Aggressive Reduplication

Kie Zuraw

University of California, Los Angeles

Kie Zuraw

UCLA Department of Linguistics

3125 Campbell Hall

Box 951543

Los Angeles, CA 90095-1543

kzuraw@yahoo.com

617-251-9547

fax: 310-206-5743

Aggressive Reduplication

Abstract

I propose that there is a purely phonological drive to construe words as containing a reduplicated portion. Such construals can lead to enhancement or preservation of word-internal self-similarity.

The case of vowel raising in Tagalog loan stems is examined in detail. Raising can be blocked in order to preserve similarity between the stem penult and the stem ultima. The more similar the penult and ultima along various dimensions, the more likely a reduplicative construal, and thus the more likely resistance to raising.

Two accounts of phonologically driven reduplicative construals are considered: the activity of a constraint Redup in generation, and the effect of *Spec in lexical learning. The proposal is compared to others that promote correspondence between similar or identical segments within a word; I conclude that correspondence between strings is necessary.

Aggressive Reduplication*

Loans and rare words are often 'repaired' by speakers to obey native phonotactics, as in the frequent pronunciation of English diphthong as di**p**thong, which eliminates a cluster of two fricatives. To give some idea of how widespread the modified pronunciation is, the numbers of web hits, using Google (www.google.com) in December 2001, are given in (1) for spellings reflecting the standard and modified pronunciations.

(1)

	web hits		web hits
di**p**thong	2,360	diphthong	11,000

There are also 'repairs' whose phonotactic motivation is unclear, as shown in (2). Some of the non-standard spellings in (2) reflect widespread pronunciations; others may be sporadic errors.

(2)[1]

Non-standard	hits	Standard	hits[2]
sherb**er**t[3]	about 12,000	sherbet	62,900
pompo**m**[4]	15,500	pompon	17,700
orangouta**ng**	6,130	orangutan	55,600
orangouta**ng**	257	orangoutan	67
Ok**ee**fenokee	1,430	Okefenokee [ˌoʊkəfəˈnoʊki]	17,400
har**i**-k**a**ri	8,430	hara-kiri	11,100

smorgasbor**g**	1,740	smorgasbord	71,500
sa**n**crosanct	201	sacrosanct	39,500
pers**er**vere	8,040	persevere	172,000
Inuktitu**k**	751	Inuktitut	23,500
Abu Dhab**u** / Ab**i** Dhabi	126 / 67	Abu Dhabi	135,000
asteris**t /** as**k**erisk	57 / 110	asterisk	613,300

What the "repairs" above have in common is that already-similar syllables are made more similar. For example, rang and tan become rang and tang. One interpretation is that speakers are improving what they construe as imperfect reduplication. Reduplication—the copying of material from the stem—is normally the manifestation of a morpheme, and we might expect speakers to seek to interpret a word as reduplicated only if it bears the proper morphosyntactic or semantic features. I propose, however, that regardless of morphosyntactic cues, speakers tend to construe words as reduplicated if they possess sufficient internal phonological similarity. I call this tendency for reduplicative misconstruals Aggressive Reduplication, after Hammond's (1999) Aggressive Suffixation. Hammond proposes that English adjectives whose final syllable resembles a suffix are stressed as though they actually were suffixed, just as I am proposing that words containing adjacent, similar portions are treated as though they actually were reduplicated.

Section 1 describes a straightforward class of cases—pseudoreduplicated roots in Tagalog—which sporadically behave as though they were truly reduplicated. Section 2 presents a subtler case, also from Tagalog, in which

Aggressive Reduplication, rather than enhancing word-internal similarity outright, preserves it by probabilistically blocking a productive alternation. Lexical statistics are used to argue that the greater the word-internal similarity to begin with, the more likely that the similarity-disrupting alternation will be blocked. Section 3 considers whether Aggressive Reduplication should be attributed to a reduplication-favouring constraint in generation, or to a structure-disfavouring constraint in lexical learning. Section 4 compares Aggressive Reduplication to other proposals involving drives for word-internal similarity or identity, and section 5 summarizes and concludes.

1. Tagalog pseudoreduplicated roots

In addition to various productive reduplicative morphemes, Tagalog has a large number of pseudoreduplicated roots—that is, roots of which one portion (the pseudoreduplicant) is identical to another (the pseudobase), but whose pseudobase cannot stand alone, and which lack the morphosyntactic or semantic characteristics of a productively reduplicated Tagalog word.

Some examples, selected at random from a database of 464 pseudoreduplicated words, gathered from English (1986), are given in (3).[5] The pseudoreduplicated roots are generally of the form C_1V_2-$C_1V_2C_3$ (3a) or $C_1V_2C_3$-$C_1V_2C_3$ (3b), though some pseudoreduplicated words (not illustrated) also have a medial vowel (bus**áː**bos 'slave'), a pseudoprefix (**gi**puspós 'very low-spirited'), or, most frequently, a pseudoinfix (p**al**igpîg 'shaking off water').[6] The misidentities seen in vowel height, presence of glottal stop, and nasal place of articulation are discussed below.

(3)	a.	lulód	‘N: shin’
		tá:tal	‘N: wood chips, splinters, or shavings’
		sú:soʔ	‘N: snail’
		lî:liw	‘N: bird sp.’
		pú:pog	‘N: pecking hard; repeated kissing’
		lulón	‘N: swallowing’
		laláʔ	‘A: acute’
		hî:hip	‘N: blow; puff’
		tá:taŋ	‘N: daddy’
		nú:noʔ	‘N: ancestor’
	b.	ʔagʔág, ʔagág	‘A: sifted as grain or powder’
		ŋikŋîk	‘N: noise made by pigs waiting to be fed’
		mismîs	‘N: remnants of food left after a meal’
		wigwîg	‘N: sprinkling of water during ironing’
		bunbón, bumbón	‘N: dam for attracting fish; clear pond’
		dasdás	‘N: planing or sanding wood or cane’
		gajgáj	‘N: travelling around’ ‘A: reached by travelling around’
		patpát	‘N: stick; piece of split bamboo’
		sagság	‘A: split; blunt; sagging; at the peak of success’
		ŋasŋás	‘N: scandal; excessive garrulousness and gesticulating’

The pseudoreduplicated words are probably not accidentally so. There are far more pseudoreduplicated words than would be expected through random phoneme combination,[7] and the pseudoreduplicated roots are phonologically exceptional in two ways. First, two occurrences of the same consonant within a root are rare except in pseudoreduplicated words. Not just any root with two identical consonants is counted here as pseudoreduplicated, so this observation is not vacuous. Second, pseudoreduplicated roots can contain consonant clusters that are otherwise rare or nonexistent root-internally (du**td**ót 'poking').

Still, there are several reasons to say that (synchronically, at least) these roots are not productively reduplicated, only pseudoreduplicated. First, in Tagalog the minimal root is disyllabic—the only monosyllabic roots are clitics and loans—so if these roots were reduplicated, it would be from a too-small root (e.g. *bak). Pseudoreduplication might be a repair strategy for just such too-small roots, but there are multiple pseudoreduplicating patterns (CV-, CVC-, and the medial-vowel, pseudoprefixed, and pseudoinfixed forms), so a monosyllabic lexical entry would still have to specify exactly how it is to be repaired; reduplication could not be just a predictable operation on monosyllabic roots. Second, although Tagalog does have productive CV- reduplication, there is no productive CVC- reduplication, nor are the pseudoprefixes and pseudoinfixes mentioned above productive. Third, although many pseudoreduplicated roots have a mimetic or pluractional flavour, there are no fixed meanings associated with the pseudoreduplicating patterns. And fourth, if there were a pseudoreduplicative morpheme, its distribution would be very restricted, since it occurs almost exclusively with monosyllabic roots (i.e., there are very few roots like *[babagid][8]).

Whether or not the pseudoreduplicated roots are related in some way, the important characteristics of these words for the proposal here are only that (i) they display a high degree of internal self-similarity, and (ii) this self-similarity influences how they are treated by the phonology, as will now be shown.

McCarthy and Prince (1995) propose correspondence between segments of the base and segments of the reduplicant as the reason for identity effects in morphological reduplication. Using the terms of Wilbur (1973) and McCarthy and Prince (1995), <u>transparent</u> cases are those in which a rule or constraint applies in all and only the expected environments, even though a misidentity between base and reduplicant may result; in <u>overapplication</u> only one half of the root is in the expected environment for a rule, but the rule applies to both; and in <u>underapplication</u> only one half of the root is in the expected environment for a rule, but the rule applies to neither. Most of the time, pseudoreduplicated roots show no signs of active reduplicative correspondence. That is, phonological phenomena apply transparently, even if the result is non-identity between the two halves of the root. But over- and underapplication do occur sporadically. The tables in (4) illustrate five types of example, giving in parentheses the number of roots that do not exhibit identity effects and the number that do. For example, Tagalog nasal substitution famously overapplies in productive reduplication (see Carrier 1979), with a nasal appearing in the base even though it is not adjacent to the triggering prefix ([**p**ulá] 'red', /paN+pula+in/ → [pa-**m**ula-hîn] 'to inflame', /paN+RED$_{CV}$+pula/ → [pa-**m**u-**m**ulá] 'ruddiness', /maN+RED$_{\sigma\sigma}$+pula/ → [ma-**m**ulá-**m**ulá] 'to glow', and even /naN+RED$_{CV}$+RED$_{\sigma\sigma}$+pula/ → [na-**m**ù:-**m**ulá-**m**ulá] 'reddish'). In most of the pseudoreduplicated words nasal substitution applies transparently, but in some it overapplies, as though the words were productively reduplicated (4a).

(4)

a. nasal coalescence (...N+T... → N and T coalesce)

most pseudoredup. (42)	handful of pseudoredup. (3)
transparent	overapplies
kamkám, ma-**ŋ**am**k**ám ‘usurpation’, ‘to usurp’	budbód, ma-**m**ud**m**ód ‘sprinkling’, ‘to sprinkle’

b. intervocalic tapping ([ɾ] / V__V, [d] elsewhere)

some pseudoredup.(7)	some pseudoredup.
transparent	overapplies (2)
dî:**ɾ**i ‘loathing’	**ɾ**ú:**ɾ**ok ‘acme’
	underapplies (1)
	dé:**d**e ‘baby bottle’

c. vowel height ([o] / __ C_0#, [u] elsewhere)

most pseudoredup. (142)	some pseudoredup. (12)
transparent	identity[9]
d**u**bd**ó**b	g**o**ŋg**ó**ŋ

'feeding a fire'	'gruntfish'

d. nasal place assimilation

some pseudoredup. (6 + 8 vary)	some pseudoredup. (3 + 8 vary)
transparent	underapplies[10]
da**n**dáŋ 'toasting'	di**ŋ**díŋ 'wall'

e. post-consonantal glottal deletion (*Cʔ)

some pseudoredup. (1 + 8 vary)	some pseudoredup. (5 + 8 vary)
transparent	underapplies
ʔutót 'flatulence'	**ʔ**ig**ʔ**íg 'shaking'

McCarthy and Prince (1995) propose that overapplication occurs when base-reduplicant correspondence constraints and a markedness constraint outrank input-output correspondence constraints. Underapplication occurs when an additional markedness constraint rules out overapplication, as illustrated in (5). The tableau in (5) considers only candidates that are construed as reduplicated, to show that such a construal forces both glottal stops to be retained (because Dep-BR >> *Cʔ); overapplication of glottal deletion is impossible because of a language-wide prohibition on vowel-initial words. The tableau is somewhat misleading, however, in that pseudoreduplicated stems' behaviour is lexically determined: some undergo glottal deletion, some do not, and some vary (and similarly for the other four phenomena listed). Section 3.1.1 addresses the question of how such behaviour becomes lexically encoded.

(5)

	/ʔigʔig/	*[V	Dep-BR	*Cʔ	Max-IO
a	☞[ʔig]$_R$[ʔig]$_B$			*	
b	[ʔig]$_R$[ig]$_B$		*!		*
c	[ig]$_R$[ig]$_B$	*!			**

If base-reduplicant correspondence is the mechanism responsible for over- and underapplication, then I interpret the over- and underapplication that occurs in these highly self-similar words to reflect reduplicative structure. If these words lack a true reduplicative morpheme, we have evidence that words that appear—phonologically—to be reduplicated can sometimes be treated as reduplicated, even in the absence of appropriate morphosyntax.

Behaving as though reduplicated seems to be common among pseudoreduplicated words.[11] Warlpiri has many words that Nash (1980) calls ‘lexically reduplicated’ (pp. 118-129). These words either bear an opaque semantic relationship to their unreduplicated counterpart, or have no reduplicated counterpart at all, and many fall into a few semantic categories (such as bird names and circularity). Lexically reduplicated words in Warlpiri can escape certain morpheme structure conditions, as though they were productively reduplicated (for example, <u>kuurrkuurrpa</u> ‘boobook owl’ escapes the prohibition on long vowels in non-initial syllables). Furthermore, most of the lexically reduplicated words pattern like morphologically reduplicated words in terms of stress.

In Manam (Lichtenberk 1983, Buckley 1997), a productive reduplication pattern that normally copies the last two light syllables of a base (salága 'be long', salaga-lága 'long (sg.)') copies only one syllable if the final two syllables of the base are identical (ragógo 'be warm'; ragogó-go, *ragogo-gógo 'warm'). Buckley argues that the two [g]s of [ragógo] correspond to a single underlying segment (and likewise the two [o]s), violating of McCarthy & Prince's (1995) Integrity. The reduplicant is therefore kept as small as possible to minimize further violations of Integrity. This means that [ragógo] is being treated by the phonology as though it has reduplicative structure.

The Warlpiri and Manam cases differ from Tagalog in that the pseudoreduplicated words in those languages generally look like possible productively reduplicated words. In Tagalog, because of the requirement that roots be minimally disyllabic, pseudoreduplicated words do not like possible productively reduplicated words. This may explain why in Warlpiri and Manam most or all pseudoreduplicated words show phonological signs of reduplicative structure, but only a minority do in Tagalog.

