Phonology **19** (2002) 395–439. © 2002 Cambridge University Press
DOI: 10.1017/S095267570300441X Printed in the United Kingdom

*Aggressive reduplication**

Kie Zuraw
University of California, Los Angeles

I propose that there is a purely phonological drive to impose a reduplication-like structure ('coupling') on words. This structure 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 coupling is, and thus the more likely resistance to raising. I attribute phonologically driven coupling to the activity of a constraint REDUP in generation, which shapes the way new words are lexicalised in the vowel-raising case, but also consider an alternative source for reduplicative construals (the effect of *SPEC in lexical learning). The proposal is compared to others that promote correspondence between similar or identical single segments within a word; I conclude that a relation between strings is necessary.

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 avoids a sequence of fricatives. To give a rough 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)

	hits		*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.[1]

* Thanks to the following people for very helpful comments and discussion on this work at various stages: Adam Albright, Tania Azores-Gunter, Bruce Hayes, Brett Kessler, Donka Minkova, Carson Schütze, Dominique Sportiche, Donca Steriade, Rachel Walker, Colin Wilson, Jie Zhang, Cheryl Zoll, audiences at the 1999 University of Alberta Workshop on the Lexicon, the 2000 LSA meeting in Chicago, MIT's Phonology Circle, UCLA and UC Santa Barbara, three anonymous reviewers and the associate editor.

[1] Thanks to Carson Schütze for *Inuktituk*, Adam Albright for *sancrosant* and *perservere* and Bruce Hayes for *Abu|i Dhabu|i*.

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

(2)

Non-standard	*hits*	*Standard*	*hits*[2]
sherb**er**t[3]	12,000	sherbet	62,900
pompo**m**[4]	15,500	pompon	17,700
orangutan**g**	6,130	orangutan	55,600
orangoutan**g**	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, in *orangutan*/*orangutang*, *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 interpret a word as reduplicated only if it bears the proper morphosyntactic or semantic features. I propose, however, that regardless of morphosyntactic and semantic cues, speakers may give words a reduplication-like structure if they possess sufficient internal phonological self-similarity. I call this tendency 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.

§1 describes a straightforward class of cases, pseudoreduplicated roots in Tagalog, which sporadically behave as though they were truly reduplicated. §2 gives the proposed analysis of aggressive reduplication: a constraint REDUP favours imposing a relation, COUPLING, between strings. Coupling, like morphologically reduplicated structure, invokes correspondence constraints (McCarthy & Prince 1995, 1999) that, depending on the constraint ranking, can enhance self-similarity. Correspondence constraints also regulate which strings are similar enough to be coupled in

[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 artificially reduced by the use of spell-checking software.

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

[4] This spelling appears in dictionaries alongside *pompon*.

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

the first place. §3 presents a subtler case, also from Tagalog, in which aggressive reduplication, rather than enhancing word-internal similarity outright, preserves it by probabilistically blocking vowel raising, a productive alternation, in loan-stems. 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, because fewer correspondence constraints stand in the way of coupling. A partly variable constraint ranking is proposed to account for the probabilistic nature of the similarity effects. §4 considers whether aggressive reduplication could be attributed to the structure-disfavouring constraint *SPEC (Prince & Smolensky 1993) in lexical learning. §5 compares aggressive reduplication to other proposals that promote correspondence or identity between single segments within a word; I conclude that a string-level relation, like coupling, is needed. §6 summarises 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 morpho-syntactic or semantic characteristics of a morphologically reduplicated Tagalog word.

Some examples, selected at random from a database of 464 pseudoreduplicated words gathered from English's (1986) dictionary of Tagalog, are given in (3).[5] The pseudoreduplicated roots are generally of the form C_1V_1-$C_1V_1C_2$ (3a) or $C_1V_1C_2$-$C_1V_1C_2$ (3b), though some pseudoreduplicated words (not illustrated in (3)) also have a medial vowel ([bus**á:**bos] 'slave'), a pseudoprefix ([**gi**puspós] 'very low-spirited') or, most frequently, a pseudoinfix ([p**al**igpíg] 'shaking off water').[6] Misidentities of vowel height, glottal stop presence and nasal place of articulation are discussed below.

[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 terms 'pseudoprefix' and 'pseudoinfix' here refer to a string of segments that 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-], [-ar-], [-ag-], [-aʔ-]) do, however, resemble productive infixes in related languages.

I have observed one interesting restructuring: pseudoinfixed [h-ag-ulhol] 'sobbing' is often pronounced and written as pseudoprefixed [ha-gul**g**ol], suggesting a preference for proximity between pseudoreduplicant material and pseudobase material.

