A Descriptive Grammar*. London & New York: Routledge.

ELBERT, SAMUEL H. 1948. *Grammar and comparative study of the languages of Kapingamarangi: texts, and word lists*. New York: Viking Fund.

ELBERT, SAMUEL H. 1953. Internal relationships of Polynesian languages and dialects. *Southwestern Journal of Anthropology* 9.147–173.

ELBERT, SAMUEL H. 1988. Echo of a culture: A grammar of Rennell and Bellona. *Oceanic Linguistics Special Publications 22*, ed. by Albert Schütz, i-205.

ELBERT, SAMUEL H. and MARY KAWENA PUKUI. 1979. *Hawaiian grammar*. Honolulu: University of Hawaii Press.

ENGLERT, SEBASTIÁN. 1978. *Idioma rapanui: gramática y diccionario del antiguo idioma de la Isla de Pascua*. Santiago de Chile: Universidad de Chile.

ETHNOLOGUE & RAYMOND GORDON. 2005. *Ethnologue: languages of the world*, Fifteenth edition. Dallas, TX: SIL International.

FONG, EUGENE. 1979. Vowel and consonant quantity in Abruzzese. *Orbis* 28.277–289.

FUENTES, JORDI. 1960. *Diccionario y gramática de la lengua de la Isla de Pascua: pascuense-castellano: castellano-pascuense*. Santiago de Chile: Editorial Andrés Bello.

GARELLEK, MARC and JAMES WHITE. 2010. The acoustic correlates of stress and their use in diagnosing syllable fusion in Tongan. *UCLA Working Papers in Phonetics* 108.35–65.

GARELLEK, MARC and JAMES WHITE. 2015. Phonetics of Tongan stress. *Journal of the International Phonetic Association* 45.13–34.

GERAGHTY, PAUL. 1990. Proto Eastern Oceanic *R and its reflexes. In Davidson, 51–93.

GOLDWATER, SHARON and MARK JOHNSON. 2003. Learning OT constraint rankings using a Maximum Entropy model. *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, ed. by Jennifer Spenader, Anders Eriksson & Östen Dahl, 111–120. Stockholm: Stockholm University.

GREENHILL, SIMON and ROSS CLARK. 2011. POLLEX-Online: The Polynesian Lexicon Project Online. *Oceanic Linguistics* 50.551–559.

GRÉZEL, ISIDORE. 1878. *Dictionnaire futunien-français: avec notes grammaticales.* Paris: Maisonneuve & cie.

HARLOW, RAY. 2007. *Māori: a linguistic introduction.* Cambridge, UK & New York: Cambridge University Press.

HAYES, BRUCE. 1985. Iambic and trochaic rhythm in stress rules. *Proceedings of BLS 11: parasession on poetics, metrics, and prosody*, ed. by Mary Niepokuj, Mary VanClay, Vassiliki Nikiforidou & Deborah Feder, 429–46. Berkeley: Berkeley Linguistic Society.

HAYES, BRUCE. 1995. *Metrical Stress Theory: principles and case studies*. Chicago: University of Chicago Press.

HAYES, BRUCE. 2004. Phonological acquisition in Optimality Theory: the early stages. In Kager, Pater & Zonneveld, 158-203.

HEWITT, MARK and MEGAN CROWHURST. 1996. Conjunctive constraints and templates in Optimality Theory. *Proceedings of the North East Linguistic Society 26*, ed. Jill Beckman, 101–116. Amherst, MA: GLSA Publications.

HOOPER, ROBIN. 1996. *Tokelauan*. Munich: Lincom Europa.

HOVDHAUGEN, EVEN. 1990. Vowel lengthening in Samoan. In Davidson, 95-103.

HOVDHAUGEN, EVEN. 1992. Phonetic vowel length in Samoan. *Oceanic Linguistics* 31.281–285.

KAGER, RENÉ. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

KAGER, RENÉ; JOE PATER; and WIM ZONNEVELD (eds.) 2004. *Constraints in Phonological Acquisition*. Cambridge: Cambridge University Press.

KENSTOWICZ, MICHAEL. 1996. Base-identity and uniform exponence: alternatives to cyclicity. *Current trends in phonology: models and methods*, ed. by Jacques Durand & Bernard Laks, 363–393. Manchester: European Studies Research Institute, University of Salford.