Analysts have ascribed reduplicative structure to pseudoreduplicated words in other languages that do not qualify morphosyntactically or semantically for a reduplicative analysis. Nash attributes reduplicative structure to the Warlpiri lexically reduplicated words discussed above. Buckley proposes that Manam pseudoreduplicated words have an underlying RED morpheme, and Golston and

Thurgood (in press) have proposed the Direct-OT equivalent (a lexically required *Echo violation) for Chumash. Gafos (1998) has proposed that Semitic vocalisms that trigger doubling of the final consonant of a biliteral root contain a RED element, despite lacking the required semantics.

There are two reasons, however, why pseudoreduplicated words do not provide absolute evidence for Aggressive Reduplication. The first is that, despite the arguments above against the presence of a reduplicative morpheme, there is also an argument in its favour—the mimetic or pluractional feel mentioned above that many of the Tagalog words have. Second, even if the words are not now morphologically reduplicated, it is possible that the cases of over- and underapplication are holdovers from a proto-language in which they were.

The following section discusses a more widespread case of underapplication in Tagalog: vowel raising in loanwords, where self-similarity gradiently blocks a productive alternation. Here, the absence of a reduplicative morpheme is clear, as most of the words have only partial self-similarity, and their meanings are diverse. And, because the stems in question are loans, there is no possibility of a reduplicated history.

2. Vowel raising in Tagalog loanwords

The English and Tagalog data discussed so far provided sporadic examples of reduplicative misconstruals. This section presents a more systematic case, with enough instances for statistical information to be meaningful.

Height in non-low Tagalog vowels is partially predictable. In most of the native vocabulary, the mid vowels [o] and [e] are found only in ultimas, and [u] is found only in non-final syllables. [i] can occur anywhere, and many words have [i] and [e] in free variation in the ultima. (6) shows some typical monomorphemic native words.

(6)	b**úː**ko	'young coconut, flower bud'
	b**i**gát	'burden'
	bíːl**o**g	'circle'
	dáːl**e**, dáːl**i**	'denunciation'
	búːk**i**d	'farm'

Suffixation induces alternation, by making final syllables non-final:[12]

(7)	káːl**o**s	'grain leveller'	kal**úː**s-in	'to use a grain leveller on'
	ʔab**ó**	'ash'	ʔab**u**-hîn	'to clean with ashes'
	babáːʔ**e**	'woman'	ka-babaʔ**iː**-han	'womanhood'
	sist**é**	'joke'	sist**i**-hîn	'to joke' (Spanish loan)

There are two classes of systematic exceptions to the generalization that mid vowels are found only in ultimas; they are not the focus of this paper, however, so they will be only briefly described. First, in non-final syllables containing an [aw] or [aj] diphthong (which may in turn derive from [aʔu] or [aʔi]), coalescence can occur, producing a long, stressed mid (or sometimes high, not illustrated) vowel of the same backness and rounding as the glide, as in [ʔajwán] ~ [ʔéːwan] 'I don't know' and [kaʔuntîʔ] ~ [kawntîʔ] ~ [kóːntiʔ] 'a little'.[13] The second systematic source of non-ultima mid vowels is V<u>ʔ</u>V sequences in which both vowels are non-low. In these sequences, the vowels must match in backness. If the vowels are back, the second is mid and the first may be high or mid ([suʔót] 'clothing', [poʔók] 'place'). If the vowels are front, either both vowels are high or both are mid ([biʔík] 'piglet', [meʔéʔ] 'bleat').

There are also seemingly unsystematic exceptions in the native vocabulary, though they are few (there are many exceptions in the loanword vocabulary, discussed below): words with non-ultima mid vowels (8a-d), words whose ultima vowels remain mid under suffixation (8e), and words with ultima [u] (8f), many of which could be loans. Many of the exceptions are baby-talk words, interjections, or onomatopoeic/mimetic words; as in other languages, some well-formedness requirements seem to be relaxed in the 'peripheral' vocabulary of Tagalog (see Itô & Mester 1995). The list in (8) is close to exhaustive: it includes all of the exceptions that were found in a database of the 4390 disyllabic, native roots in English's (1986) dictionary, as well as all the relevant longer native words that I have encountered. Note that many of the words with non-ultima mid vowels appear to have CV- or CVC- pseudoreduplication (8a,c), and that the words that fail to be raised under suffixation have a non-ultima mid vowel of the same backness as the ultima mid vowel (8e). These facts will be relevant below in explaining the distribution of exceptions.

(8) a. ʔ**óː**ʔo 'yes'

t**oto**ʔó 'true'

k**óː**kok 'crow of rooster; chickie'

	t**óː**toʔ ~ t**óː**toj	'(affectionate term of address for little boy)'
	g**o**ŋgóŋ	'gruntfish sp.'
	kat**óː**to	'comrade'
	bak**oː**ko	'fish sp.'
b.	b**o**hól	'shrub sp.'
	ʔ**óː**la	'eagerness'
	k**óː**kak	'croak of frog'
c.	d**éː**de	'baby bottle'
	m**éː**me	'beddie-bye'
	k**e**ŋkéŋ	'sound made by beating frying pan'
	n**éː**neʔ ~ n**éː**neŋ ~ nîːniʔ	'(affectionate term of address for little girl)'
	h**eː**l**e**h**éː**le	'pretence of not liking'
	h**éː**le	'lullaby'
d.	k**éː**rwe	'cricket'
	l**éː**teŋ	'cord'
	k**éː**toŋ	'leprosy'
		(raised when suffixed: [ketúːŋ-in] 'to have leprosy')
	t**éː**pok	'victimised by hooligans'
	h**éː**to ~ ʔ**éː**to	'Here it is!'
	b**éː**lat	'Serves you right!'
	l**éː**kat	'How could you?!'
	p**éː**klat	'scar'
	t**éː**kas	'swindler'
	s**éː**laŋ ~ s**éː**lan	'delicacy'

	kulá:lat ~ kul**é:**lat		‘last’	
e.	dé:d**e**	‘baby bottle’	paded**é:**-hin	‘give a baby a bottle’
	totoʔ**ó**	‘true’	totoʔ**ó:**-hin	‘to be sincere’
	poʔ**ó**t	‘hatred’	ka-poʔ**o**t-án	‘to hate’

(and all other oʔo words; found no eʔe words with suffixed derivatives)

f.	samp**ú**ʔ	‘ten’
	ʔi:m**u**s	‘cape, headland’ (from Eng. isthmus?)
	kasój~ kas**ú**j	‘cashew’ (from Eng. cashew?)
	bagkós ~ bagk**ú**s	‘on the contrary’
	bambó ~ bamb**ú** ~ banb**ú**	‘club, baton, stick’ (from Malay bambu?)
	dá:toʔ ~ dá:t**u**ʔ	‘chieftain’ (from Malay datu?)
	labîw ~ labj**ú**	‘weeds that grow in a burned field’

As in the native vocabulary, there are exceptions of each kind to vowel height phonotactics in loanwords from Spanish and English. Exceptions are much more numerous among the loanwords, however, because mid and high vowels are freely distributed in the source languages:

(9)	b**é:**nta	‘sales’	(from Spanish venta)
	k**o**ɾék	‘correct’	(from English correct)
	ʔas**ú**l	‘blue’	(from Spanish azul)
	ʔà:bak**ú**s	‘abacus’	(from English abacus)

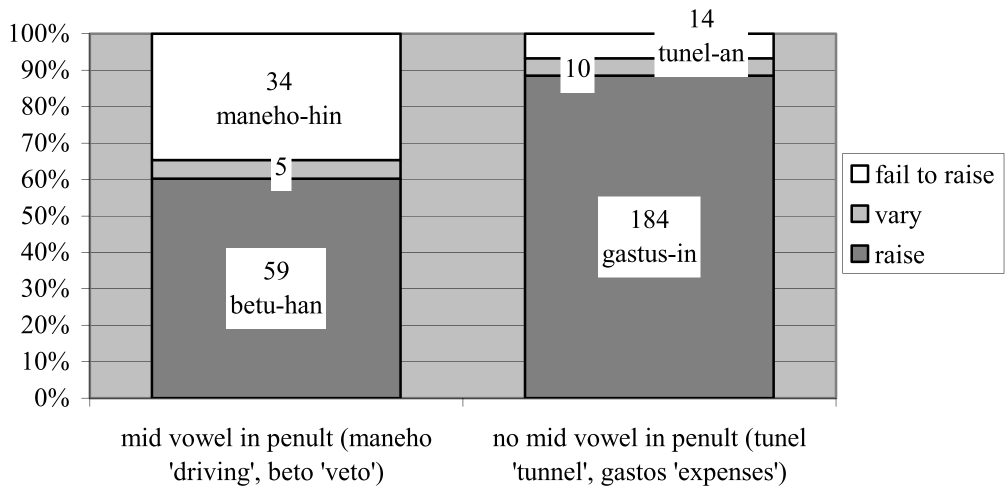
Some loanword stems with mid-vowel ultimas alternate under suffixation (10a), and some fail to alternate (10b):

(10)	a.	sab**ó**n	'soap'	sab**u**n-án	'to put soap on'
		atá:k**e**	'attack'	atak**î**:-hin	'to attack (object focus)'
		gó:lp**e**	'hit'	g**u**lp**i**-hîn	'to hit (OF)'[14]
	b.	ká:bl**e**	'cable'	kabl**e**-hán	'to send a cable to'
		mag-mané:h**o**	'to drive (AF)'	maneh**ó**:-hin	'to drive (OF)'

Because vowel height within a bare stem is usually borrowed faithfully from Spanish or English, it is of little interest—that is, a non-ultima mid vowel is present just because it was present in the Spanish or English word. What is of interest is whether a loan-stem's ultima vowel alternates when given a native suffix, because that can be determined only by the Tagalog phonology. I constructed a database from English's (1986) dictionary of all 488 Spanish and English loans with a mid vowel in the ultima and one or more listed suffixed derivatives.

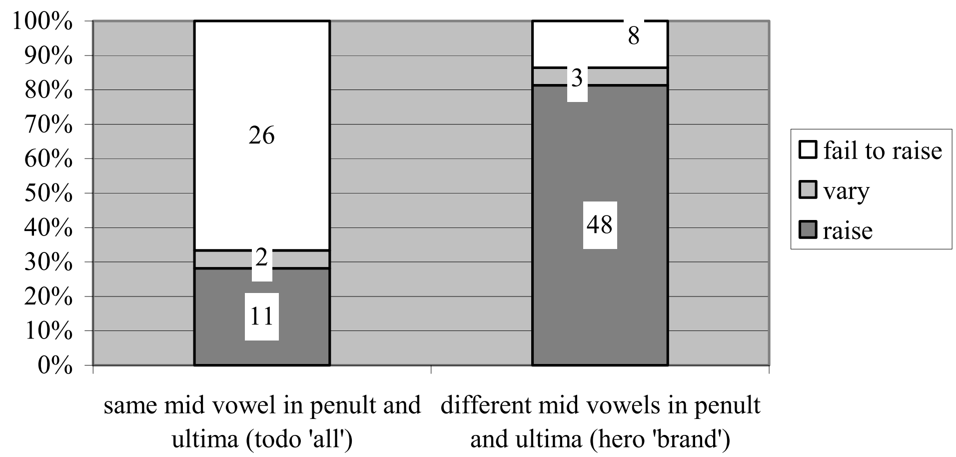
As observed by Schachter and Otanes (1972), the best predictor that a loanword stem will fail to alternate is the presence of a mid vowel in another syllable. As shown in (11), only 7% of stems with a low- or high-vowel penult fail to be raised (like t**u**n**e**l-an 'to tunnel'),[15] but 35% of those with a mid-vowel penult fail to be raised (like man**e**h**o**-hin 'to drive'). The effect is significant, with $p<.0001$, using Fisher's Exact Test (variable stems were omitted for all applications of Fisher's Exact Test).

(11) Effect of mid vowel in penult on probability of raising



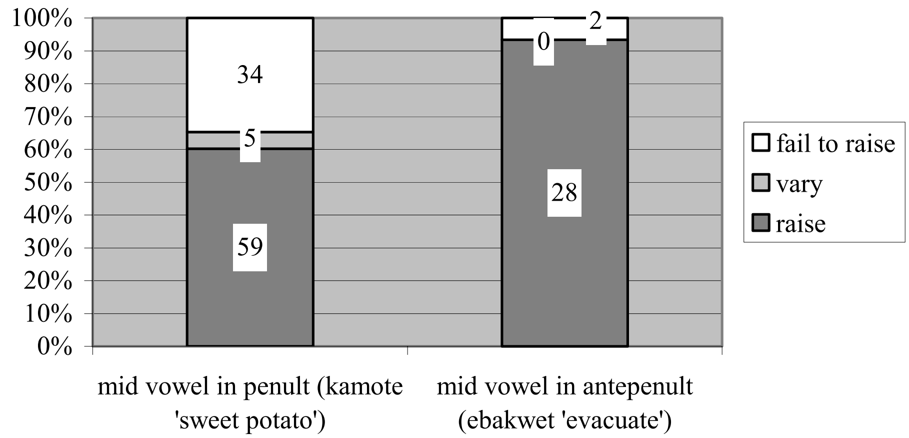
The likelihood of non-raising is enhanced if the two mid vowels match in backness (p<.0001):

(12) Effect of matching backness between penult and ultima, given a mid penult.



Proximity also has an effect. A mid vowel in the penult strongly encourages non-raising, as compared to a mid vowel in the antepenult ($\underline{p}<.005$), which in turn has no effect when compared to no mid vowel at all:

(13) Effect of proximity



There are several possible explanations for why the presence of another mid vowel discourages raising. I will present three; although two of them can account for the matching-backness and proximity effects, I will argue that an Aggressive Reduplication analysis is superior to all three, because only Aggressive Reduplication can explain an additional effect of similarity beyond the vowels themselves.

The first explanation that will ultimately be rejected is that the presence of a non-ultima mid vowel, which is unpredictably [-high], marks the whole word as contrastive for [high]. The ultima vowel would thus also be interpreted as contrastively (rather than predictably) [-high], and so remain [-high] under suffixation. This explanation could account for the backness effect if we assume that a non-ultima [e] tends to mark the whole word as contrastive for [high] in [-back] vowels only, and a non-ultima [o] tends to mark the whole word as contrastive for [high] in [+back] vowels only. To account for the proximity effect, we could assume that a non-ultima mid vowel tends to mark only adjacent syllables as contrastive for [high]. This explanation seems to work, and will be rejected because of additional effects to be discussed below that it does not capture.