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

(3) a. lulód ‘shin (N)’
tá:tal ‘wood chips, splinters, shavings (N)’
sú:soʔ ‘snail (N)’
lí:liw ‘bird sp. (N)’
pú:pog ‘pecking hard, repeated kissing (N)’
lulón ‘swallowing (N)’
laláʔ ‘acute (A)’
hí:hip ‘blow, puff (N)’
tá:taŋ ‘daddy (N)’
nú:noʔ ‘ancestor (N)’

b. ʔagʔág, ʔagág ‘sifted as grain or powder (A)’
ŋikŋík ‘noise made by pigs waiting to be fed (N)’
mismís ‘remnants of food left after a meal (N)’
wigwíg ‘sprinkling of water during ironing (N)’
bunbón, bumbón ‘dam for attracting fish, clear pond (N)’
dasdás ‘planing or sanding wood or cane (N)’
gajgáj ‘travelling around (N)’, ‘reached by travelling around (A)’
patpát ‘stick, piece of split bamboo (N)’
sagság ‘split, blunt, sagging, at the peak of success (A)’
ŋasŋás ‘scandal, excessive garrulousness and gesticulating (N)’

The pseudoreduplicated words are probably not accidentally so. There are far more pseudoreduplicated words than would be expected through random phoneme combination (see the Appendix), and the pseudoreduplicated roots are phonologically exceptional in two ways. First, it is rare to find two occurrences of the same consonant within a root 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 non-existent root-internally ([du**td**ót] ‘poking’).

Still, there are several reasons to say that (synchronically, at least) these roots are not morphologically reduplicated, only pseudoreduplicated. First, in Tagalog the minimal root is disyllabic – the only monosyllabic roots are clitics and loans – so if the words in (3) were reduplicated, it would be from a too small root (e.g. *[bak]). Pseudoreduplication might be a way to augment just such too small roots, but there are multiple pseudoreduplicating patterns (CV-, CVC- and the medial-vowel, pseudoprefixed and pseudoinfixed forms) and some unpredictability of stress (in the CV- and medial-vowel roots), so a monosyllabic lexical entry would still have to specify exactly how it is to be augmented; reduplication could not just be 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

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

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 would occur almost exclusively with monosyllabic roots (i.e. there are few uninflected words like *[babagid]).[7]

Whether or not the pseudoreduplicated roots are historically related, 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 & Prince (1995) propose correspondence between segments of the base and segments of the reduplicant as the reason for identity effects in morphological reduplication. In the terms of Wilbur (1973) and McCarthy & Prince (1995), TRANSPARENT 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 OVERAPPLICATION only the base or the reduplicant is in the expected environment for a rule, but the rule applies to both; and in UNDERAPPLICATION only the base or the reduplicant is in the expected environment for a rule, and 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. (4) illustrates five types of example, with the number of roots that do not exhibit identity effects and the number that do given in parentheses. For example, Tagalog nasal coalescence famously overapplies in morphological 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-hín] 'to inflame', /paN+RED_{CV}+pula/→[pa-**m**u-**m**ulá] 'ruddiness', /maN+$\text{RED}_{\sigma\sigma}$+pula/→[ma-**m**ulá-**m**ulá] 'to glow' and even /naN+RED_{CV}+$\text{RED}_{\sigma\sigma}$+pula/→[na-**m**ù:-**m**ulá-**m**ulá] 'reddish'). In most of the pseudoreduplicated words, nasal coalescence applies transparently, but in some it overapplies, as though the words were morphologically reduplicated (4a).

[7] Two common exceptions are [babá:ʔe] 'woman' and [lalá:ki] 'man'.

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

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

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

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

transparent:	some pseudoreduplicated words (7) **d**íːɾi ‘loathing’
overapplies:	some pseudoreduplicated words (2) **ɾ**úːɾok ‘acme’
underapplies:	some pseudoreduplicated words (1) **d**eː**d**e ‘baby bottle’

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

transparent:	most pseudoreduplicated words (142) d**u**bd**ó**b ‘feeding a fire’
identity:[8]	some pseudoreduplicated words (12) g**o**ŋg**ó**ŋ ‘gruntfish’

d. *nasal place assimilation*

transparent:	some pseudoreduplicated words (6+8 variable) da**n**dáŋ ‘toasting’
underapplies:[9]	some pseudoreduplicated words (3+8 variable) di**ŋ**díŋ ‘wall’

e. *postconsonantal glottal deletion* (*Cʔ)

transparent:	some pseudoreduplicated words (1+8 variable) **ʔ**utót ‘flatulence’
underapplies:	some pseudoreduplicated words (5+8 variable) **ʔ**ig**ʔ**íg ‘shaking’

McCarthy & 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 with reduplicative structure, to show that such structure forces both glottal stops to be retained (because Dep-BR≫*Cʔ). (What would require the

[8] If we characterise the alternation as lowering in final syllables, [goŋgóŋ] displays overapplication; if we characterise the alternation as raising in non-final syllables, it displays underapplication.

