The Samoan Reduplication System

Ethan Campbell-Taylor
Advisor: Ryan Bennett

Submitted to the faculty of the Department of Linguistics in partial fulfillment of the requirements for the degree of Bachelor of Arts

Yale University

April 20, 2016

Abstract

Reduplication is a morphophonological process wherein part or all of a stem appears twice, potentially with some modification, in the reduplicated form of the word. Although reduplication is a morphological process, in that it used to inflect and derive words, it is unique in that some (and usually all) of the phonological material it introduces originally comes from the base word. This interplay of phonology and morphology, along with its presence to some extent in virtually every natural language, has long made it a topic of interest to phonologists and to linguists more generally.

This paper is a thorough account of reduplication in Samoan. Samoan has two types of reduplication: CV reduplication and bimoraic reduplication. I use a Base-Reduplicant Correspondence Theory framework (McCarthy & Prince, 1995) to describe how the reduplicated forms of Samoan words surface by establishing a constraint ranking for each type of reduplication. Crucially to my analysis, I claim that Samoan reduplicants have moraic, not syllabic, template size. I also provide evidence for the claim by Mosel & Hovdhaugen (1992), among others, that the vowel lengthening that sometimes occurs with reduplication is, in fact, lexically determined.

Acknowledgements

I cannot sufficiently thank my advisor, Ryan Bennett, for all his help guiding my research and analysis throughout the development of this thesis. Many thanks are also due to Claire Moore-Cantwell, who loaned me both a paper Samoan dictionary and the digital Samoan database I used. I would also like to thank Sharon Inkelas: my final assignment for her morphophonology class I took at the 2015 Linguistics Summer Institute was the inspiration for this project. Finally, I must Raffaella Zanuttini and the other linguistics majors—Kyle, Alexa, Grace, Ralph, Alexandra, Rose, and Margaret—for all their helpful notes, guidance, and moral support throughout the writing process.

Contents

1 The Samoan Language

Samoan is a member of the Austronesian language family, spoken by approximately 370,000 people worldwide. Most of these speakers live in Samoa, where it is the national language, but it is also spoken in New Zealand and American Samoa (Gordon, 2005).

Samoan is notable for its two phonological registers, the colloquial *tautala leaga* and the more formal *tautala lelei*. The *tautala leaga* (meaning approximately "bad language") is used in virtually all spoken communication among native speakers, including formal contexts. The *tautala lelei* ("good language") is used for most Samoan writing, and is only obligatorily spoken when singing or speaking to God, though it is also used on the radio, in schools, and when addressing non-native speakers. The two registers have different phonological systems: the *tautala lelei* preserves all phoneme contrasts, while the *tautala leaga* collapses /t/ and /k/ to [k], /n/ and /ŋ/ to [ŋ], and /r/ and /l/ to [l] (Mosel & Hovdhaugen 1992:8). For this reason, all the data given in this paper are in the *tautala lelei* register.

The remainder of this paper is laid out as follows: the rest of Section 1 gives some of the more relevant features of Samoan, especially Samoan prosody. Section 2 is an overview of the process of reduplication generally. Section 3 describes and presents a formal analysis of CV reduplication, and Section 4 does the same for bimoraic reduplication. Section 5 describes the vowel lengthening that sometimes co-occurs with reduplication. Section 6 summarizes and concludes.

1.1 Samoan Prosody

1.1.1 Syllable Structure

In Samoan syllables, onsets are optional, while codas are forbidden. The language has a vowel length distinction as well as four "diphthongs," [ai, au, ei, ou]. Few sources agree on the exact list of diphthongs, but I have chosen to follow Zuraw et al. (2014) since they provide evidence for why those sequences are diphthongs. Namely, they demonstrate that only these vowel sequences disrupt the typical Samoan stress pattern where primary stress falls on the syllable containing the penultimate mora (further discussed in Section 1.1.2). A near-minimal pair is given in (1) and (2):

(1) ˈma.i.le
dog
'dog'

(2) ma.ˈe.la
hollow
'hollow'

I diverge from their analysis in that they are agnostic as to whether the vowel sequences in these diphthongs belong to the same syllable. I assume they do not; while there is no evidence that clearly indicates the status of these vowel sequences, not treating all vowel sequences the same way unnecessarily complicates my analysis. Bearing this in mind, I assume that [ai, au, ei, ou] are not true phonetic diphthongs, just vowel sequences that happen to disrupt the typical stress pattern of Samoan. In that sense, these are diphthongs from a prosodic standpoint, and so I hereafter refer to them as "prosodic diphthongs."

Relatedly, Zuraw et al. (2014) also remain agnostic as to whether long vowels are phonetically long, or if they are a sequence of two identical short vowels. I assume that long vowels are indeed phonetically long, and so belong to a single heavy syllable. All together, this means that Samoan syllables have four possible shapes: CV, V, CV:, and V:.

1.1.2 Syllable Weight and Stress

Samoan prosody is sensitive to syllable weight and to moras. Syllables with a long vowel are heavy, with two moras, while those with a short vowel are light, with only one mora.

Primary stress in Samoan always falls on the syllable containing the penultimate mora of the word. In most words, such as "*ŋafa*" in (3), the final two syllables are both light (CV or V), so the penultimate syllable bears primary stress:

(3) ˈŋa.fa
lineage
'lineage'

The penultimate syllable also is stressed if that syllable is heavy (i.e. the nucleus is a long vowel) and the final syllable is light (though see Section 2.3.1), as in (4):

(4) ˈmaː.nu
emerge
'to emerge'

If, however, the final syllable is heavy, as in (5), then the final syllable bears primary stress regardless of the penultimate syllable's weight:

(5) to.ˈtoː
belch
'to belch'

Finally, as seen above in (1), if the penult is the second vowel of a prosodic diphthong, then primary stress falls on the antepenult i.e. the syllable containing the first vowel of the prosodic diphthong. The tableau in (10) is a simplified Optimality Theory (Prince & Smolensky, 1993/2004) analysis of Samoan primary stress. I first define the constraints used in the ranking:

(6) FootBinarity$_\mu$ (Prince & Smolensky, 1993/2004)
Assign a violation mark for every foot that does not have exactly two moras.

(7) Rhythm-Type=Trochee (Prince & Smolensky, 1993/2004)
Assign a violation mark for every foot that does not have stress on its initial mora and only that mora.

(8) *aˈi (Zuraw et al., 2014)
Assign a violation mark for every unstressed non-high vowel immediately followed by a stressed high vowel.

(9) EDGEMOST-R (Prince & Smolensky, 1993/2004)

Assign a violation mark for every syllable between the end of the rightmost foot and the right edge of the word.

(10)

/maile/, 'dog'	FTBIN$_\mu$	RHTYPE=TROCHEE	*A'I	EDGEMOST-R
☞ a. (ˈma.i).le				*
b. ma.(ˈi.le)			*!	
c. ma.(i.ˈle)		*!		
d. (ˈma.i.le)	*!			

Secondary stress more or less follows this pattern; secondary-stressed feet are still moraic trochees. Notably, Samoan exhibits an initial dactyl effect (Prince, 1983) for stems with five or seven moras (virtually all of which are English loans):

(11) ˌte.mo.ka.ˈla.si
democracy
'democracy'

It is also important to note that bimoraic affixes (such as the reduplicants in bimoraic reduplication, see Section 4) are their own prosodic words, and so footing takes place within them without regard to the rest of the stem. Monomoraic affixes (such as the reduplicants in CV reduplication, see Section 3) belong to the same prosodic word as the stem they attach to, and so they are footed along with the rest of the stem.