A second possible explanation to be rejected is that the presence of the non-ultima mid vowel marks the whole word as belonging to a foreign stratum, subject to different constraints or to a different constraint ranking (see Itô and Mester 1995) that do not require raising under suffixation. If this is the explanation, we expect that other markers of foreignness could be found that would also discourage alternation. I examined several such predictors. Stress or length on a non-final closed syllable and prepenultimate stress or length are rare or nonexistent in the native vocabulary, so 'foreign' stress might be expected to predict non-alternation. There is a small difference in the predicted direction, with 15% of loan-stems with foreign stress failing to be raised and 12% of other loan-stems failing to be raised, but it is not significant ($p>.5$). Foreign distribution of [d] and [ɾ] (in the native vocabulary, [ɾ] is normally found intervocalically and [d] elsewhere) also has no effect on the likelihood of alternation. The difference is not in the predicted direction (10% vs. 15%), and not significant ($p>.25$). Finally, initial clusters of two or more consonants and medial clusters of three or more consonants, not found in the native vocabulary, have no effect (14% of loan-stems with such clusters fail to be raised, and 14% of loan-stems without such clusters fail to be raised). Thus, the idea that a non-ultima mid vowel serves as a general cue to foreignness does not seem to be a good explanation for why the presence of such a vowel discourages alternation, since other cues to foreignness do not discourage alternation. In addition, foreignness marking cannot explain the matching-backness effect, though it could explain the proximity effect if we assume that a contrastively [-high] vowel marks only adjacent syllables as foreign.

A third possible mechanism by which the non-ultima mid vowel could discourage alternation is vowel harmony. If a [-low] vowel must agree in height with a preceding vowel, then raising of the o in maneho under suffixation would be prevented. Vowel harmony could explain the backness effect if agreement between target and trigger for one feature encourages harmony of another feature (see Kaun 1995 for evidence that agreement in height encourages rounding harmony, although the phonetic explanation given for that phenomenon would not apply to an interaction between backness and rounding). The proximity effect could be explained if harmony requires adjacency. We might additionally expect that a mid vowel in the antepenult could block raising only if the intervening vowel is [-high], but because all but two stems with antepenult mid vowels (and no penult mid vowel) fail to be raised, we cannot test this prediction.

Although the contrastiveness-marking and vowel-harmony account are workable so far (foreignness-marking is not), I reject all three possibilities in favour of Aggressive Reduplication, because of evidence given below. I claim that a preceding mid vowel impedes raising of the stem-final vowel because a reduplicative correspondence relation has been imposed between the two syllables. Raising the second vowel would therefore violate Ident-BR(high). If the constraint against nonultima mid vowels, *NonFinalMid, and Ident-BR(high) are variably ranked, then raising will sometimes be blocked if the word is construed as reduplicated.

The tableau in (14) shows the analysis of vowel height in general (preservation of exceptional nonfinal mid vowels, and raising of final vowels). Ident-IO(hi), which could easily be replaced with output-output correspondence to the unsuffixed form, has been separated into final-syllable and nonfinal-syllable versions. (15) illustrates how raising is blocked if there is a reduplicative construal. As in (5), only reduplicatively construed candidates are shown; section 3 presents a proposal for why such candidates would be preferred.

(14)

/heto/	Ident-IO(hi) nonfinalσ	*NonUltimaMid	Ident-IO(hi) stem-finalσ
a ☞ heto		*	
b hito	*!		
/kalos+in/			
c ☞ kalusin			*
d kalosin		*!	

(15)

/CoCo + -an/	Ident-IO(hi)	Ident-BR(hi)	*NonUltimaMid	Ident-IO(hi)

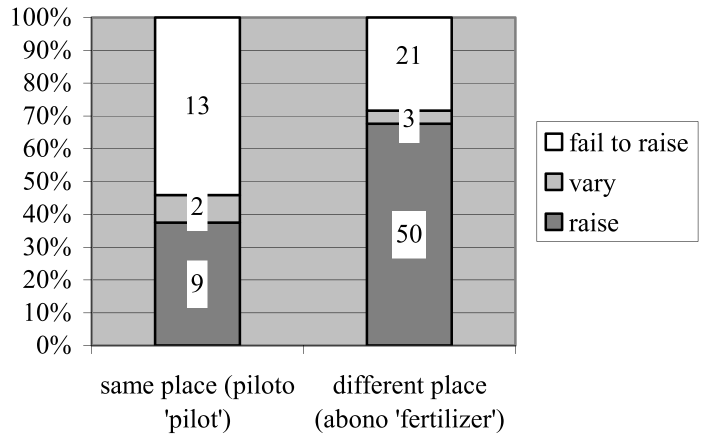
	nonfinalσ			stem-finalσ
a ☞ $[Co]_\alpha[Co]_\alpha han$			**	
b $[Co]_\alpha[Cu]_\alpha han$		*!	*	*
c $[Cu]_\alpha[Cu]_\alpha han$[16]	*!			

Aggressive Reduplication can also explain the matching-backness and proximity effects. If two vowels have the same backness, a reduplicative construal incurs no violation of Ident-BR(back). And whatever constraints prefer productive reduplicants to be adjacent to their bases (perhaps alignment, though this issue is unsettled) would also disprefer reduplicative correspondence between the ultima and a pre-penultimate syllable, explaining the proximity effect.

But Aggressive Reduplication makes an additional prediction. In words that have only partial internal similarity, whatever force encourages a reduplicative construal (see section 3 for two possibilities) is in competition with base-reduplicant correspondence constraints (abbreviated as 'Corr-BR' below). A reduplicative construal should therefore be easier to impose between the penult and the ultima when they are more similar, because fewer Corr-BR constraints are violated. Assuming variable constraint ranking (see section 3.1.1), a ranking that allows a reduplicative construal is more likely to occur if fewer Corr-BR constraints' rankings are crucial. Therefore, greater similarity in various aspects of the penult and ultima of loanstems, not just in the vowels themselves, should be correlated with a higher rate of non-raising, if non-raising reflects a reduplicative construal. The other three accounts do not make this prediction: a mid vowel in the penult's likelihood of marking an ultima mid vowel as contrastively [-high] should not be affected by the similarity between the two syllables. Nor should foreignness marking be affected by syllabic similarity. And vowel harmony should not be affected by similarity between two syllables outside of the vowels themselves. The remainder of this section argues that the prediction is correct; therefore, the Aggressive Reduplication account is preferred over the other three.

Aggressive Reduplication predicts that similarity between penult and ultima along any dimension—not just vowel backness—should encourage non-raising by making it easier to impose a reduplicative construal. For example, matching place of articulation in syllable onsets should encourage non-raising. This works the same way as the matching-backness effect: correspondence between [lo] and [to] does not violate Ident-BR(coronal), but correspondence between [bo] and [no] does. (16) shows that when the penult and ultima onsets have the same place of articulation (in complex onsets, only the first consonant was used), non-raising is more likely ($p<.05$). Note that (16), like the other charts in this section, compares stems whose penult and ultima are similar along some dimension (here, onset place) to stems whose penult and ultima are dissimilar along that dimension. The penult and ultima onsets of the words grouped with piloto must be identical in major place, but may be different or similar in voicing or manner; the penult and ultima onsets of the words grouped with abono must differ in major place, but may be different or similar along other dimensions.

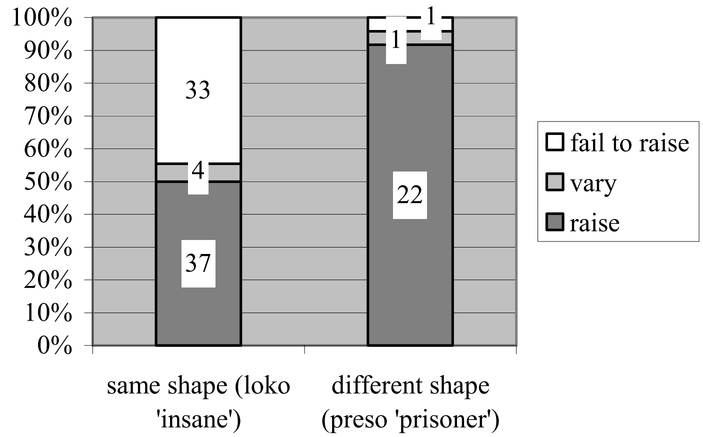
(16) Effect of onset place of articulation on rate of raising



As illustrated below, two other types of similarity—onset shape and rime shape—also encourage non-raising significantly. Two additional dimensions were considered: the effect is in the predicted direction for onset manner, but not significant; for onset voicing, the (non-significant) effect is not in the predicted direction.

When onsets match in shape (simple vs. complex), non-raising is encouraged ($\underline{p}<.0001$):[17]

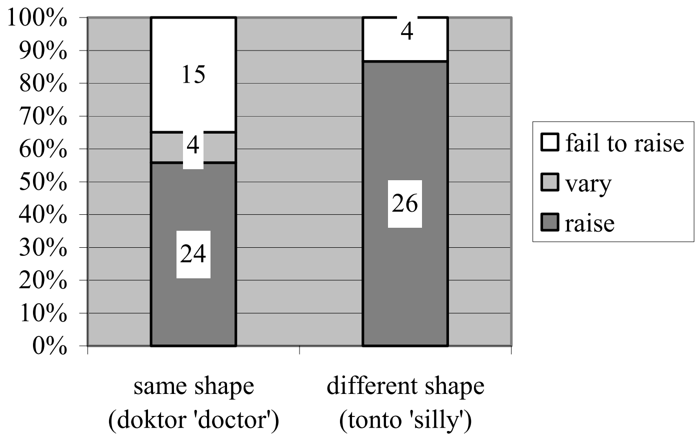
(17) Effect of onset shape on rate of raising



Here the crucial constraint is Max-BR or Dep-BR (leaving aside for the moment which syllable would count as the base and which as the reduplicant): in [pɾéːso] 'prisoner', the [ɾ] of [pɾeː] (or the [p]) lacks a correspondent in [so].

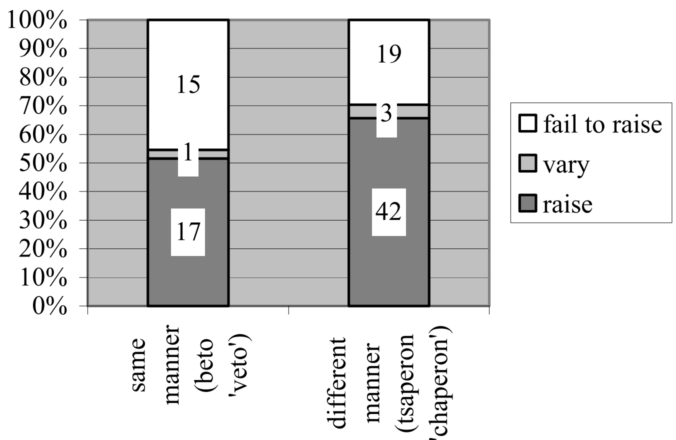
There are not enough cases in which both penult and ultima are closed to compare coda consonants themselves, but we can compare rime shape (open vs. closed), and again a match promotes non-raising (p<.05), because correspondence is possible without violating Max-BR or Dep-BR. As discussed in more detail in section 3.3, suffixation opens a final closed syllable, which alters the similarity properties of the final two syllables' rimes. Therefore, the data in (18) are restricted to words with an open ultima, so that the characterization of a stem's penult and ultima as 'same' or 'different' with respect to rime shape does not change under suffixation.

(18) Effect of rime shape on rate of raising



Identical onset manner[18] also appears to encourage non-raising (the effect is in that direction), though not significantly (p=.17).

(19) Effect of onset manner on rate of raising

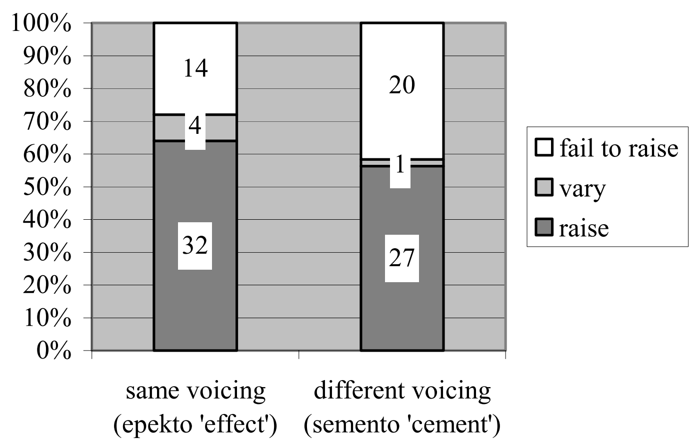


The [b] and [t] of [bé:to] ‘veto’ can correspond no matter what the ranking of Ident-BR(sonorant), Ident-BR(nasal), or any other Ident-BR(manner) constraints.

But the [p] and [ɾ] of [tʃaːpeɾón] ‘chaperon’ can correspond only if various Ident-BR(manner) constraints are ranked low.

Voicing is the one property that does not pattern in the predicted direction, though the difference observed is not significant (p=.28). It would have been desirable to examine voicing only where it is contrastive (in stops and foreign-derived affricates), but there are too few stems whose last two onsets are both obstruents.

(20) Effect of onset voicing on rate of raising



The table in (21) summarises the effects of each predictor of the Aggressive Reduplication analysis that was examined, and, for comparison, each predictor of the rejected tagging-as-foreign analysis.

(21)

Predictor	% non-raising, stems with predictor	% non-raising, stems without predictor	p	
Predictors of non-raising in Aggressive Reduplication analysis				
matching backness	65%	11%	<.0001	
matching onset place	57%	33%	<.05	si.
matching onset shape	40%	12%	<.0001	
matching rime shape	40%	22%	<.05	
matching onset manner	50%	36%	=.17	
matching onset voicing	40%	42%	=.28	
Predictors of non-raising in foreign-marking analysis (rejected)				
foreign stress	15%	12%	>.5	
foreign d/ɾ distribution	10%	15%	>.25	
large consonant clusters	14%	14%	>.5	

The Aggressive Reduplication analysis can be summarised thus: if there is a mid vowel in the penult, a reduplicative construal tends to discourage raising because raising would violate Ident-BR(high). Such a construal is more likely when the final syllable is similar to the penult, because few base-reduplicant correspondence constraints are violated. If the ranking of these constraints with respect to the driving force behind reduplicative construals is variable, then a ranking that allows a reduplicative construal is more likely when few Corr-BR constraints would be violated, and thus fewer Corr-BR constraints need to be ranked low.