[9] 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.

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

reduplicative structure is addressed below in §2.) Overapplication of glottal deletion is impossible because of a language-wide prohibition on vowel-initial words, *[V. The tableau is somewhat misleading, however, in that the behaviour of pseudoreduplicated stems is lexically determined: some stems undergo glottal deletion, some do not and some vary (and similarly for the other four phenomena listed). §3.3.1 addresses the question of how such behaviour becomes lexically encoded.

(5)

	/ʔigʔig/	*[V	Dep-BR	*Cʔ	Max-IO
underapplication	☞ a. $[ʔig]_R[ʔig]_B$			*	
transparent application	b. $[ʔig]_R[ig]_B$		*!		*
overapplication	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 occur 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.[10] Warlpiri has many words that Nash (1980: 118–129) calls 'lexically reduplicated'. 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 morphologically reduplicated (for example, *wiinywiinypa* 'grey falcon' and *piirrpiirrpa* 'half white' escape 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 1998), 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 last 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 McCarthy & Prince's (1995) Integrity. The reduplicant is therefore kept as small as possible to minimise further violations of Integrity. This means that [ragógo] is being treated by the phonology as though it has reduplicative structure, assuming that reduplication is multiple correspondence.

[10] Thanks to the associate editor for pointing out the Warlpiri and Chaha (see below) cases, and to a reviewer for the Manam case.

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

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

Analysts have ascribed reduplicative structure to pseudoreduplicated words in other languages that do not qualify morphosyntactically or semantically for a reduplicative analysis. Nash (1980) attributes reduplicative structure to the Warlpiri lexically reduplicated words discussed above. Buckley (1998) proposes that Manam pseudoreduplicated words have an underlying RED morpheme, and Golston & 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. McCarthy (1981) argues that Arabic (and other Semitic) roots of the form XYY or XYXY are derived from underlying XY by reduplication-like spreading. This accounts for the overapplication of palatalisation and labialisation to phonologically reduplicated roots in Chaha (McCarthy 1983, Banksira 2000).

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 cases of over- and underapplication are holdovers from a proto-language in which they were morphologically reduplicated.

§3 discusses a more widespread case of underapplication in Tagalog: vowel raising in loanwords, where self-similarity gradiently blocks a productive alternation. There, 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. But first, §2 proposes an analysis of aggressive reduplication.

2 Analysis of aggressive reduplication

This section proposes that there is a constraint in the grammar that prefers output forms to have a reduplication-like structure. I call this constraint REDUP (cf. Hammond's 1999 SUFFIX, which promotes treating words as suffixed). §4 considers an alternative, that reduplicative construals occur during lexical learning and are driven by *SPEC (Prince & Smolensky 1993). Before defining REDUP, I will briefly review and then extend the

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

view of productive, morphological reduplication proposed by McCarthy & Prince (1995, 1999).

McCarthy & Prince (1995, 1999) propose that there is a relation, correspondence, between segments of the reduplicant and segments of the base. Correspondence does not itself require similarity – that is governed by violable constraints. MAX-S_1S_2 and DEP-S_1S_2 are constraints that require segments of strings S_1 and S_2 to participate in the relation, and the IDENT-S_1S_2 family requires any participating segment to be similar to its correspondent. MAX-BR, DEP-BR and IDENT-BR (I will abbreviate the set as CORR-BR) are versions of these constraints that apply when S_2 is a reduplicant and S_1 is its base. McCarthy & Prince's constraint definitions could be rephrased as in (6).

(6) a. MAX-BR
If a word contains a portion labelled as a reduplicant and a portion labelled as the reduplicant's base, then every segment in the reduplicant must have a correspondent in the base.

b. DEP-BR (equivalent to MAX-RB)
If a word contains a portion labelled as a reduplicant and a portion labelled as the reduplicant's base, then every segment in the base must have a correspondent in the reduplicant.

c. IDENT-BR[F]
If a word contains a portion labelled as a reduplicant and a portion labelled as the reduplicant's base, and if a segment in the base corresponds to a segment in the reduplicant, the two segments must have the same value for the feature [F].