Zuraw et al. (2014) does not have prosodic information on trimoraic affixes—they do not seem to exist in Samoan. I still consider a few candidates where the reduplicant has three moras e.g. candidate c in (53). The stress marks and syllable boundaries for such candidates are simply my best guesses following the above generalizations. For a thorough OT analysis of Samoan stress, see Zuraw et al. (2014).

2 Reduplication

2.1 What is reduplication?

Reduplication is a morphophonological process wherein to create a new word or form of a word, some substring of the word to be reduplicated is copied, potentially modified, and then attached to the word somewhere[1]. This description is intentionally vague; reduplication displays remarkable variety cross-linguistically in both form and function. It can be used for inflectional morphology, as in (12), or derivational morphology, as in (13)...

(12) **Warlpiri** (Nash, 1980, 130)
kuɽdu-kuɽdu
PL-child
'children'

(13) **Ulithian** (Lynch et al., 2002, 799)
sif-sifu
VERB-grass.skirt
'wear a grass skirt'

...it can copy anywhere from a single syllable, as in (14), to the entire word, as in (15)...

(14) **Agta** (Marantz, 1982, 439)
tak-takki
PL-leg
'legs'

(15) **Indonesian** (Cohn, 1989)
maʃarakat-maʃarakat
PL-society
'societies'

[1] At least superficially—it is challenging to describe the process in a theory-neutral way.

...and it can occasionally show other features such as fixed segmentism (see Alderete et al., 1999), as in English shm- reduplication (e.g. *Oedipus shmoedipus*, Nevins & Vaux (2003)).

Typologically, reduplication is extremely common, occurring to some extent in virtually every natural language (Downing & Inkelas, 2015, 526). It commonly has an iconic meaning, designating plurals, pluractionals, or intensives, but it can be used for a wide variety of syntactic and semantic functions. For instance, English has "contrastive focus reduplication", seen in sentences like those in (16) and (17):

(16) I'll make the tuna salad, and you make the SALAD–salad.

(17) Is he French or FRENCH–French?

Here, reduplication is used "to focus the denotation of the reduplicated element on a more sharply delimited, more specialized, range" (Ghomeshi et al., 2004, 308). In other words, the reduplicated form carries the stem's most typical meaning (in (16), a green salad, and in (17), a native of France).

Before proceeding further, I will define some important terms that will be used throughout this paper: the word which undergoes reduplication is the *stem*, the targeted substring which gets copied is the *base of copying*, and the extra material which appears in the reduplicated word is the *reduplicant*.

As reduplication involves both phonology and morphology, it has long been a topic of intense investigation in linguistics. Consequently, there have been numerous approaches put forth that aim to describe reduplication. I use Base-Reduplicant Correspondence Theory as the framework for my analysis.

2.2 Base-Reduplicant Correspondence Theory

Base-Reduplicant Correspondence Theory, hereafter BRCT (McCarthy & Prince, 1995), is an extension of Optimality Theory (Prince & Smolensky, 1993/2004) developed to account for reduplication. Under BRCT, reduplication is considered to be a form of phonological copying. The framework proposes the existence of an abstract morpheme, RED, which has no internal structure or phonological content. RED affixes to the stem undergoing reduplication and copies all of its phonological material from that stem[2].

Typical Optimality Theory has just one type of faithfulness constraint, Input-Output or IO constraints. Broadly speaking, these constraints penalize differences between the input and the output of the grammar. BRCT still has these faithfulness constraints (though McCarthy & Prince (1995) call them Input-Base constraints), but it has two other types as well. There are Base-Reduplicant faithfulness constraints, which penalize differences between the reduplicant and the base of copying, and there are Input-Reduplicant faithfulness constraints, which penalize differences between the reduplicant and the base of copying.

An unfortunate accident of terminology is that what McCarthy & Prince (1995) term "base" is not always the same thing as the "base of copying." In BRCT, the base is the part of the output that corresponds to the stem in the input. So in cases of full reduplication (where the base of copying is the entire stem), "base" and "base of copying" refer to the same string. In partial reduplication (where the base of copying is not the entire stem), the base is a substring of the base of copying. Since the base of copying is the only part of the base with any correspondents in the reduplicant, this means that for Base-Reduplicant faithfulness constraints, the reduplicant is exactly as faithful to the base as to the base of copying. However, for Input-Base constraints, there is a crucial difference between the two. The three types of faithfulness relevant to BRCT are shown in Figure 1.

Under the full model, every faithfulness constraint has three versions, e.g. MAX-BR, MAX-IB, and MAX-IR. In practice, Faith-IR constraints are very low-ranked and are almost always inactive; as seen in Sections 3 and 4, they are never relevant for Samoan reduplication. For these simpler cases, McCarthy & Prince (1995) also provide a simplified model without IR faithfulness constraints, shown in Figure 2.

[2]Except perhaps in cases of fixed segmentism (see Alderete et al., 1999), which does not occur in Samoan.

Figure 1: Full Model of Base-Reduplicant Correspondence Theory

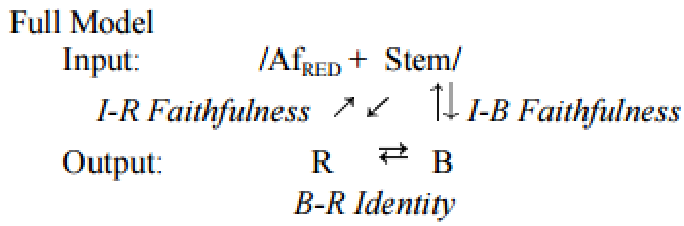
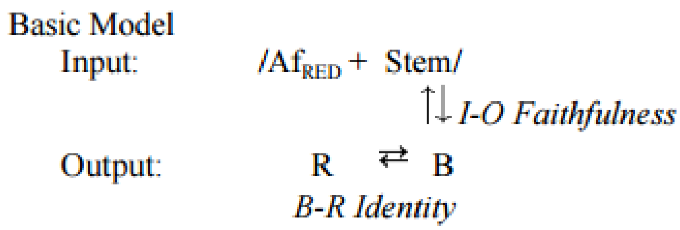


Figure 2: Basic Model of Base-Reduplicant Correspondence Theory



Example (18) is an illustrative tableau (from Downing & Inkelas, 2015, 517) that uses the Basic Model of BRCT to describe Sanskrit intensive reduplication.

(18)

RED-/svap/, 'sleep'	MAX-IO	*COMPLEX	MAX-BR
a. svaː-svap		**!	
☞ b. saː-svap		*	*
c. saː-sap	*!		

2.3 Data Collection

Before getting into the details of Samoan reduplication, a few notes about the data my analysis is based are in order. The data are taken from a digitized version of Milner's (1993) Samoan dictionary[3], generously provided to me by Claire-Moore

[3] I also used a paper version of an earlier edition, specifically Milner's (1966) dictionary, for additional reference and more specific definitions. For the purposes of this thesis, there do not seem to be many significant differences between the two. I cite both dictionaries when appropriate, though the more commonly cited one will be Milner (1993).

Cantwell. The database exists as an Excel spreadsheet, where each row is an entry for a different word and each column contains information from that entry, including page number, part of speech, and forms for various semantic functions. For more on the database, see the footnote in (Zuraw et al., 2014, 271).

While reduplication is used for several functions in Samoan, I restricted my search for words that have reduplicated forms (of either of the two modes) to just four functions: the plural forms of verbs, the frequentatives[4], the "faqa-verbs" (verbs with the prefix *faʔa-*, denoting a causative), and the derived verbs. I chose these functions because plurals and frequentatives are the most commonly used forms for CV reduplication and bimoraic reduplication, respectively. The prefix *faʔa-* is one of the only affixes that can co-occur with CV reduplication, so I decided those forms were also important to consider. As for verbs, Mosel & Hovdhaugen (1992) and personal visual inspection of the database indicated that there were a significant number of reduplicated forms in this category as well.