The three other accounts discussed, contrastiveness-marking, foreignness-marking, and vowel harmony, do not predict that similarity in anything other than the vowels themselves should promote nonraising. Therefore, I conclude that the Aggressive Reduplication analysis is superior to these three.

3. Learning or production?

The above sections have argued that self-similarity within a word can cause it to be construed as reduplicated, leading to similarity enhancement (as in the sporadic cases from English and from Tagalog pseudoreduplicated roots), or to similarity preservation (as in the avoidance of vowel raising under suffixation in Tagalog). They have also argued that such a construal is more likely the greater the self-similarity, because fewer base-reduplicant correspondence constraints are violated. But what countervailing force promotes reduplicative construals in the first place? At least two mechanisms suggest themselves: generation (i.e., mapping lexical to surface forms) and lexical learning (the reverse). This section shows how each would work, then what kind of evidence would distinguish between the two possibilities. In the Tagalog vowel raising case, the evidence is inconclusive, but some sporadic examples from Palauan suggest that at least some reduplicative construals result from generation alone.

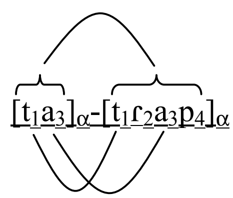
3.1. Generation analysis

One possible source of reduplicative construals is generation—that is, there is a constraint in the grammar that prefers output forms to have a reduplicated structure. I call this constraint Redup; it requires all words to have internal, base-reduplicant-type correspondence (cf. Hammond's Suffix, which promotes construing words as suffixed).

There are many ways to formulate Redup that are consistent with the data discussed here. The key is that it require internal correspondence. This proposal depends on the nature of correspondence as proposed by McCarthy and Prince (1995): correspondence is an arbitrary relation between phonological structures that does not itself require similarity. Violable constraints require the relation to have certain properties, and enforce similarity between items that are in correspondence. Thus Redup demands only that there be correspondence within a word, not that the correspondents be similar.

It may be helpful here to distinguish between two types of correspondence. First is the familiar correspondence between segments. For example, in productively reduplicated $\text{mag-}b_1i_2\text{-}b_1i_2l_3i_4$, the two [b]s are in correspondence, as indicated by the matching numerical subscripts. Second is the relation between the reduplicant and the base as a whole, which we could notate with subscript R and B: $\text{mag-}[b_1i_2]_R\text{-}[b_1i_2l_3i_4]_B$. It is this higher-level relation that drives the segmental-level correspondence—for example the constraint Max-BR (McCarthy & Prince 1995), which requires every segment of the base to have a correspondent in the reduplicant, is relevant only if some part of the word is labelled as the base. We could say that segments are the domain and range of the correspondence relation C, indicated by Arabic-numeral subscripts, and strings of segments are the domain and range of the correspondence relation C', indicated by Greek-letter subscripts. So in a hypothetical reduplicated word $[t_1a_3]_\alpha$-$[t_1ɾ_2a_3p_4]_\alpha$, the matching subscripts indicate that t_1Ct_1 and a_3Ca_3, but $[t_1a_3]_\alpha C'[t_1ɾ_2a_3p_4]_\alpha$.

(22) C' relation:



C relation:

Redup, then, requires string-level correspondence. We could define it as requiring string-level correspondence between any two syllables of a word, as in (23).

(23) Redup

If σ_1 and σ_2 are syllables belonging to the same word, then the string of segments belonging to σ_1 must be in (string-level) correspondence with the string of segments belonging to σ_2.

For example, $[b_1a_2]_\alpha[d_1a_2]_\alpha$ would not violate Redup so defined, because it has just two syllables, and they are in string-level correspondence (they share the subscript α). The segment-level correspondence relation is imperfect, violating Ident-BR(place), but this does not affect evaluation of Redup.

But this is only one possible formulation of Redup. Many of the English examples in (2) seem to involve correspondence between feet rather than syllables (e.g. [hari]$_\alpha$[kari]$_\alpha$), and productive reduplication, in Tagalog as in other languages, can involve foot copying. Productive reduplication can also place into correspondence strings that do not have the same prosodic shape, as in Ilokano [pjan.]$_\alpha$ - [pja.n]$_\alpha$o 'pianos': the [n] of the reduplicant is a coda, but the [n] of the base is an onset (variant pronunciations [pii.]$_\alpha$ - [pj]$_\alpha$a.no and [pi]$_\alpha$ - [p.j]$_\alpha$a.no also show prosodic mismatches; Hayes & Abad 1989). If Redup promotes the same range of correspondence structures that are found in productive reduplication, then it might merely require a word to contain some pair of corresponding substrings—other, violable constraints would be responsible for promoting adjacency between those substrings[19] and identity of prosodic roles between their segments (Steriade 1988 argues that identity of prosodic roles is promoted in Sanskrit reduplication). In Tagalog, prosodic identity is unviolated in productive reduplication, so constraints enforcing it should be ranked high, consistently with the apparent correspondence of syllable-shaped strings in the raising data.[20]

Although Redup itself does not require word-internal similarity (but merely string-level correspondence), the base-reduplicant correspondence constraints of McCarthy and Prince's theory do. Max-BR requires every segment of the base to have a correspondent in the reduplicant, and Dep-BR, which could equivalently be called Max-RB, requires every segment of the reduplicant to have a correspondent in the base. The Ident-BR(F) family of constraints require corresponding segments to have identical feature values.

McCarthy and Prince intend these constraints to apply to substrings morphologically labelled ‘base’ and ‘reduplicant’, but they could also be thought of as applying to any two substrings within a word that correspond. The only difficulty in that case becomes how to obtain the difference in direction made by Max-BR and Dep-BR. In cases of productive reduplication, the two constraints could be thought of as morphologically sensitive versions of a more general constraint Max-R that requires correspondence between the segments of any substrings within an output form that themselves correspond:

(24) Max-R: if S_1 and S_2 are distinct substrings of a single output form, and $S_1C'S_2$, then for every segment x of S_1, there exists some segment y of S_2 such that xCy.

Max-BR = Max-R with S_1 = base and S_2 = reduplicant.
Max-BR = Max-R with S_1 = reduplicant and S_2 = base.

Because there is no reduplicative morpheme in the vowel raising case, and thus no basis for distinguishing reduplicant and base, I will use Max-R, which penalizes both segments in the first string that lack a correspondent in the second, and segments in the second string that lack a correspondent in the first.

The Corr-BR constraints (Max-R and the Ident-BR(F) family) interact with Redup, because Redup requires substrings to correspond with each other, and Corr-BR requires those corresponding strings to be similar. That is, Redup, by putting substrings into correspondence, invokes Max-R in much the same way that Max and Dep, by putting segments into correspondence, invoke Ident-(F). Redup also interacts with the input-output correspondence constraints (abbreviated Corr-IO), including Ident-IO(F), Max-IO, and Dep-IO (corresponding segments of input and output must have identical values for feature [F]; every segment in the input must have a correspondent in the output; every segment in the output must have a correspondent in the input) by dispreferring candidates whose underlying material has been changed to achieve greater self-similarity. The Corr-BR and Corr-IO constraints interact with Redup to (i) restrict which syllables can be in correspondence and (ii) enhance the similarity of corresponding syllables. The schematic factorial typology in (25) illustrates the interaction (jagged lines between columns indicate that all rankings are to be considered). The three possible outcomes are string-level correspondence despite imperfect similarity ($[\text{tag}]_{\alpha}[\text{dag}]_{\alpha}$), enhancement of internal similarity ($[\text{tag}]_{\alpha}[\text{tag}]_{\alpha}$), and lack of string-level correspondence ($[\text{tag}]_{\alpha}[\text{dag}]_{\beta}$).

(25) Interaction of Redup with Corr-IO and Corr-BR

- Redup, Ident-IO(voice) >> Ident-BR(voice)
  reduplicated construal, despite imperfect similarity (a)
- Redup, Ident-BR(voice) >> Ident-IO(voice)
  voicing difference is ‘repaired’ (b)
- Ident-IO(voice), Ident-BR(voice) >> Redup
  non-reduplicated construal (c)

/tagdag/	Redup	Ident-IO(voice)	Ident-BR(voice)
a $[tag]_\alpha[dag]_\alpha$			*
b $[tag]_\alpha[tag]_\alpha$		*	
c $[tag]_\alpha[dag]_\beta$	*		

The interaction holds not just for Ident constraints but also for Max, which has the effect of requiring similarity of syllable shape:

(26)

/tagta/	Redup	Max-IO	Max-R
a $[tag]_\alpha[ta]_\alpha$			*
b $[ta]_\alpha[ta]_\alpha$		*	
c $[tag]_\alpha[ta]_\beta$	*		

Because there are many Corr-BR and Corr-IO constraints, a language may belong to different classes in this typology for different correspondence constraints—for example, allowing a voiced and voiceless segment to correspond in an output, but requiring correspondents to agree in place of articulation.

The typology also becomes more complicated when markedness constraints are included, as in the vowel-raising case. There, Redup interacts with Corr-BR and *NonFinalMid to determine to which syllable pairs' similarity is preserved in spite of the markedness constraint. The distinction becomes apparent there between candidates like (25a) or (26a), with a reduplicative construal but no enhancement of similarity, and (25c) or (26c), with no reduplicative construal: even without enhancement of similarity, the reduplicative construal can make itself known by blocking an alternation.

There arises the question of why, if there is such a constraint as Redup, no languages exist in which all words are reduplicated. Such a language would be very inefficient—every word's uniqueness point would be at the halfway mark, and the second half of the word would serve no contrastive function. Although the mechanism that prevents pathological grammars from arising is not understood, such a mechanism is needed to prevent many other contrast-reducing constraints from rising to the top of the grammar. For example, the silent language, in which the structure-banning constraint *Struc (Zoll 1993) dominates all faithfulness constraints, does not exist. Similarly, Prince and Smolensky 1993 propose constraints of the form *P/X that forbid X as a syllable nucleus (the less sonorous X is, the more marked it is a nucleus: *P/[t]>> *P/[n] >> *P/[u] >> *P/[a]). But there is no language in which all the *P/X except *P/[a] are undominated. Such a language, all of whose syllable nuclei would be [a], would not be nearly as dysfunctional as the silent language or the fully reduplicated language. (Reduplicated, silent, or one-vowel languages may of course be used by children in the early stages of acquisition, but they do not exist as full-fledged languages in use by any speech community.)

To summarise, Redup requires all words to be construed as reduplicated. In most words, that requirement is overridden by the Corr-BR violations that would occur if dissimilar syllables corresponded, or by the Corr-IO violations that would occur if underlyingly dissimilar syllables were made more similar to satisfy Corr-BR. But if two substrings of a word are sufficiently similar, Redup can be satisfied at little cost to Corr-IO and Corr-BR, and reduplicative over- and underapplication can result. In productively reduplicated words, of course, there is no obstacle at all to satisfying Redup, because the reduplicant has no underlying material of its own.

The tableau in (27) illustrates how the generation account of Aggressive Reduplication would explain the gradient similarity effect seen in vowel raising. Candidate b in fails because the vowels in the base and reduplicant differ in height; c makes the vowels identical, but at the expense of changing an underlying height specification (as mentioned in note 14, double raising does sometimes occur, suggesting a variable ranking for Ident-IO(hi)-nonfinalσ, as we might expect in a language with only marginally contrastive [hi]). Similarly, d makes the consonants identical at the expense of changing various underlying manner features; and e fails because it is not construed as reduplicated (even the candidates that are construed as reduplicated have two violations of Redup, because the third syllable does not correspond to the first, nor to the second). Ident-BR(manner) (shorthand for several Ident-BR(F) constraints) is ranked low enough to allow [t] and [n] to correspond.

(27) Aggressive reduplication blocks vowel-raising

/tono + -an/	Ident-IO (manner)	Ident-IO(hi) nonfinalσ	Ident-BR (hi)	Redup	*Non UltimaMid	Ident-BR (manner)	Ident-IO(hi) finalσ

a ☞ [to]$_\alpha$[no]$_\alpha$han				**	**	*	
b [to]$_\alpha$[nu]$_\alpha$han			*!	**	*	*	*
c [tu]$_\alpha$[nu]$_\alpha$han		*!		**		*	*
d [to]$_\alpha$[to]$_\alpha$han	*!			**	**		
e tonu han				***!	*		*

This type of Aggressive Reduplication can be seen a case of the emergence of the unmarked (McCarthy & Prince 1994). Because input-output faithfulness to all aspects of structure that are contrastive in a language is ranked high, outright enhancement of self-similarity, as in the sporadic cases in English, is expected to be rare.[21] But, even if Corr-IO outranks Corr-BR, preventing enhancement of self-similarity, Redup can still make its effects felt by setting up an internal correspondence relation that preserves internal similarity—here, by blocking an alternation.[22] McCarthy and Prince's schematic ranking for emergence of the unmarked cases in reduplication is Corr-IO >> *X >> Corr-BR: under this ranking, bases may freely contain the marked structure X, but reduplicants obey *X. More generally, emergence of the unmarked occurs when a markedness constraint that appears to be freely violated in one context nonetheless selects among candidates in another context. In the raising case here, the contextual difference is not in whether the constraint Redup is obeyed, but rather in whether its effects can be seen. The crucial ranking is, schematically,

Corr-IO >> Corr-BR >> Redup >> *X

where in this case *X is *NonFinalMid, and Corr-BR is Ident-BR(high). In unsuffixed [tó:no], Redup and Ident-BR(hi) can be satisfied without violating any other constraints, but Corr-IO constraints prevent internal similarity from being enhanced, so the correspondence relation imposed by Redup has no audible effect on the output. Likewise, they prevent a word like [minú:to] 'minute' from being "repaired" to *[minó:to] (though there are some cases that look like this in (8a,c)). In suffixed [tonó:han], however, Redup and Ident-BR(hi) together compel a violation of *NonFinalMid.