KENSTOWICZ, MICHAEL. 1997. Quality-sensitive stress. *Rivista di Linguistica* 9.157–187.

KISSEBERTH, CHARLES. 1970. On the functional unity of phonological rules. *Linguistic Inquiry* 1.291–306.

KUKI, HIROSHI. 1970. *Tuamotuan phonology*. Canberra: Australian National University.

KUO, GRACE and CHAD VICENIK. 2012. The intonation of Tongan. *UCLA Working Papers in Phonetics* 111.63–91.

LANYON-ORGILL, PETER. 1944. *A study of the Leuangiua language*. London: Luzac & Company.

LAZARD, GILBERT and JOHN PELTZER. 2000. *Structure de la langue tahitienne*. Paris: Peeters.

LEGENDRE, GÉRALDINE, ANTONELLA SORACE and PAUL SMOLENSKY. 2006. The Optimality Theory–Harmonic Grammar connection. *The Harmonic Mind*, ed. by Paul Smolensky & Géraldine Legendre, 339–402. Cambridge, MA: MIT Press.

LIEBER, MICHAEL and KALIO DIKEPA. 1974. *Kapingamarangi lexicon*. Honolulu: University of Hawaii Press.

LYNCH, JOHN. 1998. *Pacific languages: an introduction*. Honolulu: University of Hawaii Press.

MAYER, JOHN F. 2001. *Code-switching in Samoan: T-style and K-style.* Honolulu: University of Hawai'i dissertation.

MCCARTHY, JOHN J. and ALAN PRINCE. 1993. Generalized Alignment. *Yearbook of Morphology 1993*, 79–154.

MELLANDER, EVAN. 2005. Generating configurational asymmetries in prosodic phonology. *UG and external systems: language, brain and computation*, ed. by Anna Maria Di Sciullo, 129–151. Amsterdam & Philadelphia: John Benjamins.

MESTER, ARMIN. 1992. Some remarks on Tongan stress. MS, UC Santa Cruz.

MILNER, G. B. 1993. *Samoan dictionary: Samoan-English, English-Samoan*. Auckland: Polynesian Press.

MOSEL, ULRIKE and EVEN HOVDHAUGEN. 1992. *Samoan reference grammar*. London & Tromsø: Oxford University Press.

MOSEL, ULRIKE and AINSLIE SO'O. 1997. *Say it in Samoan*. Canberra: Department of Linguistics, Research School of Pacific and Asian Studies, Australian National University.

MOYLE, RICHARD. 2011. *Takuu grammar and dictionary: a Polynesian language of the South Pacific*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.

MOYSE-FAURIE, CLAIRE. 1993. *Dictionnaire futunien-français avec index français-futunien*. Paris: Peeters.

MUTU, MARGARET and BEN TEÌKITUTOUA. 2002. *Ùa Pou : aspects of a Marquesan dialect. (Pacific Linguistics 533)*. Canberra: Research School of Pacific and Asian Studies, Australian National University.

MYERS, SCOTT. 1987. Vowel shortening in English. *Natural Language and Linguistic Theory* 5.485–518.

NÆSS, ÅSHILD and EVEN HOVDHAUGEN. 2011. *A grammar of Vaeakau-Taumako*. Berlin & Boston: Walter de Gruyter.

PASTER, MARY. 2013. Rethinking the 'duplication problem.' *Lingua* 126.78–91.

PAWLEY, ANDREW. 1960. Samoan phonology in outline. *Te Reo* 3.47–50.

PEPERKAMP, SHARON. 2005. A psycholinguistic theory of loan adaptations. *Proceedings of the 30th Annual Meeting of the Berkeley Linguistics Society*, ed. by Marc Ettlinger, Nicholas Fleischer & Mischa Park-Doob, 341–352. Berkeley, CA: Berkeley Linguistics Society.

POSER, WILLIAM. 1985. Cliticization to NP and Lexical Phonology. *Proceedings of the West Coast Conference on Formal Linguistics* 4, 262–272.

PRATT, GEORGE. 1878. *A grammar and dictionary of the Samoan language*, 2nd edtion, ed. by S. J. Whitmee. London: Trübner & Company.