To actually collect all the words with reduplicated forms, I first converted the spreadsheet into a tab-delimited text file (after filling all the empty cells with dummy entries i.e. "!!"). I then wrote a Python script that read in each row and, given the stem, searched for several potential reduplicated outputs in the list of all the forms of the word. I tried to cover a wide range of possible outputs; in addition to standard CV and bimoraic reduplication, I also checked for forms with lengthened vowels and forms with shortened vowels. For each reduplicated form it found, the script wrote the entry number, the stem, and the reduplicated form in one of six separate text files, depending on what type of reduplication was used and if vowel lengthening, shortening, or neither was present.[5] The result was a total of 831 distinct reduplicated forms. It is possible that this misses some forms with particularly unusual semantics or phonological realizations, but I believe this is ample evidence to describe the

[4] "Frequentative" is the term used by both Milner (1966; 1993) and Mosel & Hovdhaugen (1992), though it seems to be a catchall term for true frequentatives, pluractionals, and some verbs with idiomatic semantics as well. I use the term when they do in the absence of anything more precise.

[5] 2 types of reduplication $\times$ 3 ways vowels can change in the output = 6

majority patterns.

2.3.1 Potential Data Issues

A problem facing this analysis is that while the dictionary I use (Milner, 1993) is from 1993[6], the prosody I cite is from a 2014 paper, Zuraw et al. (2014). As that paper points out, Samoan prosody appears to have changed somewhat in the past 50 years. There is certainly at least one major difference between the reported prosody in Zuraw et al. (2014) and what is actually in the data. Specifically, Zuraw et al. (2014) claim that all the words in Milner (1993) with heavy penults and light final syllables were produced by their consultants with a light penult. Moreover, the forms with heavy penults were judged as unacceptable by those speakers. However, aside from the issue of words with heavy penults, the minimal prosody described in Milner (1966) and Mosel & Hovdhaugen (1992) conforms with that in Zuraw et al. (2014), though it is much less thorough. I therefore assume that the account of Samoan prosody in Zuraw et al. (2014) is accurate for the data I use besides the few words with heavy penults and light final syllables.

2.4 Assumptions

Before getting into the details of the analysis, there are a few important assumptions I make. First, both CV and bimoraic reduplication have an extremely highly-ranked Locality constraint which no winning candidate ever violates:

(19) Locality (Lunden, 2004)
Assign a violation mark if Red copies nonadjacent phonological material.

This constraint means that we never see, for instance, /palu/ reduplicate to *[lu-palu]. Candidates violating this constraint will not be considered, and so Locality will not be listed in any tableaux. Relatedly, I assume that Red can only copy material from the structure it attaches to:

[6]Though again, I also make reference to the 1966 version of the dictionary, Milner (1966).

(20) RED + toʔulu → /to⟨RED⟩ʔulu/ → ˌto.ʔu.ˈʔu.lu,
RED + toʔulu → /to⟨RED⟩ʔulu/ → *ˌto.to.ˈʔu.lu
'they fall'

In CV reduplication, the type employed in (20), RED attaches to the primary-stressed syllable of the stem, in this case '*ʔu*'. Under this assumption RED can only copy that syllable, and not '*to*,' even though that syllable is adjacent.

Additionally, I assume that FOOTBINARITY$_\mu$, RHYTHM-TYPE=TROCHEE, and *AˈI[7] (see Section 1.1.2) are all undominated in all the tableaux in this paper. Candidates that violate these basic rules of Samoan prosody will not be considered, and these constraints will not be included in the tableaux.

Finally, though not critical to the analysis, it is worth noting that RED always infixes to the left of the base of copying. For most Samoan reduplicated forms, it is impossible to tell which string is the reduplicant and which is the base of copying. The exception are words with final long vowels that have undergone CV reduplication, like in (21):

(21) o.ˈtaː, ˌo.⟨ta⟩.ˈtaː
ripe ⟨FREQ⟩ripe
'to be ripe' 'they are ripe'[8]

It is clear from (21) that RED must have attached to the left of the base of copying; if it had attached to the right, that would mean that the vowels in the reduplicant and the base of copying had both had their length changed. For parsimony's sake, I assume that the reduplicant (which is underlined in all tableaux) lies to the left of the base of copying in all cases, not just words with final long vowels.

[7] EDGEMOST-R is another candidate that enforces standard Samoan prosody; however, as will be demonstrated in Section 3.2, it is not undominated.

[8] Samoan lacks person marking on verbs, but for readability's sake, all the examples of plural forms of verbs in this paper are translated in 3rd person. Singular verbs are given in the infinitive.

3 CV Reduplication

CV reduplication (a term I borrow from Zuraw et al., 2014) is one of the two modes of Samoan reduplication, so called because the reduplicant is always a CV syllable. It is not as productive as bimoraic reduplication, as it apparently cannot apply to loanwords (Mosel & Hovdhaugen, 1992, 219)[9]. It also has several morphological restrictions, rarely appearing in conjunction with any affixes besides the causative *faʔa-* and the ergativizing *-ina*, as seen in (22), respectively:

(22) ⟨tufa⟩tufa-ina
⟨FQV⟩share out-ERG
'to share out'[10]

3.1 Semantic Functions

CV reduplication is used for a variety of functions, but its most common use is to generate the plural forms of verbs. It also forms the frequentative for some lexical items, and it can form new verbs with unpredictable (though often related to the stem) semantics. Examples for each of these meanings are given in (23), (24), and (25):

(23) ˌte.le.ˈno.a, ˌte.le.⟨no⟩.ˈno.a
telenoa, ⟨PL⟩telenoa
'to be naked', 'they are naked'

[9]It's worth noting that Mosel & Hovdhaugen (1992) refer to CV reduplication as "partial reduplication" and bimoraic reduplication as "full reduplication." These terms are misleading; "partial reduplication" is generally used in the literature to describe any case where the base of copying is not the entire stem (Inkelas & Downing, 2015). In Samoan, both CV and bimoraic reduplication can be "partial" in this sense. "Total reduplication" is more commonly used than "full reduplication," but both mean that the base of copying is the entire stem. As will be seen in Section 4, bimoraic reduplication is only "total" in stems with two moras.

[10]Note that I have enclosed the reduplicant in angle brackets, indicating that is an infix, even though it appears at the start of the word—I assume that RED is always infixing, though in cases like the one here, only vacuously so i.e. it infixes after an empty string.

(24) ˈmi.lo, ⟨mi⟩.ˈmi.lo
twist, ⟨FREQ⟩twist
'to twist', 'twist repeatedly'

(25) ˈto.ʔo, ⟨to⟩.ˈto.ʔo
punting.pole, ⟨VERB⟩punting.pole
'punting pole', 'to lean on'

3.2 Form and Analysis

The general pattern in CV reduplication is that the base of copying is the syllable that bears primary stress (see Section 1.1.2 for more on Samoan stress), provided that syllable has an onset—words with onsetless primary-stressed syllables do not experience CV reduplication. The reduplicant is always a light syllable, even in cases where the syllable with primary stress is heavy. Since light syllables have a short vowel as their nucleus and Samoan has no complex onsets, all the copied syllables will be CV syllables.

In most cases, the final two syllables of a word are light, so the penultimate syllable bears primary stress and gets copied. An example of such a word is given in (26):

(26) ˈma.te, ⟨ma⟩.ˈma.te
die ⟨PL⟩die
'to die', 'they die'

For words of this form, we need just two constraints in our tableau. One constraint has to align RED with the primary-stressed syllable and another has to make sure RED copies no more than that syllable (more precisely, that it copies a single mora). The latter is a templatic constraint of the sort proposed in McCarthy & Prince (1993), specifically RED=σ_μ:

(27) RED=σ_μ
Assign a violation mark if the reduplicant is not a light syllable.