It is the presence of a RED morpheme in the input that requires morphological labelling of one part of the word as 'base' and another part as 'reduplicant'. We could schematise the chain of causation as in (7).

(7) RED morpheme *requires* morphological labelling of output
invokes
MAX-BR, DEP-BR *require* correspondence
invokes
ID-BR[F] *requires* similarity

I propose that this scheme be generalised so that the base–reduplicant relationship is a special, morphologically specified instance of a more general relation that can hold between substrings of a word, which I call COUPLING. Coupling is a phonological property, and regardless of the

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

morphology, Gen freely generates candidates with and without coupling. The coupling relation makes it possible to define more general correspondence constraints, in (8), that do not require reduplicative morphology. (No distinction is made here between MAX and DEP, because without reduplicative morphology, the two coupled substrings have equal status, unlike B and R.)

(8) a. MAX-κκ
If a word contains two substrings S_1 and S_2 that are coupled, then every segment in S_1 must have a correspondent in S_2 and *vice versa*.

b. IDENT-κκ[F]
If a word contains two substrings S_1 and S_2 that are coupled, and if a segment in S_1 corresponds to a segment in S_2, the two segments must have the same value for the feature [F].

('κ' stands for 'coupled string', not for any Greek word; 'c' is avoided because of possible confusion with 'consonant'.)

The CORR-BR constraints can be thought of as morpheme-specific instances of the CORR-κκ constraints that apply when one of the coupled strings is tagged as RED and the other as its base.[11] (A reduplicant and its base can always be coupled, because RED has no underlying phonological material to stand in the way.)

If the morphological tagging that invokes CORR-BR is required by a RED morpheme in the input, what requires the coupling relation that invokes CORR-κκ? I propose a constraint, REDUP, that requires all words to contain coupled substrings. Just as MAX is insensitive to the similarity of segments it requires to be in correspondence, REDUP is insensitive to the similarity of substrings it requires to be coupled. It is up to other constraints to regulate similarity. The diagram in (9), parallel to (7), schematises the chain of causation. REDUP, by putting substrings into a coupling relation, invokes MAX-κκ in much the same way that MAX, by putting segments into correspondence, invokes IDENT[F].

(9) REDUP constraint *requires* coupling of substrings in output
↓ *invokes*
MAX-κκ *requires* correspondence
↓ *invokes*
ID-κκ[F] *requires* similarity

[11] Thanks to an anonymous reviewer for suggesting this.

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

There are many ways to formulate REDUP that are consistent with the data discussed in this paper. The simplest definition, in (10), is the one that will be used in the tableaux below.

(10) REDUP

A word must contain some substrings that are coupled.

For example, ka[ba]$_\alpha$[da]$_\alpha$, where the shared Greek-letter subscript indicates coupling, satisfies REDUP, because it contains the coupled substrings *ba* and *da*. (The two substrings' dissimilarity is irrelevant to REDUP.) [kab]$_\alpha$[ada]$_\alpha$ satisfies REDUP equally well, though its violations of CORR-$\kappa\kappa$ would be severe.

Some other possible definitions of REDUP are (i) that it requires coupling between any two syllables or feet of a word (with a violation for every syllable or foot that is left out, or a violation for every pair that is not coupled), and (ii) that it requires a word to be completely partitioned into two coupled substrings (with a violation if the partitioning is not complete, or a violation for every segment that is left out). Although the definition in (10) will be used below, there is a tendency in the data discussed in this paper for coupled strings to be adjacent syllables. I leave it to future research to determine whether this tendency should be accounted for by the definition of REDUP, or by the same violable constraints that, in morphological reduplication, promote syllable- or foot-sized reduplicants, reduplicant–base adjacency and identity of prosodic roles between corresponding segments. In Tagalog, morphological reduplicants are always one or two syllables and adjacent to the base, and corresponding segments of B and R always have the same prosodic role, so these constraints should be ranked high.

The CORR-$\kappa\kappa$ constraints interact with REDUP, because REDUP requires substrings to be coupled, and CORR-$\kappa\kappa$ requires those coupled strings to be similar. REDUP also interacts with the input–output correspondence constraints (abbreviated CORR-IO), which disprefer candidates whose underlying material has been changed to achieve greater self-similarity. The CORR-$\kappa\kappa$ and CORR-IO constraints interact with REDUP to (i) restrict which strings can be coupled, and (ii) enhance the similarity of coupled strings. The schematic factorial typology in (11) illustrates the interaction (jagged lines between columns indicate that all rankings are to be considered). The three possible outcomes are coupling despite imperfect similarity ([tag]$_\alpha$[dag]$_\alpha$), coupling with enhancement of internal similarity ([tag]$_\alpha$[tag]$_\alpha$) and lack of coupling ([tagdag]).