PRINCE, ALAN. 1983. Relating to the grid. *Linguistic Inquiry* 14.19–100.

PRINCE, ALAN. 1990. Quantitative consequences of rhythmic organization. *Parasession on the Syllable in Phonetics and Phonology*, ed. by Michael Ziolkowski, Manuela Noske & Karen Deaton, 355–398. Chicago: Chicago Linguistic Society.

PRINCE, ALAN and PAUL SMOLENSKY. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA & Oxford: Blackwell.

PRINCE, ALAN and BRUCE TESAR. 2004. Learning phonotactic distributions. In Kager, Pater & Zonneveld (eds.), 245–291.

PUKUI, MARY KAWENA and SAMUEL H. ELBERT. 1986. *Hawaiian dictionary*. Honolulu: University of Hawaii Press.

RENSCH, KARL HEINZ. 1986. *Tikisionalio fakafutuna-fakafalani: dictionnaire futunien-français*. Canberra: Department of Linguistics, Research School of Pacific Studies, Australian National University.

RACKOWSKI, ANDREA and NORVIN RICHARDS (eds.) 2002. *Proceedings of AFLA 8: the Eighth Meeting of the Austronesian Formal Linguistics Association.* (MIT Working Papers in Linguistics 44).

ROSS, MALCOLM; ANDREW PAWLEY; and MEREDITH OSMOND. 1998. *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society. Volume 1: Material culture.* Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National Unversity.

ROSS, MALCOLM; ANDREW PAWLEY; and MEREDITH OSMOND. 2007. *The lexicon of Proto-Oceanic: the culture and environment of ancestral Oceanic society. Volume 2: The physical environment.* Canberra: Australian National University Press.

SCHÜTZ, ALBERT J. 1985. Accent and accent units in Maori : the evidence from English borrowings. *Journal of the Polynesian Society* 94.5–26.

SCHÜTZ, ALBERT J. 1985. *The Fijian language*. Honolulu: University of Hawaii Press.

SCHÜTZ, ALBERT J. 2001. Tongan accent. *Oceanic Linguistics* 40.307–323.

SCOTT, N. C. 1948. A study in the phonetics of Fijian. *Bulletin of the School of Oriental and African Studies* 12.737–752.

SIMONA, ROPATI; JUDITH HUNTSMAN; and ANTHONY HOOPER. 1986. *Tokelau dictionary*. Apia: Office of Tokelau Affairs.

SMOLENSKY, PAUL. 1995. On the internal structure of the constraint component Con of UG. Handout of a talk given at UCLA.

SPERLICH, WOLFGANG, ed. 1997. *Tohi vagahau Niue: Niue language dictionary*. Honolulu: Government of Niue & University of Hawaii Press.

STIMSON, J. FRANK and DONALD STANLEY MARSHALL. 1964. *A dictionary of some Tuamotuan dialects of the Polynesian language*. Salem, MA & The Hague: The Peabody Museum of Salem Massachusetts & Het Koninklijk Instituut voor Taal-, Land- en Volkenkunde.

TAUMOEFOLAU, MELENAITE. 2002. Stress in Tongan. In Rackowski & Richards (eds.), 341-354.

THIEBERGER, NICK. 2006. *A grammar of South Efate: An Oceanic language of Vanuatu*. Honolulu: University of Hawaii Press.

TSUKAMOTO, AKIHISA. 1988. *The language of Niuafo'ou Island*. Canberra: Australian National University dissertation.

ZURAW, KIE; KRISTINE MAK YU; ROBYN ORFITELLI. 2014. Word-level prosody in Samoan. *Phonology* 31.271–327.

* For material in the Samoan and Tongan sections, thanks to consultants John Fruean, Saia Moala, Piula Tonga, and Manu Tu'uholoaki; co-authors Kathleen O'Flynn, Robyn Orfitelli, Kaeli Ward, and Kristine Yu; Hilda Koopman and participants in the 2007-08 and 2009-10 linguistics field methods courses at UCLA. For discussion and comments, thanks to participants in the UCLA phonology seminar, AFLA 2014 at the University of Hawai'i, and NELS 2014 at MIT, and especially the anonymous reviewers and the associate editor Gene Buckley.

[1] That is, with stress on its first half. See section 2 for more on syllabification.