In (28), we see that Aggressive Reduplication explains the matching-backness effect straightforwardly. If Ident-BR(back) is ranked high, then a word like [tó:do] can be construed as reduplicated, but a word like [hé:ɾo] cannot, and so *NonUltimaMid causes raising.

(28) The matching-backness effect:

	/todo+in/ 'to include all'	Ident-IO(bk)	Ident-BR(bk)	Ident-BR(hi)	Redup	*NonUlt Mid
a	toduhin				***!	*
b	todohin				***!	**
c	☞ $[to]_\alpha[do]_\alpha$hin				**	**
d	$[to]_\alpha[du]_\alpha$hin			*!	**	**

/heɾo+in/ 'to brand'	Ident-IO(bk)	Ident-BR(bk)	Ident-BR(hi)	Redup	*NonUlt Mid

e	☞ heɾuhin				***	*
f	heɾohin				***	**!
g	[he]$_\alpha$[ɾo]$_\alpha$hin		*!		**	**
h	[ho]$_\alpha$[ɾo]$_\alpha$hin	*!			**	**
i	[he]$_\alpha$[ɾu]$_\alpha$hin		*!	*!	**	**

There is no ranking of these constraints that does the reverse, selecting a in the first tableau, but g in the second. Assuming variable constraint ranking, a word with non-backness-matching vowels can undergo raising, but not more often than an equivalent word with backness-matching vowels.

3.1.1. Accounting for lexical variation

The similarity effects seen in vowel raising, along with the backness and proximity effects, are only lexical tendencies, not reliable rules. It is not the case, for example, that onsets with different places of articulation never correspond (thus blocking raising), only that they are less likely to correspond. I adopt Zuraw's (2000) solution to this general problem of how speakers can learn, use, and maintain patterns in the distribution of exceptions that are merely probabilistic: existing words' properties—in this case, whether or not a stem is raised—are encoded in their lexical entries in some form (perhaps by listing suffixed and unsuffixed forms, in this case[23]). Corr-IO constraints are ranked high and ensure that the listed information is faithfully expressed. Marginal constraints like Redup that do not play a core role in the language are ranked lower and variably, because the language learner is exposed to conflicting information about where to rank them. But when the speech community is in the process of establishing what a loan-stem's new suffixed form should be, Corr-IO is irrelevant—there is no listed information to be faithful to—and the lower-ranked constraints come into play.

When the stem has a mid vowel in the penult, Redup and Ident-BR(hi) favour non-raising, but *NonUltimaMid and other Ident-BR constraints favour raising. The more dissimilar the penult and the ultima, the less likely a reduplicative construal in these early stages.[24]

For example, for [to][do] to be construed as reduplicated, Redup need only outrank Ident-BR(voice). But for ʔes[tor][bo] to be interpreted as reduplicated, Redup must outrank Ident-BR(voice), Ident-BR(place), and Max-R(C). If all four constraints—Redup, Ident-BR(voice), Ident-BR(place), and Max-R(C)) —are variably ranked in the grammar, then the stricter the ranking requirement, the less likely that any randomly chosen ranking will satisfy it. So, a ranking chosen by a speaker on any given occasion is more likely to assign reduplicated structure to todo than to ʔestorbo. This discrepancy influences speaker and hearer behaviour in the early stages of a suffixed word's life, making it more likely that todo will become listed as non-raising, and ʔestorbo as raising.

The reason why only some pseudoreduplicated roots exhibit reduplicative effects is similar: because of high-ranking Corr-IO, an individual word's lexical entry determines whether it displays such effects.[25] Lexical entries are subject to pressure from low-ranking Redup, however, which shapes the development of new affixed forms and newly coined words, and can cause occasional errors in the perception and production of established words.

In summary, under the generation account, speakers impose reduplicative structure wherever possible, because of the constraint Redup, which requires word-internal correspondence. Corr-IO and Corr-BR keep the tendency in check most of the time, but sometimes Corr-IO can be overridden, producing sporadic enhancement of word-internal similarity (e.g., English smorgasborg). In the more systematic vowel-raising case, Redup and Corr-BR preserve word-internal similarity without violating Corr-IO, by blocking an alternation.

3.2. Learning analysis[26]

An alternative to the generation account is that reduplicative construals are imposed during lexical learning. The resulting lexical entries with morphosyntactically unmotivated reduplicative structure would be similar to those proposed by Buckley (1997), Golston and Thurgood (in press), and Gafos (1998) (see Section 1). The Tagalog vowel raising case differs from the cases examined by those authors in that reduplication is imperfect, and segments must be placed in correspondence that have varying amounts of non-shared underlying material (e.g., [t] and [d], [k] and [d]) that has to be prespecified.

In lexical learning, the learner must determine, based on the overt (i.e., audible) portion of a word that she hears, the optimal lexical representation given her grammar. Prince and Smolensky (1993) propose lexicon optimisation as the mechanism by which learners establish a lexical representation. Holding fixed the surface form, the learner finds the optimal underlying-surface pairing, as illustrated in (29). Markedness constraints, such as *[spread glottis] and NoCoda, are irrelevant in lexicon optimisation, because varying the underlying form does not affect how well they are satisfied. The extent to which the underlying form matches the surface form, then, is determined by the interaction of faithfulness constraints with *Spec. *Spec (or perhaps a family *Spec-V, *Spec-[round], etc.) penalizes phonological material in the underlying form, much as *Struc (Zoll 1993) penalizes material in the surface form. Just as *Struc is inactive in lexical learning (because in all candidates, the phonological material in the surface form is held constant), so *Spec is irrelevant in generation, because the underlying material is held constant in all candidates. In (29), candidate a, with no aspiration in the underlying form, is preferred because *Spec >> Ident-IO(spread glottis).[27]

(29) Lexicon optimization

[k^{h}æt]	*Spec	Ident-IO (spread glottis)	*[+spread glottis]	NoCoda
☞ a /kæt/ → [k^{h}æt]		*	*	*
b /k^{h}æt/ → [k^{h}æt]	^{h}![28]		*	*

A reduplicative construal of the underlying form may save violations of *Spec, by representing repeated material (segments or features) once instead of twice. Without committing any particular model of imperfectly reduplicated representations, (30) uses Marantz's (1982) notation to illustrate how an imperfectly reduplicated representation for [todo] contains less phonological material than an unreduplicated representation. (Assume that the incompletely specified C and V slots are filled by association to a copy of the melody [do], with the pre-specified feature [-voice] taking precedence over [d]'s [+voice].) In the tableaux below, (30a) is written /RED([-voice]) do/, with the non-repeated material of the first 'copy' in parentheses after RED.

(30) a. imperfectly reduplicated b. unreduplicated

```
[-voice]
    \
     C V C V          C V C V
         | |          | | | |
         d o          t o d o
```

The lexical learner has one other task that is relevant to Aggressive Reduplication: determining what covert structure an overt form has. Covert structure includes all inaudible aspects of the surface form, such as syllable and foot boundaries (see Tesar 1998, 1999, 2000), or, in this case, reduplicative correspondence.

The tableaux in (31) illustrate the four candidate types that the learner must consider, given an overt form (shown in quotes): the underlying form may be reduplicated (a, b) or not (c, d), and the surface form may have correspondence between the non-identical segments (a, c) or not (b, d). The tableaux illustrate that the ranking of *Spec with respect to base-reduplicant correspondence constraints determines how much deviation from perfect identity is to be tolerated in a reduplicative construal.[29] Although base-reduplicant correspondence constraints are constraints on outputs, they are relevant in lexicon optimisation because their evaluation depends on covert structure. The highly self-similar overt form “todo” is construed as reduplicated because *Spec >> Ident-BR(voice). The *Spec violations “saved” are for the shared features of [t] and [d], and all the features of the second [o]. A less self-similar word (hypothetical “kodo”) is not construed as reduplicated under this ranking, because of the Ident-BR(place) violation that would result.

(31) Reduplicative construal in lexicon optimization

“todo”	Max-BR	Ident-BR(place)	*Spec	Ident-BR(voice)
☞ a /RED([-voice]) do/ → [$[\text{to}]_R[\text{do}]_B$]			[-voice]do	*
b /RED([-voice]) do/ → [$\text{t}[\text{o}]_R[\text{do}]_B$]	*!		[-voice]do	
c /todo/ → [$[\text{to}]_R[\text{do}]_B$]			todo!	*
d /todo/ → [todo]			todo!	
“kodo”				
e /RED([-voice, dorsal]) do/ → [$[\text{ko}]_R[\text{do}]_B$]		*!	[-voice][dorsal]do	*
f /RED([-voice, dorsal]) do/ → [$\text{k}[\text{o}]_R[\text{do}]_B$]	*!		[-voice][dorsal]do	
g /kodo/ → [$[\text{ko}]_R[\text{do}]_B$]		*!	kodo	*
☞ h /kodo/ → [kodo]			kodo	

Although the choice between candidates a and b that Max-BR makes is irrelevant to lexical learning itself (a and b have the same underlying form), Max-

BR must outrank Ident-BR(voice) because otherwise, in generation, underlying reduplicated structure would be ignored if a violation of base-reduplicant correspondence constraints could thereby be avoided. As shown in (32), if Max-BR outranks the base-reduplicant correspondence constraints violated by the imperfect reduplication (here, Ident-BR(voice)), then reduplicative structure in the input is preserved, and if Max-BR >> *NonFinalMid, vowel raising is prevented, as in the Tagalog case.

(32) Generation from a reduplicated underlying form

/RED([-voice]) do+an/	Max-R	Ident-BR(voice)	Ident-BR(hi)	*NonFinalMid
☞ a [$[to]_R[do]_B$han]		*		*
b [$[to]_R[du]_B$han]		*	*!	
c [$t[o]_R[do]_B$han]	*!			*
d [$t[o]_R[du]_B$han]	*!			

As with the generation account, we must allow for variable constraint ranking, so that a word that is self-similar along many dimensions has a higher probability of getting a reduplicated underlying representation than a word that is self-similar along few dimensions. In the Tagalog vowel-raising case, the self-similarity of an unsuffixed form would influence whether individual speakers create a reduplicated or unreduplicated lexical representation for it. Whether or not a stem undergoes vowel raising when suffix would still need to be separately encoded, however, at least in those few words that lack a mid vowel in the penult and yet resist raising (tunél 'tunnel', tunel-án 'to tunnel').

And as with the generation account, there is a gap in the factorial typology. There is no language that ranks *Spec (and Max-BR) so high that every word with a repeated feature anywhere receives a partially reduplicated lexical entry (to satisfy *Spec), and that ranks all Corr-BR constraints above all markedness constraints. In such a language, no alternation could change a feature to give it a different value than another occurrence of that feature in the word, perhaps with an additional requirement that the two features occur in similar prosodic positions in successive syllables. For example, /tiwa/ could undergo a postvocalic spirantisation rule (/a+tiwa/ → [asiwa]), because even if the /t/ and /w/ are in correspondence, changing /t/ to [+continuant] does not violate Ident-BR(cont). But, /tiga/ could not undergo spirantisation. Similarly, /mob/ could undergo final devoicing, but /ladmob/ could not, and so on for all rules of the language.

To conclude this section, similarity preservation of the type seen in Tagalog vowel raising can be straightforwardly explained by lexical learning, as long as reduplicative structures are allowed in monomorphemic underlying forms. Similarity enhancement, however (e.g., orangutang), cannot be straightforwardly explained. The reason is that lexicon optimisation, as currently understood, requires the underlying form that the learner chooses to generate the observed surface form. Speakers who innovated the form orangutang after hearing orangutan would have to have constructed a lexical entry that does not generate the surface form they were exposed to. Therefore, we must allow for the possibility of mishearing (or misremembering), or of constructing underlying forms that do not quite generate the desired surface form. Neither possibility seems implausible—mishearing influenced by top-down expectations certainly occurs—but both are beyond the scope of current OT learning theories, and, because the cases of outright enhancement presented here are sporadic and not systematic, I will leave them treating them under the lexical learning account as a topic for future research.

3.3. Distinguishing between learning and production

How can we determine whether Aggressive Reduplication is caused by *Spec's activity in lexical learning or by Redup's activity in generation? This section discusses first some theory-internal considerations briefly, then some empirical ones at greater length.

3.3.1. Theory-internal considerations

The generation account proposed in 3.1 requires a new constraint, Redup, whereas *Spec is generally necessary in lexical learning. Introducing new constraints is potentially dangerous because of gaps in factorial typology that may result. Both accounts, however, suffer from a gap in the factorial typology, as discussed in 3.1 and 3.2 respectively. In the generation account, there is no language in which Redup and all the Corr-BR constraints outrank all the Corr-IO constraints, causing all words to be produced reduplicated. In the learning account, there is no language in which *Spec outranks all the Corr-BR constraints, which in turn outrank all markedness constraints, so that any word with even one repeated feature value is construed as reduplicated, blocking alternation.

The lexical-learning account uses the same mechanisms to represent pseudoreduplicated words and morphologically reduplicated words (underlying RED), as well as the same mechanisms to perform generation on them (Max-BR, Ident-BR[F], etc.). Is the same true of the generation account? In the generation account, the lexical entry of a pseudoreduplicated word like bakbak is just /bakbak/, and does not contain anything like a RED morpheme (/RED+bak/). It is Redup that drives correspondence between the two baks. In the case of productive reduplication, we can view RED as a morpheme that requires some phonological material to appear in the word, without caring precisely what. When there is a RED in the input, Redup can be satisfied at little cost to Corr-BR constraints, because the reduplicant can be made as similar to the corresponding portion of the base as the grammar allows, without incurring Corr-IO violations. This is illustrated schematically for a hypothetical language in (33). Candidate g fails to realize the RED morpheme at all; in f, the realization of RED contains less-marked material, but at the cost of violating Redup. Among the remaining candidates, the emergence-of-the-unmarked ranking Ident-IO(Vplace) >> *VPlace >> Ident-BR(VPlace) causes vowel quality to be unfaithfully copied, but the ranking Ident-IO(Cplace), Ident-BR(Cplace) >> *CPlace causes consonant quality to be faithfully copied.