The alignment constraint is technically ALIGN(RED, R, $\acute{\sigma}$, L) (following example (15) from McCarthy & Prince, 1993, 34), but I abbreviate the constraint as simply ALIGN-TO-$\acute{\sigma}$ to improve legibility:

(28) ALIGN-TO-$\acute{\sigma}$

Assign a violation mark if the right edge of RED does not coincide with the left edge of the primarily stressed syllable.

A tableaux with constraints (27) and (28) is given below:

(29)

RED-/toʔulu/, 'to fall, drop'	ALIGN-TO-$\acute{\sigma}$	RED=σ_μ
☞ a. (ˌto.ʔu).(ˈʔu.lu)		
b. (ˌto.to).(ˈʔu.lu)	∗!	
c. (ˌto.ʔu).(ˈlu.lu)	∗!	
d. to.(ˌʔu.lu).(ˈʔu.lu)		∗!

Note that these constraints are unranked—either ranking produces the output we expect to surface.

In words that have heavy final syllables (i.e. the final syllable has a long vowel as its nucleus), the reduplicant is the shortened, CV version of that syllable, as seen in example (30):

(30) pa.ˈʔoː, ˌpa.⟨ʔo⟩.ˈʔoː
clash ⟨PL⟩clash
'to clash' 'they clash'

Our current constraint set cannot capture this generalization, as seen in the tableau in (31):

(31)

RED-/paʔoː/	ALIGN-TO-$\acute{\sigma}$	RED=σ_μ
a. (ˌpa.ʔo).(ˈʔoː)		
b. (ˌpa.pa).(ˈʔoː)	∗!	
c. pa.(ˌʔoː).(ˈʔoː)		∗!
d. pa.(ˈʔo.ʔo)		

This constraint set currently has no way of penalizing unfaithfulness in the stem, so candidate d, which has a shortened final vowel, is just as acceptable as the actual output, candidate a. In order for our ranking to generate the correct outuput, we need to add a faithfulness constraint, specifically IDENT(LEN)-IB:

(32) IDENT(LEN)-IB

Assign a violation mark for every vowel in the input that has a different feature value for length than its correspondent in the base.

Adding this constraint produces the correct output, as shown in (33):

(33)

RED-/paʔoː/	ALIGN-TO-$\acute{\sigma}$	RED=σ_μ	IDENT(LEN)-IB
☞ a. (ˌpa.ʔo).(ˈʔoː)			
b. (ˌpa.pa).(ˈʔoː)	∗!		
c. pa.(ˌʔoː).(ˈʔoː)		∗!	
d. pa.(ˈʔo.ʔo)			∗!

Once again, none of these constraints are ranked relative to each other since the winning candidate has no violation marks.

Words with a light final syllable but a heavy penult are rare, (though see Section 2.3.1), but they do have forms with CV reduplication. As with words with heavy final syllables like '*paʔoː*' in (30), in these forms the base of copying is shortened to CV. An example of a word with this structure is given in (34):

(34) ˈnaː.mu, ⟨na⟩.ˈna.mu
smell.of, ⟨FQV⟩smell.of
‘to smell of’, ‘to smell of’[11]

Our current constraint set does not generate the proper output for these words. A completely faithful CV reduplication performs better since it does not shorten any of the vowels in the input[12]:

(35)

RED-/naːmu/	ALIGN-TO-$\acute{\sigma}$	RED=σ_μ	IDENT(LEN)-IB
a. **.<u>na</u>.(ˈna.mu)**			∗!
b. (ˌ<u>naː</u>).(ˈnaː).mu		∗!	
☞ c. <u>na</u>.(ˈnaː).mu			
d. (ˌnaː).(ˈ<u>mu</u>.mu)	∗!		
e. (ˌ<u>naː</u>).(ˈna.mu)		∗!	∗

Clearly, we need to add a constraint that penalizes candidate c. The constraint I use has already appeared in this analysis in Section 1.1.2, EDGEMOST-R:

(36) EDGEMOST-R
Assign a violation for every syllable between the end of the rightmost foot and the right edge of the word.

As shown in (37) below, EDGEMOST-R penalizes candidate c, and so generates the correct output. It also establishes a constraint ranking, provided above the tableau in (37):

[11] This is not a typo—Milner (1966, 154) lists both of these forms with the same definition, even though one is marked as a frequentative.

[12] A brief notation definition: **bolded** candidates in tableaux are candidates we expect to win which are not optimal under the current ranking. Optimal candidates for the particular ranking are always marked with a ☞symbol.

(37) Align-to-$\acute{\sigma}$, Red=σ_μ, Edgemost-R >>Ident(len)-IB

Red-/naːmu/	Align-to-$\acute{\sigma}$	Red=σ_μ	Edgemost-R	Ident(len)-IB
☞ a. na.(ˈna.mu)				*
b. (ˌnaː).(ˈnaː).mu		*!	*	
c. na.(ˈnaː).mu			*!	
d. (ˌnaː).(ˈmu.mu)	*!		*	
e. (ˌnaː).(ˈna.mu)		*!		*

3.3 Ineffability with CV Reduplication

One final case of CV reduplication to consider is that of words with onsetless primary-stressed syllables. These words, like the one in (38), systematically do not undergo CV reduplication (Mosel & Hovdhaugen, 1992, 220):

(38) ˌsaː.ˈu.ni
prepare
'to prepare'

This is a case of ineffability—when a particular input lacks an output by virtue of its phonology. This appears to be in conflict with the fundamental Optimality Theory notion that constraints are violable: for any input, there is always some candidate which performs the best given a constraint ranking. There is no way to eliminate all candidates from contention.

The solution proposed by Prince & Smolensky (1993/2004) is that among the candidates Gen produces is the null parse, ⊙. The null parse clearly violates no traditional markedness constraints, but it also is defined to violate no faithfulness constraints either (for a detailed explanation of how this is possible, see McCarthy & Wolf, 2007). However, there still must be some constraint for it to violate so that it is not the optimal candidate every time. This constraint is MParse:

(39) MParse
Assign a violation mark if the candidate is the null parse.

There is another constraint we need to add. Even with MParse, the candidate we expect to win is one that faithfully copies the onsetless penult (in (41, candidate c). To penalize this candidate, we must add the constraint:

(40) NoBreaking (Zuraw et al., 2014)
Assign a violation mark for every sequence of adjacent identical vowels within a single prosodic word.

Adding MParse and NoBreaking to our constraint ranking produces the correct output, the null parse, for words with onsetless penults. In the new ranking, MParse is ranked below all the other constraints[13]:

(41) Align-to-$\acute{\sigma}$, Red=σ_μ, Edgemost-R, NoBreaking >>MParse

Red-/saːuni/	Align-$\acute{\sigma}$	Red=σ_μ	Edge-R	NoBreak	MParse
☞ a. ⊙					*
b. (ˌsaː).(ˈu.ni)		*!			
c. (ˌsaː).u.(ˈu.ni)				*!	
d. (ˌsaː).(ˈuː).ni			*!		
e. (ˌsaː).u.(ˈni.ni)	*!	*			
f. (ˌsaː).(ˌsaː).(ˈu.ni)	*!				
g. sa.(ˌsaː).(ˈu.ni)	*!				
h. (ˌsaː).(ˌu.ni).(ˈu.ni)		*!			

Some important notes on the above tableau: in candidate b, Red copies no material. The reduplicant, then, is not a light syllable, so that candidate violates Red=σ_μ. In candidate d, by contrast, the sequence /uu/ has been reanalyzed as [u:], so the reduplicant is still just one syllable.