[2] Strictly speaking, the two 'tolerating HL#' candidates could be either HL or LLL, depending on syllabification. What is crucial is that placing stress on the antepenultimate mora requires violating either FOOTBINARITY or ALIGN.

[3] Geraghty (1990:73) gives, for the Proto-Central Pacific form, both *takuu- and *takuRi- (with no deletion of *R and an irregular change in vowel quality). As Geraghty discusses (Geraghty 1990:89-90), *R did not always delete between Proto-Eastern Oceanic and Proto-Central Pacific; deletion may have been partly conditioned by vowel quality, with deletion less likely between identical vowels or vowels of the same height more generally.

[4] Hovdhaugen (1990:97,102) states that although (C)V̄(C)V words exist, trisyllabic words ending HL, that is (C)V(C)V̄(C)V or (C)V̄(C)V̄(C)V, do not exist, and even proposes a phonotactic restriction against such words.

[5] Hovdhaugen (1992:283) does mention the scarcity of (C)V:CV# words 'both in my and Condax's data (and in Samoan!)', suggesting that some of these words might have been included in the duration measurements. But 'my data' could also refer to previous fieldwork not included in the measurements: towards the beginning of the paper Hovdhaugen says that '[m]y data (beside participant observations on Samoa during the last eight years) consist of 27 words and sentences' (Hovdhaugen 1992:282).

[6] A reviewer points out that awkwardly placed long vowels earlier in the word can be informative too. The constraints introduced so far have no objection to footings like [(H)L(LL)] or [L(H)(LL)]. ALIGN(PWord, R; Foot, R) only requires the word to end with a foot. But other foot-alignment constraints may have opinions about such footings. Using loan data, Zuraw, Yu and Orfitelli (2014) propose that there is a preference in Samoan for the word to begin with a foot, obeying ALIGN(PWord, L; Foot, L). Thus, when /paaŋotaa/ 'prisoner' is footed as [(pàa)ŋo(táa)], this is just

what would be expected even if there were no long vowel in play, and tells us nothing new about the constraint ranking. But when /tamaaloa/ ‘man’ (etymologically a compound, ‘child’+‘long’, but plausibly monomorphemic synchronically) is footed as [ta(màa)(lóa)] rather than *[(tàma)a(lóa)], this is unexpected.

If the correct analysis of ‘man’ actually has input /tama:loa/ and output [ta(mà:)(lóa)]—with a single, long vowel—then the explanation is that long vowels attract stress. That is, WEIGHTTOSTRESS (Prince 1990, Prince & Smolensky 2004) outranks ALIGN(PWord, L; Foot, L); *[(tàma)a(lóa)], which gets around WEIGHTTOSTRESS by eliminating the long vowel, is ruled out by NoBreaking. If we believe instead that the correct underlying form is /tamaaloa/, (or that a rich-base input like /tamaaloa/ would surface as [ta(màa)(lóa)]), then *ViV́i is insufficient to rule out *[(tàma)a(lóa)], because both successive [a]s are unstressed. (The dictionary contains hardly any monomorphemes of the shape LHH or LHLL, and ‘man’ is the only such item in our data set, so it is unknown whether there could be a contrast between words where the long vowel attracts stress and words where it does not.) We would need a constraint *$V_i)V_i$, penalizing a foot boundary between unstressed identical vowels. This constraint and *$V_iV́_i$ could be folded together into a single constraint *HETEROPODICV_iV_i, which penalizes any foot boundary (ending, beginning, or both) between identical vowels. Given the scarcity of data, I will leave this digression as a speculation.

[7] I have found just one, possibly archaic, counterexample elsewhere: Mosel and Hovdhaugen (1992:237) give [aala], alongside [feala], as a possible plural of [ala] ‘to be awake’. But Milner’s dictionary, where these counts come from, does not give any plural for that verb (not all verbs have a distinct plural form). Mosel and Hovdhaugen classify [aala] as derived by vowel lengthening rather than reduplication, along with examples like [faalute], plural of [falute] ‘to gather together’, and state that ‘[m]any of these [vowel-lengthening] plural forms are quite formal and they are usually not found in modern Samoan.’