(33)

/RED+bili/	Realize Morph[30]	Redup	Ident-IO(Cplace)	Ident-IO(Vplace)	Ident-BR(Cplace)	*VPlace	Ident-BR(Vplace)	*CPlace
☞ a [bə]$_\alpha$–[bi]$_\alpha$li						**	*	***
b [bi]$_\alpha$-[bi]$_\alpha$li						***!		***
c [ʔə]$_\alpha$-[bi]$_\alpha$li					*!	**	*	**
d [bə]$_\alpha$-[bə]$_\alpha$li				*!		*		***
e [ʔə]$_\alpha$-[ʔi]$_\alpha$li			*!			**	*	*
f ʔə-bili		*!				**		**
g bili	*!					**		**

The main difference between this account of productive reduplication and McCarthy and Prince's (1995) is that it is Redup here, not Max-BR, that rules out candidates like (33f). ((33f) satisfies the morphologically insensitive Max-R vacuously.) Otherwise, matters such as emergence of the unmarked, and the size and shape of the reduplicant receive similar treatments.

A final theory-internal consideration is that in the generation account, outright similarity enhancement in self-similar words receives straightforward account using previous approaches to morphological reduplication. In the lexical-learning account, however, similarity enhancement does not follow from lexicon optimisation as currently understood, but requires further development of the theory of lexical learning, which is necessary anyway to allow for mislearning.

3.3.2. Empirical considerations

To distinguish between the learning and generation accounts on empirical grounds, we need cases in which the self-similarity characteristics that are available in lexical learning are different from those that are apply when the reduplicative construal is to make itself apparent. In the Tagalog vowel-raising case, for example, lexical learning, and the possibility of a reduplicative construal being encoded in the lexical entry, presumably takes place based on the unsuffixed form, because (i) that is the form in which the loan stems enter the language,[31] and (ii) the suffixed form itself provides evidence to the learner as to whether the stem should raise or not, so that her construal of the stem is constrained by the lexical entry of her interlocutor (i.e., to the extent that speakers treat a word consistently, the next generation is not in a position to decide whether the word should be construed reduplicatively). In the lexical-learning account, it is therefore the self-similarity of the unsuffixed form that matters for a reduplicative construal. By contrast, in the generation account, although unsuffixed forms are subject to Redup, the evidence for a reduplicative construal comes from whether the suffixed form displays raising. Therefore, the distribution of raising should be sensitive to the self-similarity of the suffixed form, not of the unsuffixed form, under the generation account.

When can the unsuffixed and suffixed forms differ in self-similarity?[32] There are a few cases in Tagalog, though unfortunately none of them offer enough relevant tokens to distinguish between the generation and learning accounts. First is rime shape. Because suffixation always opens the final syllable of the stem (a final consonant becomes an onset), identical rime shapes (open vs. closed) in the penult and ultima of an unsuffixed stem may become non-identical when the stem is suffixed, and vice-versa.

Which has a greater effect on raising, then: similarity of rime shape in the unsuffixed form, or similarity in the suffixed form? The crucial words are those ending in a consonant, because their similarity properties change under suffixation (rows b and d in (34)). Because stems from English have higher rates of non-raising than stems from Spanish, and are all consonant-final, counts in (34) are shown separated by language of origin and totalled. As predicted by both the generation and learning accounts, the 'same-same' (a) cells have significantly more nonraising than the 'different-different' (c) cells, both for Spanish origin ($p<.05$), and overall (there are no English-origin tokens). Holding constant the self-similarity of the suffixed forms, self-similarity of the unsuffixed forms has an effect: although there is no significant difference between rows a and b, the rate of non-raising is significantly higher in row d than in row c, as predicted by the lexical learning account, if the total counts are used ($p<.05$), but not if the counts are restricted to Spanish. Holding constant the self-similarity of the unsuffixed forms, there is no significant difference between rows a and d, but the rate of non-raising is significantly higher in row b than in row c, as predicted by the generation account, whether using total counts ($p<.001$) or Spanish counts alone ($p<.005$). Which has a stronger effect, similarity in the unsuffixed form or similarity in the suffixed form? The overall rate of nonraising is higher when only the suffixed form is self-similar (b) than when only the unsuffixed form is self-similar (d), but the difference is not significant. Rime shape fails, therefore, to distinguish between the lexical-learning and generation accounts, and suggests that both play a role.

(34)

	unsuffixed shapes	suffixed shapes	%nonraising (Spanish origin)	%nonraising (English origin)	total %nonraising
a	same (t**o**.d**o**)	same (t**o**.d**o**.-hin)	15/41	0/0	37%
b	different (k**o**.ɾ**ek**)	same (k**o**.ɾ**e**.k-in)	6/9	3/5	64%
c	different (t**on**.t**o**)	different (ka.-t**on**.t**o**.-han)	4/32	0/0	12%
d	same (d**ok**.t**or**)	different (d**ok**.t**o**.ɾ-in)	2/8	4/4	50%

Another similarity property that can change under suffixation is stress. (Vowel length is omitted from the transcriptions in this paragraph but discussed below.) Native Tagalog roots can have stress on the penult or the ultima. In either case, stress shifts one syllable to the right in most suffixal constructions (some constructions require a different stress pattern). If the penult is stressed, it must be open (búlaʔ 'lie', suffixed as buláʔ-an 'liar'); if the ultima is stressed, the penult may be closed or open (buláʔ 'froth', pa-bulaʔ-în 'to cause to froth'; bulták 'overfed', bultak-în 'to overfeed'). In loans, a stressed, closed penult or a stressed antepenult may also occur. When such roots are suffixed, primary stress shifts to the final syllable, with the originally stressed syllable sometimes receiving secondary stress (dilíɾjo ‘delirium’, ka-dilìɾju-hán ‘state of delirium’; telépono ‘telephone’, telèpono.-hán ‘to telephone’), and sometimes not (tónto ‘stupid’, ka-tonto-hán ‘stupidity’).

Except in stems with antepenult stress, the stress of the penult and the ultima are always different in the unsuffixed stem (35a-e). The stresses may be different (a,c) or the same (b,d,e) in the suffixed form. Comparing a to b, the rate of non-raising is higher in b, but not significantly so; and comparing c to d and e, the rate of non-raising is higher in d and e, but not quite significantly so (p=.052). The (non-significant) effect is in the direction predicted by the generation account, but the stress data cannot test the prediction of the lexical-learning account (that would require comparing a-e with f, which has only two tokens).

(35)

	unsuffixed	stress	suffixed	stress	%non-raising
a	t**ó**.d**o**	different	t**o**.d**ó**.-hin	different	12/39 = 31%
b	k**o**.ɾ**é**k	different	k**o**.ɾ**e**.k-în	same	2/5 = 40%
c	di.l**í**ɾ.j**o**	different	ka.-di.l**ì**ɾ.j**u**.-hán	different	0/9 = 0%
d	t**ó**n.t**o**	different	ka.-t**o**n.t**o**.-hán	same	5/20 = 25%
e	d**o**k.t**ó**ɾ	different	d**o**k.t**o**.ɾ-ín	same	5/11 = 45%
f	te.lé.p**o**.n**o**	same	te.lè.p**o**.n**o**.-hán	same	2/2 = 100%

With the respect to a third property, length, it is unclear how to divide the data. Stressed, nonfinal syllables are definitely long, but there is disagreement about final syllables. Final syllables have been treated as phonemically short (Schachter & Otanes 1972, who derive stress from length), but Zhang (2001) finds them to be phonetically long, even when unstressed. The similarity of the suffixed forms, shown in (36), is the same under either view, because the syllables in question are non-final. Under Schachter and Otanes' phonemic view, (36d) would be the only row in which the properties of the unsuffixed and suffixed forms differ, so we cannot compare the predictions of the lexical-learning and generation accounts. Under Zhang's phonetic view, the properties of the unsuffixed and suffixed forms differ in all rows but (36d). We can compare the rate of non-raising in the 'different-same' rows (36b,e,f), which the generation account predicts to be higher (it is 50%), with the rate of non-raising in the 'same-different' rows (36a,c), which the lexical-learning account predicts to be higher (it is 25%). The difference favours the generation account, but not significantly so (p=.052).

(36)

	unsuffixed	length-phonemic	length-phonetic	suffixed	length	%non-raising
a	t**ó**.d**o**	different	same	t**o**.d**ó**.-hin	different	12/39 = 31%
b	k**o**.ɾ**é**k	same	different	k**o**.ɾe.k-ín	same	2/5 = 40%
c	di.l**ǐ**ɾ.j**o**	different	same	ka.-di.l**ǐ**ɾ.j**u**.-hán	different	0/9 = 0%

d	t**ó**n.t**o**	different	same	ka.-t**o**n.t**o**.-hán	same	5/20 = 25%
e	d**o**k.t**ó**ɾ	same	different	d**o**k.t**o**.ɾ-ín	same	5/11 = 45%
f	te.lé.p**o**.n**o**	same	different	te.lè.p**o**.n**o**.-hán	same	2/2 = 100%

A fourth example would be initial tapping. Prefixation with a vowel-final prefix such as ka- causes a stem-initial consonant to be intervocalic. If this consonant was [d], it often becomes [ɾ] (dáːmot 'stinginess', ka-ɾamúːt-an 'stinginess'). If the stem is disyllabic (so that the initial consonant is the penult's onset), flapping changes the similarity of the penult and ultima onsets. There are only two disyllabic stems beginning with [d] in the database (doktoɾ 'doctor' and doble 'double'), however, and neither occurs with a vowel-final prefix. We could also look at stem-final [d]s flapping under suffixation alone, but there are none in the database.

A fifth and final case would be nasal substitution. Certain prefixes optionally cause a stem-initial obstruent to become a homorganic nasal, which, in a two-syllable stem, could cause the penult's onset to become more or less similar to the ultima's (**k**upás 'faded', ma-**ŋ**úːpas 'to become faded'). There are, however, no two-syllable stems in the database that take a potentially nasal-substituting prefix.

There are some sporadic examples from another Western Austronesian language, Palauan, that show that a reduplicative construal can be conditioned solely by similarity properties of the affixed form, suggesting that Aggressive Reduplication does take place in generation, though of course generation need not be the sole site. In Palauan, when suffixation shifts stress, the formerly stressed vowel undergoes reduction (Wilson 1972; Flora 1974; Josephs 1975, 1990). Short vowels usually either reduce to schwa or delete, depending in part on the consonantal context. There is variation in some words in how the formerly stressed vowel reduces, and in whether other vowels in the word undergo additional reduction. When the third-person possessive suffix [–ɛl] is attached to kləkədáll, 'parts', the result has variants kləkədəll-ɛ́l, with reduction of the stressed [a] to schwa; and kdəkdəll-ɛ́l, with additional deletion of the preceding [ə]. What is unexpected in the second variant is that the [l] of the onset cluster changes to [d], so that the first two consonant clusters are identical.[33] No other instances of [l] changing to [d] were found in a database of 1,019 suffixed nouns, extracted from Josephs 1990. This change cannot be conditioned by a reduplicated lexical entry for the bare stem, because it is the vowel deletion, found only in the suffixed form, that sets up the condition of internal similarity (by creating a sequence [kləkdə...]), which is then enhanced (by turning the first [l] into a [d]).

Diphthongs in Palauan reduce to their fronter or higher member when unstressed (in a conflict, the frontness preference prevails): orɛ̯ómǝl 'forest', orɛmǝl-ɛ́l 'her/his/its forest'. There are three cases of an unstressed [ǝ] in the unsuffixed form becoming [ɛ] when a following diphthong reduces to [ɛ] (no other instances of schwa changing to [ɛ] were found in the database):

(37)	mǝŋǝrɛ̯ómǝl	'preserve'	mǝŋɛrɛmǝl-ɛ́l	'its captain'
	ʔǝllɛ́o̯ʔ	'spoiling'	ʔɛllɛʔ-ɛ́l	'her/his spoiling (of a child)'

bləkɛ́u̯	‘bravery’	**blɛkɛŋ**-ɛ́l	‘her/his bravery’

Again, the change from [ə] to [ɛ] cannot be the result of a reduplicated lexical entry for the stem, because the change does not occur in the unsuffixed form; indeed, the reason for the change occurs only when suffixation causes reduction of the following nucleus to [ɛ] (apparently, a nucleus [ɛo] or [ɛu] is not sufficiently similar to [ə] to cause the change to [ɛ]).

4. Other approaches to word-internal similarity

This section reviews some related proposals of constraints or conditions that promote word-internal similarity.

4.1. MacEachern’s BeIdentical

MacEachern (1999) surveys languages with laryngeal co-occurrence restrictions and finds that many languages forbid segments with similar laryngeal features from co-occurring within a morpheme. For example, Cuzco Quechua allows only one aspirated stop per morpheme and only one ejective per morpheme (there are additional restrictions). In some languages, however, laryngeally similar segments are allowed just in case they are identical. For example, Peruvian Aymara[34] allows only one ejective per morpheme (appearing on the leftmost eligible consonant), unless the ejectives are identical ([k’ink’u] ‘clay’), and only one aspirated stop per morpheme, unless the aspirated stops are identical ([p^husp^hu] ‘boiled beans’). MacEachern finds that in languages with such an ‘escape clause’ for identical segments, laryngeally dissimilar homorganic consonants are forbidden or rare (*tʰata, *t’ata).

MacEachern attributes the identity escape clause to a constraint BeIdentical, which requires all consonant pairs within a morpheme to be identical. If Ident-IO(place) >>BeIdentical >>*LaryngealSimilarity >> Ident-IO(laryngeal),[35] then homorganic consonants will always be identical, and non-homorganic consonants must not be laryngeally similar:

(38)

/k’at’a/	Ident-IO (place)	Be Identical	*Laryngeal Similarity	Ident-IO (laryngeal)
a ☞ k’ata		*		*
b k’at’a		*	*!	
c t’at’a	*!		*	
/t’ata/				
d ☞ t’at’a			*	*
e t’ata		*!		