The tableau in (41) excludes Ident(len)-IB, as it is inactive there. However, it is important to establish the ranking between Ident(len)-IB and MParse; an incorrect ranking would block an output from surfacing in a case where we do expect

[13] I have abbreviated some of the constraint names in the tableau for the sake of space. This will be done at various points throughout the thesis, but will not be explcitly pointed out any more.

an output. To see how MParse is ranked relative to Ident(len)-IB, we add it and the null parse to the earlier tableau from (37):

(42) MParse >>Ident(len)-IB

Red-/naːmu/, 'smell of'	Align-$\acute{\sigma}$	Red=σ_μ	Edge-R	MParse	Id(len)-IB
☞ a. <u>na</u>.(ˈna.mu)					*
b. ⊙				*!	
c. (ˌ<u>naː</u>).(ˈnaː).mu		*!	*		
d. <u>na</u>.(ˈnaː).mu			*!		
e. (ˌnaː).(ˈ<u>mu</u>.mu)	*!		*		
f. (ˌ<u>naː</u>).(ˈna.mu)		*!			*

(42) shows that MParse dominates Ident(len_IB). In (43), a tableau with the opposite ranking, Ident(len)-IB >>MParse, is provided. If this were the actual constraint ranking, we would expect words with heavy penults and light final syllables (as well as those with onsetless penults) to systematically not undergo CV reduplication:

(43) Ident(len)-IB >>MParse

Red-/naːmu/	Align-$\acute{\sigma}$	Red=σ_μ	Edge-R	Id(len)-IB	MParse
a. **<u>na</u>.(ˈna.mu)**				*!	
☞ b. ⊙					*
c. (ˌ<u>naː</u>).(ˈnaː).mu		*!	*		
d. <u>na</u>.(ˈnaː).mu			*!		
e. (ˌnaː).(ˈ<u>mu</u>.mu)	*!		*		
f. (ˌ<u>naː</u>).(ˈna.mu)		*!		*	

Again, this is not the generalization we see in the data: MParse must outrank Ident(len)-IB. Therefore, the final constraint ranking for CV reduplication is:

(44) Align-to-$\acute{\sigma}$, Red=σ_μ, Edgemost-R, NoBreaking >>MParse >>Ident(len)-IB

3.3.1 CV Reduplication with Prosodic Diphthongs

I was unable to determine what happens in cases where CV reduplication applies to words with a prosodic diphthong (see Section 1.1.1) before a final short syllable, like the one in (45):

(45) ˈma.i.le
dog
'dog'

It is not clear if there are no words with this prosodic shape that are eligible for partial reduplication, or if these potential forms simply do not surface due to ineffability. Neither of these explanations is inconceivable—for the first, many of the words that have penultimate diphthongs are loans from English e.g. '*paipa*', 'pipe', and loanwords do not undergo CV reduplication (Mosel & Hovdhaugen, 1992). Moreover, even non-loans, such as '*maile*' above, may not experience CV reduplication simply because they are not verbs.

As for the second explanation, it is possible that the base of copying for CV reduplication is not the syllable with primary stress, but rather the syllable containing the penultimate mora. These are always the same syllable except in words with a prosodic diphthong immediately preceding the final (light) syllable. In such words, the syllable containing the penultimate mora has no onset (in '*maile*', this syllable is just '*i*'). If that syllable is the base of copying, the ranking in (44) predicts there to be no form with CV reduplication for that word. With the data I have, however, there is no way to tell which is the base of copying.

4 Bimoraic Reduplication

The second type of Samoan reduplication is bimoraic reduplication. It is very productive for all of its semantic functions, as it does apply to loanwords. Unlike CV reduplication, it has very few morphological restrictions, only not appearing with the ergativizing suffix *-(C)ia*[14].

4.1 Semantic Functions

By far the most common function of bimoraic reduplication is forming the frequentative of verbs. It also is used to form the plural of some verbs, along with a variety of miscellaneous derivational meanings. Three examples with different semantics are given below:

(46) ma.ˈso.fa, ma.⟨ˌso.fa⟩.ˈso.fa
collapse, ⟨FREQ⟩collapse
'to collapse', 'to collapse utterly'

(47) ˌsaː.ˈu.ni, ˌsaː.⟨ˌu.ni⟩.ˈu.ni
prepare, ⟨PL⟩prepare
'to prepare', 'they prepare'

(48) ˈpa.ʔe, (⟨ˌpa.ʔe⟩).ˈpa.ʔe
bleach, ⟨DERG⟩bleach
'to bleach', 'to be pale'[15]

4.2 Form and Analysis

Unlike CV reduplication, bimoraic reduplication copies moraic trochees (two light syllables or one heavy syllable) rather than a single CV syllable. It always copies either the first two or the last two moras of the base word, depending on the presence of certain structures in the stem.

The simplest cases of bimoraic reduplication are those where the stem is a CVCV bimoraic trochee, such as *sina*, 'white':

[14] "(C)" here indicates a thematic consonant, which is a lexically determined (and potentially null) consonant that appears with a number of suffixes. It is assumed that these consonants once belonged to the stems, but were reanalyzed as the onsets of these suffixes as the language lost codas (Mosel & Hovdhaugen, 1992, 192)

[15] DERG means de-ergative, as this seems to be the transformation from 'bleach' to 'be pale'.

(49) ˈsi.na, ⟨ˌsi.na⟩.ˈsi.na
white, ⟨FREQ⟩white
'to be white', 'to be dazzling white'

At this point in the analysis, the only constraint we need is a templatic constraint along the lines of RED=σ_μ from Section 3.2. Since the reduplicant is two moras, we use RED=$\mu\mu$:

(50) RED=$\mu\mu$
Assign a violation mark if the reduplicant does not have exactly two moras.

Among the candidates we consider is the null parse. As stated in Section 3.3, ineffability results for some inputs to CV reduplication—we should consider it as a possibility for bimoraic reduplication too. We therefore include MPARSE (see (39)) in the tableau as well:

(51)

RED-/sina/	MPARSE	RED=$\mu\mu$
☞ a. (ˌsi.na).(ˈsi.na)		
b. ⊙	*!	
c. si.(ˈsi.na)		*!

Either ranking of these constraints gives the correct output. As there are no Samoan stems which, by virtue of their phonology, exhibit ineffability in the output of bimoraic reduplication, I therefore assume that MPARSE is undominated. To that end, MParse will be left off of future tableaux.

Slightly more complex are stems like (52) which consist of three CV syllables:

(52) ma.ˈte.le, ma.⟨ˌte.le⟩.ˈte.le
hurry, ⟨FREQ⟩hurry
'to hurry', 'to hurry repeatedly, to be rash'

As in CV reduplication, in general, RED attaches to the left of the primary-stressed syllable. We therefore need to include ALIGN-TO-$\acute{\sigma}$ (defined earleir in (28)) in our tableaux for bimoraic reduplication:

(53)

RED-/matele/	RED=$\mu\mu$	ALIGN-TO-$\acute{\sigma}$
☞ a. ma.(ˌ<u>tele</u>).(ˈtele)		
b. (ˌ<u>ma.te</u>).ma.(ˈte.le)		∗!
c. <u>ma.(ˌte.le</u>).ma.(ˈte.le)	∗!	∗

Either ranking of RED=$\mu\mu$ and ALIGN-TO-$\acute{\sigma}$ gives the correct output, so they are unranked for the moment.

I mentioned earlier that the majority pattern is for RED to align to the primary-stressed syllable. One exception is if the primary-stressed foot contains a hiatus:

(54) ma.ʔa.i, ⟨ˌma.ʔa⟩.ma.ˈʔa.i
sharp, ⟨FREQ⟩sharp
'to be sharp', 'to be fierce'

In (54), RED aligns with the start of the word, not the primary-stressed syllable. There is a systematic avoidance of copying hiatuses motivating this non-standard alignment. In order for our tableaux to produce the correct output for these words, we need to add the constraint *HIATUS:

(55) *HIATUS[16]
Assign a violation mark for every sequence of two adjacent vowels.