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

(11) *Interaction of* REDUP *with* CORR-IO *and* CORR-$\kappa\kappa$

a. REDUP, IDENT-IO[voice] ≫ IDENT-$\kappa\kappa$[voice]
coupling despite imperfect similarity

b. REDUP, IDENT-$\kappa\kappa$[voice] ≫ IDENT-IO[voice]
voicing difference is 'repaired'

c. IDENT-IO[voice], IDENT-$\kappa\kappa$[voice] ≫ REDUP
no coupling

/tagdag/	REDUP	ID-IO[vce]	ID-$\kappa\kappa$[vce]
a. $[tag]_\alpha[dag]_\alpha$			*
b. $[tag]_\alpha[tag]_\alpha$		*	
c. tagdag	*		

Notice how the CORR-$\kappa\kappa$ constraint cuts both ways: it can work with REDUP to increase similarity between coupled strings, or work with CORR-IO to prevent coupling. The interaction holds not just for IDENT constraints but also for MAX, which has the effect of requiring similarity of syllable shape:

(12)

/trata/	REDUP	MAX-IO	MAX-$\kappa\kappa$
a. $[tra]_\alpha[ta]_\alpha$			*
b. $[ta]_\alpha[ta]_\alpha$		*	
c. trata	*		

Because there are many CORR-$\kappa\kappa$ 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 discussed below. The distinction will become apparent there between candidates like (11a) or (12a), with coupling but no enhancement of similarity, and (11c) or (12c), with no coupling: even without enhancement of similarity, the coupling makes itself known there by blocking an alternation.

Before moving on to the raising case, a further note on typology. Proposing the constraint REDUP might seem to predict that there should exist languages in which all words are reduplicated (this is predicted if REDUP and all CORR-$\kappa\kappa$ constraints outrank all CORR-IO constraints). 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. This typological overprediction is not unique to REDUP, but is shared by many contrast-reducing constraints. For example, the silent language, in which the structure-banning constraint *STRUC (Zoll 1993) dominates all faithfulness constraints, does not exist. Similarly,

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

Prince & Smolensky (1993) propose constraints of the form *PEAK/X that forbid X as a syllable nucleus (the less sonorous X is, the more marked it is a nucleus: *PEAK/[t] ≫ *PEAK/[n] ≫ *PEAK/[u] ≫ *PEAK/[a]). But there is no language in which all the *PEAK/X except *PEAK/[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 a 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. Although the (presumably extralinguistic) mechanism that excludes pathological grammars is not understood, such a mechanism is needed to prevent many other contrast-reducing constraints from being at the top of the grammar.[12]

To summarise, REDUP requires all words to contain coupled substrings. Like a reduplicative morpheme, coupling invokes correspondence constraints that enhance similarity between the coupled substrings. In most words, the coupling requirement is overridden by the CORR-$\kappa\kappa$ violations that would occur if dissimilar strings corresponded, or by the CORR-IO violations that would occur if underlyingly dissimilar strings were made more similar to satisfy CORR-$\kappa\kappa$. But if two substrings of a word are sufficiently similar underlyingly, REDUP can be satisfied at little cost to CORR-IO and CORR-$\kappa\kappa$, and reduplicative over- and underapplication can result. In morphologically reduplicated words, of course, there is no obstacle at all to satisfying REDUP, because the reduplicant has no underlying material of its own.

3 Vowel raising in Tagalog loanwords

The English and Tagalog data discussed so far provide sporadic examples of aggressive reduplication. This section presents a more systematic case, with enough instances for statistical information to be meaningful. §3.1 gives the basic facts on vowel height, §3.2 examines the distribution of vowel raising in loans and gives an aggressive reduplication account, §3.3 gives a partly variable constraint ranking to account for the probabilistic nature of the pattern and §3.4 rejects some other possible accounts of the distribution of raising.

3.1 Vowel height: basic facts

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-ultimas. [i] can occur anywhere, and many

[12] This issue is not limited to OT. In a rule-based framework where languages select from a universal set of rules, extralinguistic factors must explain limits on deletion and neutralisation rules within adult languages.

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

words have [i] and [e] in free variation in the ultima. Typical monomorphemic native words are shown in (13).