[8] Mosel and Hovdhaugen (1992:31) describe shortening as happening ‘sometimes’, and give examples of words that shorten, such as *mālōloga* ‘rest’ (*g* spells [ŋ]); words that optionally shorten, such as *pelega/pelēga*, where our consultant has either shortening or breaking, as shown below; and, in a footnote, ceremonial words that cannot shorten: *umusā-ga* ‘ceremonial feast’. But, as mentioned above, they describe long penult vowels as pronounceable with stress on either the first or the second half, so it is possible that by *pelēga* they intend [peleéŋa], which matches our data.

[9] E.g. Stochastic Optimality Theory (Boersma 1998), Noisy Harmonic Grammar (Boersma 1998, Boersma & Pater to appear), and Maximum Entropy Optimality Theory (Goldwater & Johnson 2003).

[10] Examples in a compound: [(mèa)]$_{\text{PWord}}$-[a(ʔóŋa)]$_{\text{PWord}}$ 'homework' ('thing'+'learn'), [(vài)]$_{\text{PWord}}$-[(ínu)]$_{\text{PWord}}$ 'drinking water' ('water'+'drink'). With a bimoraic suffix: [(tùsi)]$_{\text{PWord}}$-[(ína)]$_{\text{PWord}}$ 'read-ERGATIVE'. With a causative prefix: [(fàʔa)]$_{\text{PWord}}$-[(ào)(ŋáa)]$_{\text{PWord}}$ 'use'. With two-syllable reduplication: [(ùsu)]$_{\text{PWord}}$-[(úsu)]$_{\text{PWord}}$ 'sing all day'.

[11] This is close to Paster 2013's 'loss of a static generalization' category of diachronic change. Paster discusses how these cases can be analyzed in OT with a patch such as a Positional Faithfulness constraint (Beckman 1998). In this case, we would need special faithfulness to stem-final vowels (as opposed to stem-penult vowels), which is not one of the positions Beckman documents as showing increased faithfulness. This is why I have opted for base-affixed faithfulness instead.

[12] Dixon cites reduplicated *ða-ðaa* 'lots of bad things' as evidence for underlying /ðaa/ rather than /ða/, but it is possible that reduplicant material does not contribute towards the word-size minimum, and the stem here still has to be lengthened to satisfy minimality.

[13] Dixon's mini-dictionary includes separate entries for *dono* 'right, correct, agreed' as a verb that cannot be suffixed and *doonu-ya* 'coincide with (e.g. time)' as a verb that is obligatorily suffixed. Although the meanings are similar (and both fall under the range of meanings that Capell gives for *donu*), the difference in the stem-final vowel may have prompted Dixon to treat them as unrelated.

Scott's other two alternating verbs do not appear at all in Dixon's mini-dictionary, though Dixon does have an entry for the noun *dravu* 'ashes'.

[14] Churchward (1953:13) reports it is possible to lengthen a penult for emphasis: *kúumi* (instead of normal *kúmi* 'to seek or search'), *káai* (instead of normal *kái* 'eat'), and gives some examples from legends.

[15] Specifically, one speaker tended to place the pitch rise on the second vowel in the sequence, and the other two varied in having the rise sometimes on the first vowel or sometimes on the second, with one of those two speakers also sometimes having the rise in the middle of the vowel sequence.

[16] From *taau-poqou (Greenhill & Clark 2011). If the morpheme boundary were synchronically active, we'd expect *[ta(àu)-po(óu)].

[17] Others are morphologically derived: 19 end with *-ŋa* and thus could be suffixed, with accompanying lengthening (*oka* 'to husk a coconut', *okāga* 'place where coconuts are husked'); in 25 of them the long vowel is word-initial or preceded by a vowel, and so the length could be the result of reduplication (*au* 'current', *āu* '(of current) be strong'); two are prefixed (*mā+ana* 'his/her/its' > *māna* 'him/her/it'). The remaining six are proclitic particles which are probably not final in their phonological word (*nāi* 'a few', precedes a noun).

[18] Proto-Polynesian *faqa-si.a 'part, side, half' > Nuclear Polynesian *waasi.

[19] Also listed as a derivative: *haatina* 'measure of distance using joints of body'.

[20] Davletshin (2014:4-5) states that 'a lexical word is always attested as part of either a noun phrase or a verb phrase', but the examples given are all of one- or two-syllable nouns and verbs, so it is not clear whether this contradicts Moyle's claim.