However, there are other ways of avoiding hiatus beyond selecting a different base of copying. Namely, one could insert segments (perhaps /ʔ/) to separate the vowels, or simply delete one of the vowels. To discourage these strategies, I inroduce two faithfulness constraints, MAX-BR and DEP-BR:

[16]Technically, every instance of hiatus is also an instance of an onsetless syllable, so in principle, ONSET could also be used to avoid copying hiatuses. In fact, because Samoan does not have codas, any onsetless non-initial syllable is part of a hiatus. Furthermore, because DEP-BR is undominated (shown in (58)), you could replace *HIATUS with ONSET in every tableau in this paper and still get the correct outputs. However, ONSET seems like a unlikely markedness constraint for Samoan; Samoan has a large number of syllables without onsets, but it never attempts to "repair" any of them.

(56) MAX-BR

Assign a violation mark for every segment in the input with no correspondent in the output.

(57) DEP-BR

Assign a violation mark for every segment in the output with no correspondent in the input.

Adding these three constraints to our ranking gives the following tableau:

(58) *HIATUS >>ALIGN-TO-σ́

RED-/maʔai/	DEP-BR	MAX-BR	RED=$\mu\mu$	*HIATUS	ALIGN-σ́
☞ a. (ˌma.ʔa).ma.(ˈʔa.i)				∗	∗
b. ma.(ˌʔa.i).(ˈʔa.i)				∗∗!	
c. ma.(ˌʔa.i).ma.(ˈʔa.i)			∗!	∗∗	∗
d. (ˌma.ʔa).ma.(ˈʔa.ʔi)	∗!				∗
e. (ˌma.ʔa).(ˈma.ʔa)		∗!			∗

The tableau in (58) shows that *HIATUS outranks ALIGN-TO-σ́, indicating that it is better to not copy a hiatus than to have optimal alignment of RED. Although (58) does not explicitly demonstrate that DEP-BR and MAX-BR outrank the other constraints, I assume that they do since neither one is ever violated by a winning candidate for any stem. Like MPARSE, they will be left off of future tableau for simplicity.

It is also important to note that we have not actually established a ranking between RED=$\mu\mu$ and ALIGN-TO-σ́, despite the appearance of the tableau. We also have yet to rank RED=$\mu\mu$ and *HIATUS.

There is a similar tendency in bimoraic reduplication to avoid copying long vowels:

(59) ˌma.ŋa.ˈŋuː, ⟨ˌma.ŋa⟩.ˌma.ŋa.ˈŋuː
mutter, ⟨FREQ⟩mutter
'to mutter', 'to talk all at once'

Just like in (54), the presence of a certain structure (in this case, a heavy syllable) causes RED to align with a the left edge of the word instead of the primary-stressed syllable. This pressure to avoid copying heavy syllable seems to be stronger than the pressure against copying hiatuses seen earlier:

(60) ˌŋa.o.ˈsaː, ⟨ˌŋa.o⟩.ˌŋa.o.ˈsaː
untidy, ⟨FREQ⟩untidy
'to be untidy', 'to be in disorder'

(61) ˌmaː.ˈva.e, ⟨ˌmaː⟩.ˌva.e.ˈva.e
apart, ⟨FREQ⟩apart
'to be apart', 'to be cracked'

This suggests that the constraint penalizing copying long vowels outranks *HIATUS. That constraint is *CLASH:

(62) *CLASH
Assign a violation mark for every sequence of adjacent stressed[17] syllables.

Because Samoan feet are moraic trochees, Samoan heavy syllables are always their own foot, and so are always stressed. Therefore, when a heavy syllable gets copied by bimoraic reduplication, it produces a new stressed syllable next to the old one, creating a stress clash. Adding *CLASH to the constraint ranking produces the following tableaux:

(63) *CLASH >>*HIATUS

RED-/ŋaosaː/	*CLASH	RED=$\mu\mu$	*HIATUS	ALIGN-TO-$\acute{\sigma}$
☞ a. (ˌŋa.o).(ˌŋa.o).(ˈsaː)			**	*
b. (ˌŋa.o).(ˌsaː).(ˈsaː)	*!		*	
c. ŋa.o.(ˌsaː).o.(ˈsaː)		*!	**	

[17]Primary or secondary stress.

(64) *Clash >>*Hiatus

Red-/maːvae/	*Clash	Red=$\mu\mu$	*Hiatus	Align-$\acute{\sigma}$
☞ a. (ˌmaː).(ˌva.e).(ˈva.e)	*		**	
b. (ˌmaː).(ˌmaː).(ˈva.e)	**!		*	*
c. (ˌmaː).va.(ˌmaː).(ˈva.e)	*	*!	*	*

As expected, *Clash outranks *Hiatus and, by transitivity, Align-to-$\acute{\sigma}$. Red=$\mu\mu$ remains unranked relative to all three of the other constraints.

For trimoraic Samoan words that end in a heavy syllable, the base of copying is the entire final syllable:

(65) ŋa.ˈpaː, ŋa.⟨ˌpaː⟩.ˈpaː
crack, ⟨freq⟩crack
'to crack (of a whip)', 'to crackle'

This is inconsistent with our current constraint ranking; copying the heavy syllable will always produce a stress clash, and *Clash is undominated. For the ranking to capture this generalization about words with final heavy syllables, it requires a constraint to dominate *Clash. This constraint is Ident(len)-BR:

(66) Ident(len)-BR
Assign a violation mark for every vowel in the base that has a different feature value for length than its correspondent in the reduplicant.

Effectively, this penalizes copying the first two moras of words like '*ŋapaː*' in (65), resulting in a shortened version of the stem's final vowel in the reduplicant. Adding this constraint gives the following tableau:

(67) Ident(len)-BR, Red=$\mu\mu$ >>*Clash

Red-/ŋapaː/	Id(len)-BR	Red=$\mu\mu$	*Clash	*Hiatus	Align-$\acute{\sigma}$
☞ a. ŋa.(ˌpaː).(ˈpaː)			*		
b. ŋa.(ˌpaː).ŋa.(ˈpaː)		*!			*
c. ŋa.(ˌpa).ŋa.(ˈpaː)	*!				*

This tableau demonstrates that RED=μμ and IDENT(LEN)-BR are ranked higher than *CLASH. This establishes the final constraint ranking for bimoraic reduplication, given in (68):

(68) IDENT(LEN)-BR, RED=μμ >>*CLASH >>*HIATUS >>ALIGN-TO-σ́

What follows are a number of prosodic shapes not yet examined that this ranking provides the correct output for. For readability, within each tableau I omit some constraints that assign no violations, though I list those constraints above the tableau.

First are words with all light syllables where the penult lacks an onset:

(69) Omitted Constraints: IDENT(LEN)-BR

RED-/fatuaʔi/, 'to consider'	RED=μμ	*CLASH	*HIATUS	ALIGN-TO-σ́
☞ a. (ˌfa.tu).(ˌfa.tu).(ˈa.ʔi)			*	*
b. (ˌfa.tu).(ˌa.ʔi).(ˈa.ʔi)			**!	
c. fa.tu.(ˌa.ʔi).tu.(ˈa.ʔi)	*!		**	*
d. (ˌfa.tu).a.(ˌfa.tu).(ˈa.ʔi)	*!		**	*

*HIATUS is the crucial constraint here, although it operates a little differently than it does in tableaux such as (58). Rather than avoiding copying a hiatus outright, it avoids copying an onsetless syllable so that a new hiatus will not arise in the output, like it does in candidate b.