(13) b**ú:**ko 'young coconut, flower bud'
b**i**gát 'burden'
bí:l**o**g 'circle'
dá:l**e**, dá:l**i** 'denunciation'
bú:k**i**d 'farm'

Suffixation induces alternation, by making final syllables non-final:[13]

(14) ká:l**o**s 'grain leveller' kal**ú:**s-in 'to use a grain leveller on'
ʔab**ó** 'ash' ʔab**u**-hín 'to clean with ashes'
babá:ʔ**e** 'woman' ka-babaʔ**í:**-han 'womanhood'
sist**é** 'joke' sist**i**-hín 'to joke' (Spanish loan)

Besides two classes of systematic exceptions to the generalisation that mid vowels are found only in ultimas,[14] 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 (15a, b), words whose ultima vowels remain mid under suffixation (15c) and words with ultima [u] (15d). 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). Note that many of the words with non-ultima mid vowels appear to have CV- or CVC-pseudoreduplication (15a), and that the words that fail to be raised under suffixation have a penult mid vowel of the same backness as the ultima mid vowel (15c). These facts will be relevant below in explaining the distribution of exceptions.

[13] 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 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, vowel length (if any) and stress are shifted one syllable to the right. 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.

[14] 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] ~ [ʔé:wan] 'I don't know' and [kaʔuntíʔ] ~ [kawntíʔ] ~ [kó:ntiʔ] 'a little'. Jie Zhang (personal communication) finds that diphthong coalescence is blocked in pseudoreduplicated roots (i.e. [bajbáj] 'edge' cannot be pronounced *[bé:baj]). This would be another case of underapplication like those seen in §2.

The second systematic source of non-ultima mid vowels is VʔV 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').

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

(15) a. ʔ**óː**ʔo ‘yes’
g**o**ŋgóŋ ‘gruntfish’
d**éː**de ‘baby bottle’
k**e**ŋkéŋ ‘sound made by beating frying pan’
(and 10 more)[15]

b. b**o**hól ‘shrub sp.’
ʔ**óː**la ‘eagerness’
k**éː**rwe ‘cricket’
h**éː**to ~ ʔ**éː**to ‘Here it is!’
(and 10 more)

c. déːd**e** ‘baby bottle’ paded**éː**-hin ‘give a baby a bottle’
totoʔ**ó** ‘true’ totoʔ**óː**-hin ‘to be sincere’
poʔ**ó**t ‘hatred’ ka-poʔ**o**t-án ‘to hate’
(and all other [oʔo] words – see note 15)

d. samp**ú**ʔ ‘ten’
dáːtoʔ ~ dáːt**u**ʔ ‘chieftain’ (?from Malay *datu*)
labíw ~ labj**ú** ‘weeds that grow in a burned field’
(and 4 more)

As in the native vocabulary, there are exceptions of each kind to vowel-height restrictions 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:

(16) b**éː**nta ‘sales’ (from Spanish *venta*)
k**o**ɾék ‘correct’ (from English *correct*)
ʔas**ú**l ‘blue’ (from Spanish *azul*)

The majority (79%, based on the corpus discussed below) of loanword stems with mid-vowel ultimas are raised under suffixation (17a), indicating that the alternation is productive, but a substantial minority (16%; 5% vary) fail to be raised (17b):

(17) a. sab**ó**n ‘soap’ sab**u**n-án ‘to put soap on’
atáːk**e** ‘attack’ atak**íː**-hin ‘to attack (OBJ FOCUS)’
góːlp**e** ‘hit’ g**u**lp**i**-hín ‘to hit (OBJ FOCUS)’[16]

b. káːbl**e** ‘cable’ kabl**e**-hán ‘to send a cable to’
mag-manéːh**o** ‘to drive (ACTOR FOCUS)’ manehó**ː**-hin ‘to drive (OBJ FOCUS)’

[15] Among all of the exceptions in a database of the 4390 disyllabic, native roots in English’s (1986) dictionary, and relevant longer native words that I have encountered.

[16] Occasionally a non-ultima 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. The fact is consistent with the aggressive reduplication analysis of exceptions to vowel resists raising: although in most of the examples seen here, it

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

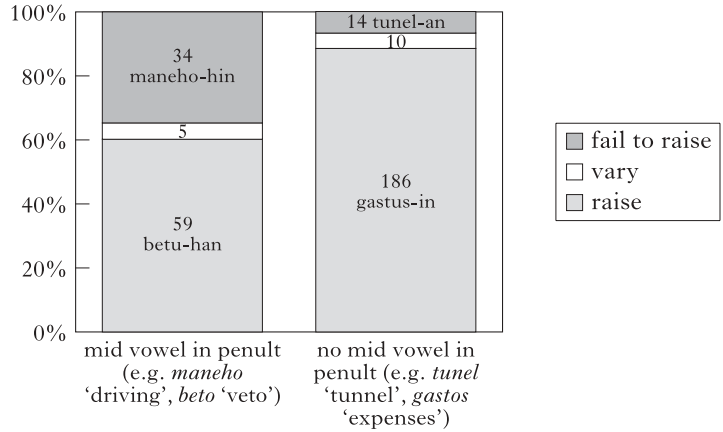


Figure 1
Effect of mid vowel in penult on probability of raising.