[21] Original: 'prononcées successivement (et sans glottale intermédiaire)'. Glosses are also translated.

[22] Here and in 49, the sources do not mark stress. Given the disagreement in the description of stress, I have not attempted to add stress marks.

[23] If a long vowel is really simply a sequence of two short vowels that happen to be identical—which should be true at least at the level of the rich base—then failing to stress a 'long vowel' is actually placing a foot boundary between two identical vowels: *[(mána)aki], *[(máta)a]. See fn. 6.

[24] Clark gives [máwa] as a narrower transcription, noting that there must be a contrast between underlying /u/, which counts for stress even if it becomes [w], and underlying /w/ (as in [ée-tawa] in 50), which does not count for stress.

[25] In which case the overall stress pattern could be described as a trochaic foot aligned to the right, with the final mora extrametrical. This would contradict Hayes 1995's 'somewhat [tentative]' ruling out of mora extrametricality, because of the 'absence of plausible cases' (Hayes 1995:58). Buckley (1994) argues in favor of an unpublished proposal of Steriade's to allow mora extrametricality.

[26] Glosses translated.

	/ma:li/	FOOTBINARITY-mora	ALIGN (PWord, R; Foot, R)	NO BREAKING	IDENT (long)
tolerating HL#	(má:li)	*			
tolerating HL#	(má:)li		*		
breaking	ma(áli)			*	
trochaic shortening	(máli)				*

<Tableau for example 2>

	...LH root /kul**i**:/	...LH root w/ suffix /kul**i**:-ŋa/	...HL root /m**a**:li/	...HL root w/ suffix /m**a**:li-ŋa/	HL# on surface?	$V_iÝ_i$ on surface?	lowest-ranked constraint
Trochaic shortening language	*no change* [ku(l**í**:)]	*shortening* [ku(l**í**-ŋa)]	*shortening* [(m**á**li)]	*de-shortening* [(m**à**:)(lí-ŋa)]	no	no	IDENT(long)
Breaking language		*breaking* [(kùl**i)(í**-ŋa)]	*breaking* [m**a(á**li)]	*de-breaking* [(m**à**:)(lí-ŋa)]	no	yes	NOBREAKING
Tolerating language		*no change* [ku(l**í**:)-ŋa] *or* [ku(l**í**:-ŋa)]	*no change* [(m**á**:)li] *or* [(m**á**:li)]	*no change* [(m**à**:)(lí-ŋa)]	yes	no	FOOTBINARITY or ALIGN

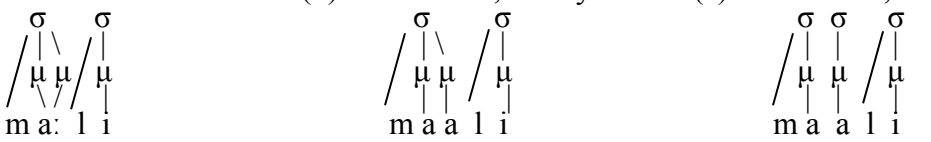
TABLE 1. Predicted typology

underlying structures

(a) one vowel: /ma:li/ (b) two vowels: /maali/

surface structures (syllabified)

(c) one vowel (d) two vowels, one syllable (e) two vowels, two syllables



<Example 3>

- *$V_iÝ_i$: a sequence of identical short vowels with stress on the second is prohibited

	/maali/	FOOTBINARITY-mora	ALIGN (Foot, R; Word, R)	*$V_iÝ_i$	MAX-V
tolerating HL#	(máali)	*			
tolerating HL#	(máa)li		*		
breaking	ma(áli)			*	
trochaic shortening	(máli)				*

<Tableau for example 4>

	...LH root /kul**ii**/	...LH root w/ suffix /kul**ii**-ŋa/	...HL root /m**aa**li/	...HL root w/ suffix /m**aa**li-ŋa/	HL# on surface?	$V_iV́_i$ on surface?
Trochaic shortening language	*no change* [ku(l**ii**)]	*shortening* [ku(l**í**-ŋa)]	*shortening* [(m**á**li)]	*de-shortening* [(m**àa**)(lí-ŋa)]	no	no