The final constraint ranking also gives the correct output for words consisting of two heavy syllables. Here, copying either syllable will result in a (second) stress clash, so the syllable with primary stress is copied:

(70) Omitted Constraints: *HIATUS

RED-/maːluː/, 'to be cooling'	ID(LEN)-BR	RED=μμ	*CLASH	ALIGN-σ́
☞ a. (ˌmaː).(ˌluː).(ˈluː)			**	
b. (ˌmaː).(ˌmaː).(ˈluː)			**	*!
c. (ˌmaː).(ˌluː).(ˌmaː).(ˈluː)		*!	***	*
d. (ˌma.lu).(ˌmaː).(ˈluː)	**!		*	*

Finally, the ranking works for words with a sequence of three vowels, where the final vowel is either short or long:

(71) Omitted Constraints: Ident(len)-BR

Red-/tuai/, 'to be late'	Red=μμ	*Clash	*Hiatus	Align-to-σ́
☞ a. (ˌtu.a).tu.(ˈa.i)			***	*
b. tu.(ˌa.i).(ˈa.i)			****!	
c. tu.(ˌa.i).tu(ˈa.i)	*!		****	*

(72)

Red-/louaː/, 'to be rough'	Id(len)-BR	Red=μμ	*Clash	*Hiatus	Align-σ́
☞ a. (ˌlo.u).(ˌlo.u).(ˈaː)				***	*
b. (ˌlo.u).(ˌaː).(ˈaː)			*!	***	
c. lo.u.(ˌaː).u.(ˈaː)		*!		****	*
d. lo.(ˌu.a).u.(ˈaː)	*!			****	*

4.2.1 Exceptions to the Ranking

The final constraint ranking in (68) does not work for every stem though. Three mora stems with a heavy penult and a light final syllable consistently give the wrong output under this ranking:

(73)

Red-/saːlo/, 'to wipe'	Id(len)-BR	Red=μμ	*Clash	*Hiatus	Align-σ́
a. **(ˌsaː).lo.(ˈsaː).lo**		*!			
☞ b. (ˌsaː).(ˈsaː).lo			*		
c. (ˌsa.lo).(ˈsaː).lo	*!				

(74)

Red-/oːi/, 'to groan'	Id(len)-BR	Red=μμ	*Clash	*Hiatus	Align-σ́
a. **(ˌoː).i.(ˈoː).i**		*!		***	
☞ b. (ˌoː).(ˈoː).i			*	**	
c. (ˌo.i).(ˈoː).i	*!			***	

As a reminder, the ☞symbol indicates which candidates wins under the final ranking, but the bolded candidate is the actual output. All stems of this prosodic shape fail similarly under the final ranking.

My proposed explanation is that the words listed in Milner (1966, 1993) with heavy penults actually have light penults. As noted in Section 2.3.1, modern Samoan does not have heavy penults with final light syllables (Zuraw et al., 2014). If this proposal is correct, than the actual stems in (73) and (74) are '*salo*' and '*oi*.' Their tableaux are below:

(75)

RED-/salo/	ID(LEN)-BR	RED=$\mu\mu$	*CLASH	*HIATUS	ALIGN-$\acute{\sigma}$
☞ a. (ˌsa.lo).(ˈsa.lo)					
b. sa.(ˈsa.lo)		*!			

(76)

RED-/oi/	ID(LEN)-BR	RED=$\mu\mu$	*CLASH	*HIATUS	ALIGN-$\acute{\sigma}$
☞ a. (ˌo.i).(ˈo.i)				***	
b. o.(ˈo.i)		*!		**	

There are a few other stems for which the proper output does not arise under the final ranking, but none do so systematically; these other stems are just lexical exceptions. We can see this is the case by comparing two prosodically identical stems, one of which the ranking gives the wrong output for, and another it gives the right output for:

(77) Omitted Constraints: IDENT(LEN)-BR

RED-/folau/, 'to travel'	RED=$\mu\mu$	*CLASH	*HIATUS	ALIGN-TO-$\acute{\sigma}$
a. **fo.(ˌla.u).(ˈla.u)**			**!	
☞ b. (ˌfo.la).fo.(ˈla.u)			*	*
c. fo.(ˌla.u).fo.(ˈla.u)	*!		**	*

(78) Omitted Constraints: IDENT(LEN)-BR

RED-/maʔai/, 'to be sharp'	RED=$\mu\mu$	*CLASH	*HIATUS	ALIGN-TO-$\acute{\sigma}$
☞ a. (ˌma.ʔa).ma.(ˈʔa.i)			*	*
b. ma.(ˌʔa.i).(ˈʔa.i)			**!	
c. ma.(ˌʔa.i).ma.(ˈʔa.i)	*!		**	*

5 Reduplication With Lengthening

For both CV and bimoraic reduplication, there are certain words for which one of the output's vowels is lengthened. There are some differences between how lengthening works in the two types of reduplication, but by and large they are very similar.

5.1 CV Reduplication with Lengthening

When CV reduplication is accompanied by lengthening, the lengthened vowel is always the first vowel of the output. Two examples are given in (79) and (80):

(79) ˈsu.lu, ⟨ˌsuː⟩.ˈsu.lu
flee, ⟨PL⟩flee
'to flee', 'they flee'

(80) ma.ˈti.va, ˌmaː.⟨ti⟩.ˈti.va
poor, ⟨PL⟩poor
'to be poor', 'they are poor'

Note that the lengthened syllable is not necessarily the reduplicant: in (79), the reduplicant is lengthened, but in (80), the lengthened syllable is just the first syllable of the word.

Initial lengthening does not have any special semantics, except that it only occurs with plural forms[18], shown by the (I suspect cognate) pair in (81) and (82):

[18]But not *all* plural forms, e.g. (30).

(81) ˈte.ʔi, ⟨teː⟩.ˈte.ʔi
start, ⟨PL⟩start
'to start, to awake suddenly', 'they start, they awake suddenly'

(82) ˈte.ʔi, ⟨te⟩.ˈte.ʔi
rise, ⟨ERG⟩rise
'to rise (of tides)', 'to put up (an umbrella, mosquito net, etc.)'

Mosel & Hovdhaugen (1992) say that which words experience lengthening with reduplication is purely lexically determined; all my evidence supports that claim. The best argument in favor of that is that there is no prosodic basis for which words experience lengthening—that is, a word's syllable structure does not predict whether the reduplicated form will have initial lengthening. Consider the following near-minimal pair of CVCVCV words:

(83) ta.ˈʔa.lo, ˌtaː.⟨ʔa⟩.ˈʔa.lo
play, ⟨PL⟩play
'to play', 'they play'

(84) to.ˈʔu.lu, ˌto.⟨ʔu⟩.ˈʔu.lu
fall, ⟨PL⟩fall
'to fall, drop', 'they fall, drop'

Despite their identical prosodic structure and similar phonology, only one of these words, '*taʔalo*,' has lengthening in the reduplicated form.

There are some potential motivations for initial lengthening. First, it is possible that in some cases, initial lengthening is used to distinguish otherwise homophonous plurals. Consider the pair given in (85) and (86):

(85) ˈte.ʔa, ˌ⟨teː⟩.ˈte.ʔa
pass, ⟨PL⟩pass
'to pass (by)', 'they pass (by)'

(86) ˈte.ʔa, ⟨te⟩.ˈte.ʔa
throw, ⟨PL⟩throw
'to throw', 'they throw'

Without lengthening, these two plural forms would be indistinguishable. This cannot be the only explanation for lengthening, as there are not many pairs for which homophony avoidance could be a factor motivating lengthening. Additionally, there is at least one case where two semantically distinct (though admittedly perhaps cognate, according to Milner 1966) homophonous stems both experience lengthening:

(87) ˈsu.lu, ˌ⟨suː⟩.ˈsu.lu
flee, ⟨PL⟩flee
'to flee', 'they flee'

(88) ˈsu.lu, ˌ⟨suː⟩.ˈsu.lu
fall.headlong, ⟨PL⟩fall.headlong
'to fall headlong', 'they fall headlong'

Another challenge to the homophony avoidance proposal is that Samoan has ways besides CV reduplication to mark the plural form of a verb, such as the prefix *ta-*:

(89) ˈi.li, ta-ˈi.li
blow, PL-blow
'to blow (a horn)', 'they blow (a horn)'

(90) ˈlo.u, ta-ˈlo.u
fetch.down, PL-fetch.down
'to fetch down (oranges, etc.)', 'they fetch down (oranges, etc.)'