3.2 Aggressive reduplication and raising in loans

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. This section argues that aggressive reduplication influences the distribution of non-raising among loans.

To examine the distribution of raising and non-raising, I constructed a database from English's (1986) Tagalog dictionary of all 306 Spanish and English loan-stems with a mid vowel in the ultima and one or more listed suffixed derivatives (the total number of suffixed forms is 488). As observed by Schachter & Otanes (1972), the best predictor that a loanword stem will fail to alternate is the presence of a mid vowel in another syllable. The graph in Fig. 1 shows that only 7% of stems with a low- or high-vowel penult fail to be raised (like [t**u**nel-an] 'to tunnel'),[17] 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 $p<\cdot0001$, using Fisher's Exact Test

will be argued that the stem-ultima vowel resists raising in order to remain similar to the stem-penult vowel, in [gó:lpe] the reverse happens – the stem-penult vowel and stem-ultima vowel remain similar by both being raised. 'Double raising' cases like [gó:lpe] are not included in the statistical analysis because they are too rare. Aggressive reduplication would predict that double raising, like non-raising, is more likely when the stem ultima and stem penult are more similar.

[17] 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 than individual words that are or are not raised. For that reason, multiple derivatives of a single stem are collapsed in the bar charts below. The cases listed as varying vary not from derivative to derivative, but within each derivative (more than one pronunciation is attested).

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

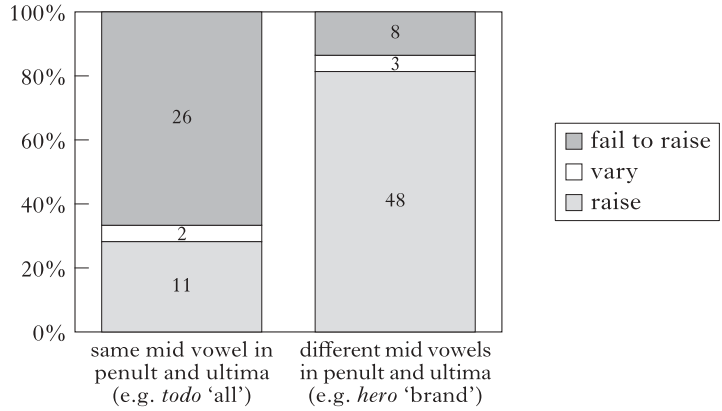


Figure 2
Effect of matching backness between penult and ultima, given a mid penult.

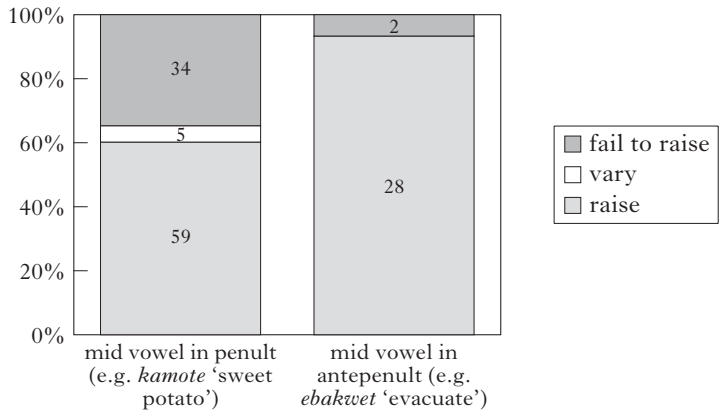


Figure 3
Effect of proximity.

(variable stems were omitted for all applications of Fisher's Exact Test; all p-values below are also from Fisher's Exact Test).

As shown in Fig. 2, given a mid vowel in the penult, the likelihood of non-raising is greatly enhanced if the two mid vowels match in backness ($p<\cdot0001$).

Proximity also has an effect, as shown in Fig. 3. A mid vowel in the penult strongly encourages non-raising, as compared to a mid vowel in the antepenult ($p<\cdot005$), which in turn has no effect when compared to no mid vowel at all.

To summarise, a vowel in the stem's ultima is less likely to be raised if there is another mid vowel in the immediately preceding syllable. The

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

effect is stronger if the two mid vowels match in backness. I claim that a preceding mid vowel impedes raising of the stem-ultima vowel because a coupling relation has been imposed between the last two syllables of the stem. Raising the ultima vowel would therefore violate IDENT-ĸĸ[high].