In terms of Aggressive Reduplication, this is equivalent to the ranking Ident-IO(place), Ident-BR(place), Ident-BR(laryngeal) >> Redup >> *LaryngealSimilarity >> Ident-IO(laryngeal). That is, consonants that already have the same place of articulation (high-ranked Ident-IO(place) and Ident-BR(place) require this) are forced to have the same laryngeal features also:

(39)

/k’at’a/	Ident-IO (place)	Ident-BR (place)	Ident-BR (laryngeal)	Redup	*Laryngeal Similarity	Ident-IO (laryngeal)
a k’at’a				*	*!	*
b ☞ k’ata				*		*
c [k’a][ta]		*!	*!			*
d [k’a][t’a]		*!			*	
e [t’a][t’a]	*!				*	
/t’ata/						
f t’ata				*!		
g [t’a][ta]			*!	*		
h ☞ [t’a][t’a]					*	*

How can we test whether the escape clause is due to categorical BeIdentical, which requires only that consonants be identical, or to Aggressive Reduplication, which requires maximal similarity between corresponding syllables? The Aggressive Reduplication analysis predicts that, if Ident-BR constraints on vowel features are variably ranked, identical ejectives or aspirated stops should be followed by identical vowels more often than would otherwise be expected—that is, we should find many words like [p^husp^hu] and few words like [k’ink’u]. MacEachern (p.c.) reports that this does seem to be the case, and a dictionary count confirms her observation. In Deza Galindo’s (1989) dictionary, the vowels are identical in 21 of 25 $C^h{}_1V_2$(C)$C^h{}_1V_3$... roots (84%), and in 18 of 23

$C'_1V_2(C)C'_1V_3$... roots (78%) . By contrast, in a sample of the first consonant-initial word (if any) on each page of the Aymara-to-Spanish portion of the dictionary, only 72 of 139 (52%) words have identical vowels in the first two syllables ($p<.0005$ for comparing the first two root types with the controls by Fisher's Exact Test). In Ayala Loayza's (1988) dictionary, the counts are similar, with the first two vowels being identical in 18 of 22 $C^h_1V_2(C)C^h_1V_3$... roots (82%), and 18 of 18 $C'_1V_2(C)C'_1V_3$... roots (100%), but only 54 of 114 of controls (47%) ($p<.0001$).

Mester (1988) handles identity escape clauses of this type by treating identical segments as a single feature bundle, multiply associated to two segmental slots. (This raises some problems for those cases where another consonant intervenes between the two identical consonants.) Such structures would then escape a ban on multiple [constricted glottis] or [spread glottis] features within a root, because there is only one such feature, but it belongs to a multiply associated consonant. This autosegmental account does not, however, explain why identity of following vowels would encourage such multiple linking of consonants.

4.2. Suzuki's IdentOns

Suzuki (1999), in reanalysing data from Cohn (1992), proposes a constraint family IdentOns that requires onsets of adjacent syllables to be similar—in this case, to have the same value for [rhotic]. This constraint explains why dissimilation of r...r to r...l is blocked in Sundanese when the two /r/s are onsets of adjacent syllables. Cohn explains the phenomenon, which occurs both under –arinfixation and within roots, by linking two adjacent-onset /r/s to a single [-lateral] specification. Suzuki's proposal differs from MacEachern's in requiring only consonants that are in a particular prosodic arrangement to be similar (as Suzuki points out, in MacEachern's data, ejectives and aspirated consonants are restricted to onset position anyway, ruling out distinctions based on syllable position).

A more general Aggressive Reduplication analysis of Sundanese would predict that other similarities between adjacent /r/-initial syllables should further discourage dissimilation. This prediction can be tested, however, only if there is some variation in dissimilation. Cohn does report variation in –ar- infixation: in 2 of 22 infixed /CVrCV.../ words, both speakers she consulted unexpectedly infix –al-, and one speaker find both –ar- and –al- acceptable for an additional 5 words. Cohn does not however, list these words. Examining the 105 roots found in a dictionary (Lembaga Basa & Sastra Sunda 1985) with more than one /r/, Cohn finds that in 57 of 67 $rV_1rV_2...$ roots, the vowels are identical (e.g., rorod 'pull in'), and in 19 of 20 $rV_1C_1rV_2C_2$ roots, $V_1C_1 = V_2C_2$ (e.g. ribrib 'arms overly full'). Among the 18 roots of other shapes, 17 are recent loans. The data from roots appear to support the prediction of Aggressive Reduplication: successive liquid onsets that escape a general dissimilation process are likely to belong to syllables that are similar in other ways.

4.3. Yip's Repeat

Yip (1998) proposes a family of Repeat constraints. For example, $\text{Repeat}_{\text{Plural}}$, active in Javanese, requires an output to contain two identical elements when the input contains a plural morpheme. Because members of the Repeat family are specific to various input morphemes (which are thereby reduplicative morphemes), Repeat constraints, unlike Redup, drive reduplication only when it is morphologically called for.

4.4. Walker and Rose's Consonantal Correspondence

Walker and Rose (Walker 2000a, Walker and Rose to appear) propose a family of constraints that require consonants to enter into correspondence if they already share certain feature values. This constraint family is similar to Redup in that perfect identity is not required—only a correspondence relation is required, and it is left to other constraints to enforce similarity (partial or total) along other dimensions. Walker and Rose's proposal, which I will refer to as Consonantal Correspondence, does not predict that other segments should have any effect on encouraging correspondence between consonants. They do propose a constraint that requires corresponding consonants to be in the same syllabic position, however.

Aggressive Reduplication and Consonantal Correspondence differ in two ways. First, Aggressive Reduplication brings about correspondence between substrings, so that similarity between one pair of segments (e.g., two nuclei) can cause similarity enhancement in another (e.g., their onsets). The data from Tagalog (where onsets and codas affect nuclei), Aymara (where nuclei affect onsets), and Sundanese (where, again, nuclei affect onsets) support this feature of Aggressive Reduplication over correspondence that is restricted to pairs of segments.

The second way in which Aggressive Reduplication and Consonantal Correspondence differ is that Redup on its own penalizes non-correspondence regardless of prior similarity, whereas Consonantal Correspondence constraints apply only of some threshold of similarity is met. Aggressive Reduplication is able to require prior similarity by constraint interaction, however, for example by ranking both Ident-IO(F) and Ident-BR(F) high. Therefore, these two types of correspondence-requiring constraint make largely overlapping empirical predictions. There is, however, one exception.[36] Only Consonantal Correspondence can produce a system in which all consonants that are similar to at least some degree become identical, and less-similar consonants do not assimilate at all. For example, given the Consonantal Correspondence constraint ranking in (40), /daba/ → $[[da]_{\alpha}[da]_{\alpha}]$, and /data/ → $[[da]_{\alpha}[da]_{\alpha}]$ (because the input consonants are already identical in either voicing or place), but /dapa/ → $[[da]_{\alpha}[pa]_{\beta}]$ (no change).

(40)

/daba/	Ident-BR (place)	Ident-BR (voice)	CorrIf Ident (place)	CorrIf Ident (voice)	Ident-IO (place)	Ident-IO (voice)
$[da]_{\alpha}[ba]_{\beta}$				*!		
$[da]_{\alpha}[ba]_{\alpha}$	*!					
☞ $[da]_{\alpha}[da]_{\alpha}$					*	
/data/						
$[da]_{\alpha}[ta]_{\beta}$			*!			

$[da]_{\alpha}[ta]_{\alpha}$		*!				
☞ $[da]_{\alpha}[da]_{\alpha}$						*
/dapa/						
☞ $[da]_{\alpha}[pa]_{\beta}$						
$[da]_{\alpha}[ba]_{\beta}$				*!		*
$[da]_{\alpha}[ta]_{\beta}$			*!		*	
$[da]_{\alpha}[pa]_{\alpha}$	*!	*!				
$[da]_{\alpha}[ba]_{\alpha}$	*!					*
$[da]_{\alpha}[ta]_{\alpha}$		*!			*	
$[da]_{\alpha}[da]_{\alpha}$					*!	*!

In Aggressive Reduplication, by contrast, if Redup and the Ident-BR(F) constraints are ranked high enough to force the violations of Ident-IO(place) and Ident-IO(voice) in /daba/ → $[[da]_{\alpha}[da]_{\alpha}]$ and /data/ → $[[da]_{\alpha}[da]_{\alpha}]$, respectively, then they are high enough to force violations of both Ident-IO constraints in /dapa/ → $[[da]_{\alpha}[da]_{\alpha}]$:

/daba/	Ident-BR (place)	Ident-BR (voice)	Redup	Ident-IO (place)	Ident-IO (voice)
$[da]_{\alpha}[ba]_{\beta}$			*!		
$[da]_{\alpha}[ba]_{\alpha}$	*!				
☞ $[da]_{\alpha}[da]_{\alpha}$				*	
/data/					

$[da]_\alpha[ta]_\beta$			*!		
$[da]_\alpha[ta]_\alpha$		*!			
☞ $[da]_\alpha[da]_\alpha$					*
/dapa/					
$[da]_\alpha[pa]_\beta$			*!		
$[da]_\alpha[pa]_\alpha$	*!	*!			
$[da]_\alpha[ba]_\alpha$	*!				*
$[da]_\alpha[ta]_\alpha$		*!		*	
☞ $[da]_\alpha[da]_\alpha$				*	*

Which of these predictions better matches the actual typology is unclear. There are languages in which consonants must either match in both laryngeal features and place, or differ in both (MacEachern 1999), but in those cases there are independently attested constraints against laryngeal similarity. These cases can be analysed with either Consonantal Correspondence or Aggressive Reduplication.

5. Conclusion

I have proposed that there is a purely phonological drive for words to be reduplicated. The proposal is supported by the ambiguous behaviour of pseudoreduplicated words in Tagalog—sometimes they display reduplicative over- and under-application, sometimes they fail to—and by the distribution of exceptions to vowel raising. Vowel raising tends to be blocked when it would create dissimilarity between the penult and ultima vowels. There are several possible explanations for this, but only Aggressive Reduplication explains why raising is even less likely when the penult and ultima are similar in other ways: greater similarity results in fewer impediments to a reduplicative construal, and a reduplicative construal blocks raising.

Two possible sources of morphosyntactically unmotivated reduplicative construals were considered: a constraint in generation that requires word-internal correspondences, and a structure-minimizing drive in lexical learning. Tagalog data provide insufficient basis for distinguishing between these two explanations, but some sporadic examples from Palauan point to the generation account rather than the learning account.

The idea of purely phonological drives for imposing morphological structure has been proposed by other researchers. As mentioned in the introduction, Hammond (1999) argues that English adjectives are treated as suffixed for purposes of stress if the final syllable is segmentally identical to some real adjectival suffix. Such phonologically based detection of affixes must be fairly widespread, since it is a necessary step in back-formation, though semantic factors play a role there too.[37] Baroni (2000) and Goldsmith (2001) demonstrate that a substantial amount of morphology can be learned from a corpus without morphosyntactic information, by relying on distributional information (the recurrence of substrings) alone. Perhaps morphosyntax is only one of the clues that language learners and users rely on to detect morphological structure.

References

Archangeli, Diana, Laura Moll, and Kazutoshi Ohno (1998). Why not *NÇ̥? CLS 34. 1-26.

Ayala Loayza, Juan Luis (1988). Diccionario español-aymara, aymara-español. Lima, Peru: Editorial J. Mejia Baca.

Baroni, Marco (2000). Distributional cues in morpheme discovery: a computational model and empirical evidence. PhD dissertation, UCLA.

Boersma, Paul (1998). Functional phonology: formalizing the interactions between articulatory and perceptual drives. The Hague: Holland Academic Graphics.

Buckley, Eugene (1997). Integrity and correspondence in Manam double reduplication. NELS 28. 59-67.

Carrier, Jill (1979). The interaction of morphological and phonological rules in Tagalog. PhD dissertation, MIT.

Cohn, Abigail (1992). The consequences of dissimilation in Sundanese. Phonology 9. 199-220.

Deza Galindo, Juan Francisco (1989). Jaya mara aru: nuevo diccionario aymara-castellano, castellano-aymara. Lima, Peru: Graphos 100 Editores.

English, Leo James (1986). Tagalog-English dictionary. Manila: Congregation of The Most Holy Redeemer.

Flora, Marie (1974). Palauan Phonology and Morphology. PhD dissertation, University of California, San Diego.

Gafos, Diamandis (1998). Eliminating long-distance consonantal spreading. Natural Language and Linguistic Theory 16. 223-278.

Notes

Goldsmith, John (2001). Unsupervised learning of the morphology of a natural language. Computational Linguistics 27. 153-198.

Golston, Chris and Elzbieta Thurgood (in press). Reduplication as echo: evidence from Bontok and Chumash. Ritva Laury, Gerald McMenamin, Shigeko Okamoto, Vida Samiian, and Karumuri V. Subbarao (eds.) Papers in honor of P.J. Mistry. New Delhi: Indian Institute of Language Studies.

Hammond, Michael (1999). English stress and cranberry morphs. Paper presented at the LSA, Los Angeles.

Hayes, Bruce and May Abad (1989). Reduplication and syllabification in Ilokano. Lingua 77. 331-374.

Hayes, Bruce, Bruce Tesar, and Kie Zuraw (2000). OTSoft. Software package, http://www.linguistics.ucla.edu/people/hayes/otsoft/.

Itô, Junko and Armin Mester (1995). Japanese phonology. In John Goldsmith (ed.) The handbook of phonological theory. Cambridge, Mass.: Blackwell. 817-838.

Josephs, Lewis (1975). Palauan Reference Grammar. University Press of Hawaii, Honolulu.

Josephs, Lewis (1990). New Palauan-English Dictionary. University Press of Hawaii, Honolulu.

Jun, Jongho (1995). Perceptual and articulatory factors in place assimilation: an Optimality-theoretic approach. PhD dissertation, UCLA.

Kaun, Abigail (1995). The typology of rounding harmony: an optimality theoretic approach. PhD dissertation, UCLA.

Kessler, Brad (2001). The significance of word lists. Stanford, Cal.: CSLI.

Lembaga Basa & Sastra Sunda (1985). Kamus umum basa Sunda, cetatan kelima. Bandung, Indonesia: Penerbit Tarate Bandung

Lichtenberk, Frantisek (1983). A grammar of Manam. Honolulu: University of Hawaii Press.

MacEachern, Margaret (1999). Laryngeal cooccurrence restrictions. New York: Garland.

Marantz, Alec (1982). Re reduplication. Linguistic Inquiry 13. 435-482.

McCarthy, John and Alan Prince (1994). The emergence of the unmarked: optimality in prosodic morphology. In Mercè Gonzàlez (ed.) NELS 24. 333–379.

McCarthy, John and Alan Prince (1995). Faithfulness and reduplicative identity. In Jill Beckman, Suzanne Urbanczyk and Laura Walsh Dickey (eds.) University of Massachusetts occasional papers in linguistics 18: papers in optimality theory. 249–384.