TABLE 2. Schematic trochaic shortening language

	/maali/	FOOTBINARITY-mora	ALIGN (Foot, R; Word, R)	*$V_iÝ_i$	MAX-V
tolerating HL#	(máali)	*!			
tolerating HL#	(máa)li		*!		
breaking	ma(áli)			*!	
trochaic shortening	☞ (máli)				*

<Tableau for example 9>

plural type	examples with C-initial base	# items	examples with V-initial base	# items	%V-initial
CV redup.	***sg.* lafi** ***pl.* lalafi** **'hide'**	**122**	***no examples***	**0**	**0%**
bimoraic redup.	*sg.* motu *pl.* motumotu 'break'	17	*sg.* saauni *pl.* saauniuni 'prepare'	1	6%
fe- and/or *-(C)i*	*sg.* toŋi *pl.* fetoŋi 'throw'	19	*sg.* oso *pl.* feoso(f)i 'jump'	8	30%
ta-	*sg.* sulu *pl.* tasulu 'insert'	16	*sg.* ili *pl.* taili 'blow'	4	20%
CV redup. or *ta-* (ambiguous)	*sg.* tanu *pl.* tatanu 'cover over'	13	*NA*	*NA*	*NA*
first syll. lengthened	*sg.* palaluu *pl.* paalaluu 'flap'	11	*sg.* ŋaosi *pl.* ŋaaosi 'prepare food'	3	21%
zero	*sg.* sili *pl.* sili 'put up'	12	*sg.* ulu *pl.* ulu 'go into'	3	20%
other (variation, removal of redup., multiple marking)	*sg.* sui *pl.* fesui ~ tasui 'change' *sg.* ʔote *pl.* feʔoteʔotei 'scold'	15	*sg.* uliuli *pl.* uli 'be black'	3	17%
no listed plural	*sg.* vase *pl. none* 'draw'	804	*sg.* usi *pl. none* 'melt'	145	15%
total		***1030***		***167***	***14%***

TABLE 3. Plural-marking patterns: CV reduplication is never used for a vowel-initial base

/pelee+ŋa/ or /pele:+ ŋa / base: [pelée]	FOOTBIN-μ	ALIGN (PWd,R; Foot, R)	DON'TSHORTEN-V́:-BA	*$V_iV́_i$	MAX-V or IDENT-(long)
tolerating HL# pe(léeŋa)	*!				
tolerating HL# pe(lée)ŋa		*!			
breaking ☞ pele(éŋa)				*	
trochaic shortening ☞ pe(léŋa)			*		*

<Tableau for example 12>

	...LH root /kul**ii**/	...LH root w/ suffix /kul**ii**-ŋa/	...HL root /m**aa**li/	...HL root w/ suffix /m**aa**li-ŋa/	HL# on surface?	$V_iV́_i$ on surface?
Breaking language	*no change* [ku(l**íi**)]	*breaking* [(kùl**i)(í**-ŋa)]	*breaking* [m**a(á**li)]	*de-breaking* [(m**àa**)(lí-ŋa)]	no	yes

TABLE 4. Schematic breaking language

	/maali/	FOOTBINARITY-mora	ALIGN (Foot, R; Word, R)	MAX-V	*$V_iV́_i$
tolerating HL#	(máali)	*!			
tolerating HL#	(máa)li		*!		
breaking	☞ ma(áli)				*
trochaic shortening	(máli)			*!	

<Tableau for example 18>

/peesi/	FOOTBINARITY-mora	ALIGN (PWord, R; Foot, R)	MAX-V	*$V_iÝ_i$	*AÍ
(péesi)	*!				
(pée)si		*!			
☞ pe(ési)				*	
(pési)			*!		

<Tableau for example 26a>

/paito/	FOOTBINARITY-mora	ALIGN (PWord, R; Foot, R)	MAX-V	*$V_iÝ_i$	*AÍ
(páito)	*!				
(pái)to		*!			
☞ pa(íto)					*
(páto)			*!		