Note that *ta-* is used to pluralize words with onsetless penults like (89) because those words cannot undergo CV reduplication (see Section 3.3). However, it can also be used to pluralize words that in principle could have CV reduplicated forms, such as (90). This indicates that vowel lengthening is not the only employable strategy to avoid homophony in plurals.

Moreover, there are some words for which reduplicated forms with and without lengthening are both possible outputs:

(91) RED + /otaː/, ‘to be ripe’
ˌo.⟨ta⟩.ˈtaː, ˌoː.⟨ta⟩.ˈtaː
⟨PL⟩ripe, ⟨PL⟩ripe
‘they are ripe’, ‘they are ripe’

A second possibility for why lengthening might be occurring is to serve as an additional phonological cue for reduplication. The vowels in CV reduplicants are very short, particularly in fast speech, and are often dropped entirely (Mosel & Hovdhaugen, 1992, 34). However, since Samoan does not have geminate consonants, the derived geminate is shortened to a regular consonant, and so the reduplication is not audible:

(92) RED + *nofo*, ‘to live’ → *nonofo*, ‘they live’ → *n:ofo* → *nofo*

Lengthening the initial vowel could potentially prevent the reduplicated vowel from being dropped. This explanation works for stems with more than two moras as well, though instead of helping retain the reduplicated vowel, it helps distinguish the singular from the otherwise identical plural:

(93) RED + *mativa*, ‘to be poor’ → *ma:titiva*, ‘they are poor’ → *ma:t:iva* → *ma:tiva*

Though this seems like a reasonable explanation for why lengthening could occur, it does not explain why the majority of words do not experience lengthening.

5.2 Bimoraic Reduplication with Lengthening

As with CV reduplication, some Samoan words have one of their vowels lengthened when they undergo bimoraic reduplication:

(94) va.ˈʔi.li, ˌvaː.⟨ˌʔi.li⟩.ˈʔi.li
search.for, ⟨FREQ⟩search.for
‘search for’, ‘extricate, question closely’

Like initial lengthening in CV reduplication, this has no special semantics. A key difference, though, is that it is not confined to only a certain meaning, as it is in CV reduplication. Another difference is that some words have lengthening on the final vowel, rather than the first, as in (95):

(95) ˈpa.la, ⟨ˌpa.la⟩.pa.ˈlaː
rotten, ⟨FREQ⟩rotten
'be rotten', 'be muddy'

Again, which words have lengthening with bimoraic reduplication is lexically determined. Like CV reduplication, prosodic structure is not a factor:

(96) ˈfo.no, ⟨ˌfoː.no⟩.ˈfo.no
mend, ⟨FREQ⟩mend
'to mend', 'to mend frequently'

(97) ˈmo.no, ⟨ˌmo.no⟩.ˈmo.no
caulk, ⟨FREQ⟩caulk
'to plug a small opening, to caulk', 'to slip (into the hand)'

There are fewer good potential motivations for lengthening with bimoraic reduplication than with CV reduplication. It does seem to be able to distinguish otherwise homophonous forms:

(98) ˈta.la, ⟨ˌtaː.la⟩.ˈta.la
tell, ⟨FREQ⟩tell
'to tell', 'to talk, chat'

(99) ˈta.la, ⟨ˌta.la⟩.ˈta.la
unfold, ⟨FREQ⟩unfold
'to unfold', 'to pay out (a net), to space out'

However, the proposal that lengthening is an additional phonological cue for reduplication does not work for bimoraic reduplication since the syllables copied by bimoraic reduplication are not shortened like they are in CV reduplication.

6 Conclusion

In summary, Samoan has two different types of reduplication. They are semantically as well as phonologically distinct, requiring different constraint rankings to generate the outputs we see. In the course of this analysis, I have provided evidence that the unit of timing in Samoan is the mora, and not the syllable. I also showed that whether lengthening occurs with reduplication is lexically determined, and not conditioned by prosodic structure, phonological content, or semantics.

References

Alderete, John, Jill Beckman, Laura Benua, Amalia Gnanadesikan, John McCarthy & Suzanne Urbanczyk. 1999. Reduplication with fixed segmentism. *Linguistic Inquiry* 30(3). 327–364.

Cohn, Abigail C. 1989. Stress in indonesian and bracketing paradoxes. *Natural Language & Linguistic Theory* 7(2). 167–216.

Downing, Laura J. & Sharon Inkelas. 2015. What is reduplication? Typology and analysis part 2/2: The analysis of reduplication. *Language and Linguistics Compass* 9(12). 516–528.

Ghomeshi, Jila, Ray Jackendoff, Nicole Rosen & Kevin Russell. 2004. Contrastive focus reduplication in English (the salad-salad paper). *Natural Language & Linguistic Theory* 22(2). 307–357.

Gordon, Jr. (ed.), Raymond G. 2005. *Ethnologue: Languages of the world, fifteenth edition*. Dallas: SIL International. http://www.ethnologue.com/15.

Inkelas, Sharon & Laura J Downing. 2015. What is reduplication? Typology and analysis part 1/2: The typology of reduplication. *Language and Linguistics Compass* 9(12). 502–515.

Lunden, Anya. 2004. Reduplicant placement, anchoring, and locality. *University of California, Santa Cruz.* .

Lynch, John, Malcolm Ross & Terry Crowley. 2002. *The Oceanic languages*, vol. 1. Psychology Press.

Marantz, Alec. 1982. Re reduplication. *Linguistic Inquiry* 13(3). 435–482.

McCarthy, John J & Alan Prince. 1993. Prosodic morphology I: Constraint interaction and satisfaction. *Linguistics Department Faculty Publication Series* 14.

McCarthy, John J. & Alan Prince. 1995. Faithfulness and reduplicative identity. *University of Massachusetts Occasional Papers in Linguistics* 249–384.

McCarthy, John J & Matthew Wolf. 2007. Less than zero: Correspondence and the null output. In Curt Rice (ed.), *Modeling ungrammaticality in optimality theory*, Equinox Publishing.

Milner, G. B. 1966. *Samoan dictionary*. Pasifika Press.

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

Mosel, Ulrike & Even Hovdhaugen. 1992. *Samoan reference grammar*. Scandinavian University Press.

Nash, David George. 1980. *Topics in Warlpiri grammar.*: Massachusetts Institute of Technology dissertation.

Nevins, Andrew & Bert Vaux. 2003. Metalinguistic, shmetalinguistic: the phonology of shmreduplication. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, vol. 39 1, 702–721. Chicago Linguistic Society.

Prince, Alan & Paul Smolensky. 1993/2004. *Optimality Theory: constraint interaction in generative grammar*. Malden, MA: Blackwell. Revision of 1993 technical report, Rutgers University Center for Cognitive Science. Available online as ROA-537, Rutgers Optimality Archive, http://roa.rutgers.edu/.

Prince, Alan S. 1983. Relating to the grid. *Linguistic Inquiry* 19–100.

Zuraw, Kie, Kristine M. Yu & Robyn Orfitelli. 2014. The word-level prosody of Samoan. *Phonology* 31(02). 271–327.