Mester, R. Armin (1988). Studies in tier structure. New York: Garland.

Nash, David (1980). Topics in Warlpiri grammar. PhD dissertation, MIT.

Prince, Alan and Paul Smolensky (1993). Optimality theory: constraint interaction in generative grammar. Ms, Rutgers University and University of Colorado at Boulder.

Schachter, Paul and Fe Otanes (1972). Tagalog reference grammar. Berkeley: University of California Press.

Steriade, Donca (1988). Reduplication and syllable transfer in Sanskrit and elsewhere. Phonology 5. 73-155.

Steriade, Donca (1997). Phonetics in phonology: the case of laryngeal neutralization. Ms., UCLA.

Steriade, Donca (1999). Alternatives to the syllabic interpretation of consonantal phonotactics. In Osamu Fujimura, Brian Joseph, and Bohumil Palek (eds.) Proceedings of LP '98 [= 4th Linguistics and Phonetics Conference]: Item Order in Language and Speech. Prague: Karolinum Press. 205-242.

Suzuki, Keiichiro (1999). Identity ? similarity: Sundanese, Akan, and tongue twisters. Paper presented at the LSA, Los Angeles.

Tesar, Bruce (1998). An Iterative Strategy for Language Learning. Lingua 104.131-145.

Tesar, Bruce (1999). Robust Interpretive Parsing in Metrical Stress Theory. WCCFL 17. 625-639.

Tesar, Bruce (2000). Using Inconsistency Detection to Overcome Structural Ambiguity in Language Learning. RuCCS Technical Report TR-58, Rutgers University.

Walker, Rachel (2000a). Long distance consonantal identity effects. WCCFL 19. 532-545.

Walker, Rachel (2000b). Nasal reduplication in Mbe affixation. Phonology 17. 65-115.

Walker, Rachel and Sharon Rose (submitted). A typology of consonant agreement as correspondence.

Wilbur, Ronnie Bring (1973). The phonology of reduplication. Bloomington, Ind.: Indiana University Linguistics Club.

Wilson, Helen (1972). The Phonology and Syntax of Palauan Verb Affixes. PhD dissertation, University of Hawaii.

Yip, Moira (1998). Identity avoidance in phonology and morphology. In Steven Lapointe, Diane Brentari, and Patrick Farrell (eds.) Morphology and its relation to phonology and syntax. Stanford, Cal: CSLI.

Zhang, Jie (2001). The contrast-specificity of positional prominence: evidence from diphthong distribution. Handout from talk given at LSA, Washington, DC.

Zoll, Cheryl (1993). Directionless syllabification and ghosts in Yawelmani. Transcript of talk given at ROW-1, Rutgers University.

Zuraw, Kie (2000). Patterned exceptions in phonology. PhD dissertation, UCLA.

---

[*] [Acknowledgements suppressed for now to maintain anonymity]

[1] [suppressed acknowledgements for some of these words]

[2] Some of the hits may be from other languages in which the same lexical drifts and errors have taken place (possibly for the same reasons), and from non-native writers of English. The number of hits for non-standard spelling is probably artificially reduced by the use of spell-checking software.

[3] 25,200 hits, but about half (based on inspection of the first few dozen) were personal names.

[4] This spelling appears in dictionaries.

[5] Tagalog data throughout this paper are from Schachter & Otanes (1972), English's (1986) dictionary, and my own fieldwork. Where lexical statistics are

given, they are, to avoid bias, calculated only from the words contained in English (1986).

[6] The term 'pseudoaffix' here refers to a string of segments does not form a true morphological affix, but that looks phonologically like an affix (in this case, because it is attached to something that looks phonologically like a reduplicated root). I am agnostic as to whether these pseudoaffixes are treated by speakers as affixes, and whether they derive from historical affixes. The pseudoinfixes present in pseudoreduplicated words (-al-, -aɾ-, -ag-, aʔ-) do, however, resemble productive infixes in related languages.

I have observed one interesting restructuring: pseudoinfixed [h-ag-ulhol] 'sobbing' often appears as pseudo-prefixed [ha-gulɡol], suggesting a preference for proximity between pseudoreduplicant material and pseudobase material.

[7] In a database of 4,390 disyllabic, non-loan roots from English (1986) (roots with pseudoaffixes were excluded), 96 were CV-reduplicated, 256 were CVC-reduplicated, and 47 were ambiguous because the final syllable was either open or closed by a glottal stop, which must delete preconsonantally. How likely is this number of pseudoreduplicated roots to occur if phonemes are combined randomly? To answer the question directly, we should perform every possible reshuffling of the phonemes in the database and determine what how many reshufflings yield that high a number of pseudoreduplicated roots. The number of possible reshufflings is too high, however, to examine them all ($4390^6 \approx 7 \times 10^{22}$). Using Kessler's (2001) approach, a sample of 10,000 reshufflings was taken instead. (Thanks to Brad Kessler for discussion his technique.) The distribution

obtained was approximately normal, with medians well below the actual numbers; even the highest number of pseudoreduplicated roots obtained in 10,000 trials (in bold) was never as high as the actual number. In the case of CVC- reduplication, it never even comes close. As shown in the table below, the results were similar for shuffling onset and rime; onset+nucleus ("head") and coda; and whole syllables. Because the distribution of re-shufflings has been only sampled (though extensively), we cannot say exactly what is the probability that such a large number of pseudoreduplicated words would be obtained through random combination of sounds, but we can safely conclude that $\underline{p}<.01$.

Number of pseudoreduplicated roots obtained

		unit of shuffling				
		phoneme	onset/rime	head/coda	syllable	actual
CV-	highest	**91**	**89**	**90**	**89**	96
	median	59	58	58	58	
	lowest	31	33	33	32	
CVC-	highest	**10**	**10**	**11**	**12**	256
	median	2	2	2	3	
	lowest	0	0	0	0	
ambiguous	highest	**38**	**38**	**37**	**33**	47
	median	17	17	17	16	
	lowest	4	2	4	4	

[8] Two common exceptions are [babáːʔe] ‘woman’ and [laláːki] ‘man’.

[9] If we characterize the alternation as lowering in final syllables, goŋgóŋ displays overapplication; if we characterize the alternation as raising in nonfinal syllables, goŋgóŋ displays underapplication

[10] Nasal assimilation across morpheme boundaries is productive in Tagalog. It is therefore not obvious why underapplication occurs instead of overapplication (see McCarthy & Prince’s 1995 discussion of underapplication). Underapplication is analysable, however, if we let faithfulness to non-preconsonantal nasal place outrank faithfulness to nasal place in general, with the markedness constraint(s) favouring assimilation ranked in between.

[11] Thanks to the associate editor for pointing out the Warlpiri case, and to a reviewer for the Manam case.

[12] Tagalog has just two native suffixes, -in and -an, whose most common and productive function is to form verbs. These suffixes are also used alone and in with prefixes in other morphological constructions. There are some loan suffixes such as Spanish-derived -eɾo and -ista that can combine with native stems. In most suffixal constructions, length (if any) and stress are shifted one syllable to the right (see section 3.3). The [h] that appears when a vowel-final stem is suffixed can be thought of as (i) epenthetic, (ii) part of a postvocalic allomorph of the suffix, or (iii) part of the suffixed allomorph of the stem.

---

[13] Jie Zhang (p.c.) finds that diphthong coalescence is blocked in pseudoreduplicated roots (i.e., [bajbáj] 'edge' cannot be pronounced *[béːbaj]). This would be another case of underapplication like those seen in Section 2.

[14] Occasionally a nonultima mid vowel such as the [o] in góːlpe becomes high under suffixation. I know of no cases in which this happens without the ultima mid vowel's also being raised. That fact lends is consistent with the Aggressive Reduplication analysis of exceptions to vowel raising: although in most of the examples seen here, it will be argued that the stem-ultima vowel resists raising in order to remain similar to the stem-penult vowel, in góːlpe the reverse happens—the stem-penult vowel and stem-ultima vowel remain similar by both being raised. 'Double raising' cases like góːlpe are not included in the statistical analysis because they are too rare. Aggressive Reduplication would predict that double raising, like nonraising, is more likely when the stem ultima and stem penult are more similar.

[15] The behaviour of a stem's derivatives is uniform (all are raised, all vary, or all fail to be raised), so we can speak of stems that are or are not raised, rather individual words that are or are not raised. The cases listed as varying vary not from derivative from derivative but within each derivative (more than one pronunciation is attested).

[16] As mentioned in note 14, some stems do undergo double raising under suffixation, suggesting a variable ranking between Ident-IO(high) and *NonFinalMid.

[17] Because there are only two words with a mid vowel in the penult whose final syllables both have complex onsets (and thus the "same shape" category is overwhelmingly cases with two simple onsets), we might wonder whether the presence of a complex onset somehow encourages raising. As mentioned above in the discussion of markers of foreignness (the foreignness-marking analysis predicts that a complex onset would discourage raising), the presence of a complex syllable margin in general in the dataset has no effect on raising.

[18] Onsets were counted as having identical manner if they were both stops, both affricates, both fricatives, both liquids, both nasals, or both glides. This makes for a small set of 'same' tokens, and thus a low level of significance.

[19] The restructuring of hagulhol to hagulgol (see note 6) might reflect adjacency requirements.

[20] If similarity between the last two consonants in ...V.CVC stems encouraged non-raising, we would have evidence for correspondence between non-syllable strings. As seen in 3.3.2, there are only 26 consonant-final words with a mid vowel in the penult, too few to test this prediction, so it is possible that non-syllable correspondences occur. I leave as a question for future research whether languages like Ilokano that do not respect prosodic identity are more likely to put non-syllable strings in correspondence.

[21] Most of the English examples, moreover, involve modifications to codas rather than to onsets or nuclei. This could reflect the reduced perceptibility of codas (Jun 1995; Steriade 1997, 1999).

---

[22] Some casual data suggest that similar cases of similarity preservation through rule blocking (rather than outright enhancement) may exist in other languages: some English speakers feel that tapping of /d/ is almost obligatory in the proper names Quasimodo and Frodo, but only optional in pseudoreduplicated dodo. Similarly, Zulu allows either light or dark [l], but pseudoreduplicated Lulu requires two light [l]s. In French, [ɔ] is usually found instead of [o] in nonfinal syllables (e.g. [dɔdy] 'chubby'), but is not possible in baby-talk reduplicated words like [dodo] 'beddie-bye' (even though the source word, [dɔʁmiχ] 'to sleep', has [ɔ]). Thanks to Bruce Hayes for the English observations and to Roger Billerey for the French observation.

[23] Some additional unpredictable properties of suffixed stems may justify listing suffixed allomorphs. Syncope, as in /bukas + an/ → [buksán] 'to open', is lexically determined and occurs in a minority of stems, where it is often variable. The consonant clusters created by syncope can undergo further modifications that are also lexically determined, such as metathesis (/tanim + an/ → [ta**mn**án] 'to plant') and hardening (/halik + an/ → [halikán] ~ [halkán] ~ [ha**g**kán] 'to kiss'). At least these irregular stems, then, must have listed suffixal allomorphs.

[24] I have carried out preliminary learning simulations using Boersma's (1998) Gradual Learning Algorithm, as implemented in Hayes, Tesar, and Zuraw (2000). The output of the algorithm is a stochastic grammar: a ranking value is learned for each constraint, and in any actual utterance each constraint is assigned a selection point, which tends to be close to the ranking value. The result is probabilistic variation in the overall constraint ranking: if C_1's ranking value is much higher

---

than C_2's, C_1 will virtually always outrank C_2, but if C_1's ranking value is only somewhat higher than C_2's, C_2 will often outrank C_1.

Because of over- and under-application in productive reduplication for nasal coalescence, nasal assimilation, and glottal deletion (see (4)), the algorithm ranks Ident-BR(place), Ident-BR(nasal), and Max-R high enough that they often outrank Redup in an utterance. Ident-BR[voice], on the other hand, is ranked so low that Redup always outranks it. As a result, the mismatch in voicing between penult and ultima onsets is no impediment to a reduplicative construal (and thus blocking of raising). This is consistent with the facts among actual loanstems in (16).

[25] Lexicalisation of reduplicative behavior can also explain why in Tagalog, some pseudoreduplicated words display stronger reduplicative identity effects than are seen in productive reduplication: tapping does not under- or over-apply in productive reduplication, but can do either in pseudoreduplication (4b). Because the grammar of Tagalog applies tapping transparently, the under- and over-applications must be lexically encoded.

[26] I'm indebted to an anonymous reviewer and to [suppressed for now to maintain anonymity] for suggesting and pressing this line of explanation.

[27] How the child would learn the ranking of *Spec constraints is uncertain. Plausibly, *Spec is ranked at the top of the grammar by default, and demoted only when the winning underlying form in an optimisation fails to generate the desired surface form.

[28] Violations for material present in all candidates are omitted.

[29] In addition, *Spec could be exploded into several differently ranked constraints so that, for example, certain repeated features would be tolerated but not others.

[30] Walker (2000b)

[31] Thanks to Michael Wagner for discussion of this point: the suffixes in question are derivational, not inflectional, so it is not necessary for a new loanstem to be suffixed right away. Moreover, the majority of loanstems do not have a suffixed form in the dictionary, whereas only a handful of loanstems fail to occur unsuffixed.

[32] [suppressed acknowledgements]

[33] The syllable boundaries here are uncertain. Palauan allows codas, but also allows complex onsets, even those of flat sonority. So, the two clusters may or may not both be onsets.

[34] I follow MacEachern in using 'Peruvian Aymara' as a label of convenience for the dialect described in the dictionaries of Ayala Loayza 1988 and Deza Galindo 1989, both published in Peru.

[35] MacEachern presents a full theory of laryngeal similarity. *LaryngealSimilarity is my shorthand for any constraints discouraging laryngeally similar consonants within a morpheme. MacEachern also argues for featural Max and Dep constraints, in part to rule out /t'ata/ → [tata]; Ident-IO(laryngeal) is again a shorthand.

[36] Factorial typologies for the two approaches were calculated using Hayes, Tesar, and Zuraw (2000).

---

[37] Thanks to the anonymous associate editor for pointing out the connection between affix-detection and back-formation.