<Tableau for example 26b>

/taaupoou/	FOOTBINARITY-mora	ALIGN (PWord, R; Foot, R)	MAX-V	*$V_iÝ_i$	*AÍ	PARSE-σ
(tàa)(ùpo)(óu)				*	*!	
☞ (tàa)upo(óu)				*		**
ta(àu)po(óu)				**!		**
(tàu)po(óu)			*!	*		*

<Tableau for example 26c>

/maalu/ no base		FOOTBIN-μ	ALIGN (PWd,R; Foot, R)	NOBREAKING-BA	MAX-V	*$V_iV́_i$
tolerating HL#	(máalu)	*!				
tolerating HL#	(máa)lu		*!			
breaking	☞ma(álu)					*
troch. short.	(málu)				*!	

<Tableau for example 34a>

/he+taa+ʔi/ base: [táa]		FOOTBIN- μ	ALIGN (PWd,R; Foot, R)	NOBREAKING- BA	MAX-V	*V_iÝ$_i$
tolerating HL#	he(táaʔi)	*!				
tolerating HL#	he(táa)ʔi		*!			
breaking	heta(áʔi)			*!		*
troch. short.	☞ he(táʔi)				*	

<Tableau for example 34b>

	...LH root /kul**ii**/	...LH root w/ suffix /kul**ii**-ŋa/	...HL root /m**aa**li/	...HL root w/ suffix /m**aa**li-ŋa/	HL# on surface?	$V_iV́_i$ on surface?
Tolerating language	*no change* [ku(l**íi**)]	*breaking* [ku(l**íi**)-ŋa] *or* [ku(l**íi**-ŋa)]	*no change* [(m**áa**)li] *or* [(m**áa**li)]	*no change* [(m**àa**)(lí-ŋa)]	yes	no

TABLE 5. Schematic tolerating language

	/maali/	MAX-V	*V_iV_i	FOOTBINARITY-mora	ALIGN (Foot, R; Word, R)
tolerating HL#	☞ (máali)			*?	
tolerating HL#	☞ (máa)li				*?
breaking	ma(áli)		*!		
trochaic shortening	(máli)	*!			

<Tableau for example 35>

	*...HL# and *V_iV_i enforced, mainly through shortening	*...HL# enforced, mainly through breaking → *V_iV_i violated	*...HL# violated (so no need to violate *V_iV_i); length alternations may persist	Stress pattern changes, ...HL# no longer problematic; length alternations may persist	Length reinterpreted as stress, faithful outputs pose no challenge to *...HL or *V_iV_i
Marquesic			Hawaiian Ùa Pou		
Tahitic			Tuamotuan Tahitian	Maori	
Rapa Nui					Rapa Nui
Samoan	Samoan				
Tokelauan			Tokelauan		
E. Uvean-Niuafo'ou		Niuafo'ou			
Ellicean		Tuvaluan	Nukuoro Kapingamarangi Takuu Ontong Java		
Futunic		Rennell-Bellona	East Futuna (Emae)	Mele-Fila (Emae)	Futuna-Aniwa Vaeakau-Taumako
Tongic		Tongan Niuean			
E. Fijian	Fijian				
W. Fijian-Rotuman					Rotuman

TABLE 6. Languages by affiliation and behavior

family						*language*	*country*
E. Fijian-Polynesian	Polynesian	Nuclear Polyn.	East Polyn.	Central Polyn.	Marquesic	Hawaiian	USA (Hawaii)
						Ùa Pou dialect of N. Marquesan	French Polynesia
					Tahitic	Maori	New Zealand
						Tuamotuan	French Polynesia
						Tahitian	French Polynesia
				Rapanui		Rapa Nui	Chile (Easter Island)
			Samoic-Outlier	E. Uvean-Niuafo'ou		Niuafo'ou	Tonga
				Ellicean		Tuvaluan	Tuvalu
						Nukuoro	Micronesia
						Kapingamarangi	Micronesia
						Takuu	PNG
						Ontong Java	Solomon Islands
				Futunic		East Futuna	Wallis and Futuna
						Rennell-Bellona	Solomon Islands
						Mele-Fila	Vanuatu
						Emae	Vanuatu
						Futuna-Aniwa	Vanuatu
						Vaeakau-Taumako	Solomon Islands
				Samoan		Samoan	Samoa
				Tokelauan		Tokelauan	Tokelau
		Tongic				Niue	Niue
						Tongan	Tonga
	E. Fijian					Fijian	Fiji
W. Fijian-Rotuman	Rotuman					Rotuman	Fiji

TABLE 7. Languages surveyed