The tableau in (18) shows the analysis of vowel height in general (preservation of exceptional non-final mid vowels, as in [heto], and raising of final vowels, as in [kalusin]). IDENT-IO[high], which could easily be replaced with output–output correspondence to the unsuffixed form, has been separated into stem-ultima and non-stem-ultima versions (the ranking is revised somewhat in §3.3).

(18)

	/heto/	ID-IO[hi] (non-stem-ult)	*NON-ULT[mid]	ID-IO[hi] (stem-ult)
a.	☞ i. heto		*	
	ii. hito	*!		
b.	/kalos+in/			
	☞ i. kalusin			*
	ii. kalosin		*!	

The tableau in (19) for /todo+in/ 'to include all' illustrates how raising is blocked if there is coupling. Candidate (b) fails because the vowels in the coupled strings differ in height; (c) makes the vowels identical, but at the expense of changing the height of the non-stem-ultima vowel. Similarly, (d) makes the consonants identical at the expense of changing underlying [voice], and (e) and (f) fail because they lack coupling. Even though self-similarity is not enhanced, as in (c) or (d) (to do so would violate CORR-IO), REDUP emerges to preserve similarity by blocking raising.[18]

(19)

/todo+in/	ID-IO [vce]	ID-IO[hi] (non-stem-ult)	ID-ĸĸ [hi]	REDUP	*NON-ULT [mid]	ID-IO[hi] (stem-ult)
☞ a. [to]$_\alpha$[do]$_\alpha$hin					**	
b. [to]$_\alpha$[du]$_\alpha$hin			*!		*	*
c. [tu]$_\alpha$[du]$_\alpha$hin		*!				*
d. [to]$_\alpha$[to]$_\alpha$hin	*!				**	
e. toduhin				*!	*	*
f. todohin				*!	**	

In (20), we see that aggressive reduplication explains the matching-backness effect straightforwardly. If IDENT-ĸĸ [back] is ranked high, then

[18] As mentioned in note 16, some stems do undergo double raising (overapplication) under suffixation, suggesting a variable ranking between IDENT-IO[high](non-stem-ultima) and *NON-ULTIMA[mid].

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

a word like [tó:do] can have coupling, but a word like [hé:ɾo] cannot, and so *NON-ULTIMA[mid] causes raising.

(20) *The matching-backness effect*

		ID-IO [bk]	ID-ĸĸ [bk]	ID-ĸĸ [hi]	REDUP	*NON-ULT [mid]
a.	/todo+in/					
	☞ i. $[to]_{\alpha}[do]_{\alpha}$hin					**
	ii. $[to]_{\alpha}[du]_{\alpha}$hin			*!		*
	iii. toduhin				*!	*
	iv. todohin				*!	**
b.	/heɾo+in/					
	i. $[he]_{\alpha}[ɾo]_{\alpha}$hin		*!			**
	ii. $[ho]_{\alpha}[ɾo]_{\alpha}$hin	*!				**
	iii. $[he]_{\alpha}[ɾu]_{\alpha}$hin		*!	*!		*
	☞ iv. heɾuhin				*	*
	v. heɾohin				*	**!

There is no ranking of these constraints that does the reverse, selecting (a.iii) for /todo+in/, but (b.i) for /heɾo+in/ 'to brand'. Assuming variable constraint ranking (see §3.3 below), a word with non-backness-matching vowels can undergo raising, but not as often as an equivalent word with backness-matching vowels.

As for proximity, whatever constraints prefer morphological reduplicants to be adjacent to their bases (perhaps alignment, though this issue is unsettled) would also disprefer coupling between the ultima and an antepenultimate syllable.

There may be other ways to account for the mid-vowel, matching-backness and proximity effects (see §3.4). But aggressive reduplication makes additional predictions. In words that have only partial internal similarity, REDUP is in competition with CORR-ĸĸ constraints. Coupling should therefore be easier to impose between the penult and the ultima when they are more similar along any dimension – not just vowel backness – because fewer CORR-ĸĸ constraints are violated. Therefore, greater similarity in various aspects of the penults and ultimas of loan-stems, not just in the vowels themselves, should encourage non-raising.

For example, matching place of articulation in syllable onsets could encourage non-raising. This works the same way as the matching-backness effect: correspondence between [to] and [no] does not violate IDENT-ĸĸ[place], but correspondence between [bo] and [no] does. The graph in Fig. 4 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 ($p<\cdot 05$). Note that Fig. 4, like the other charts in this section, compares stems whose penult and ultima are similar

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

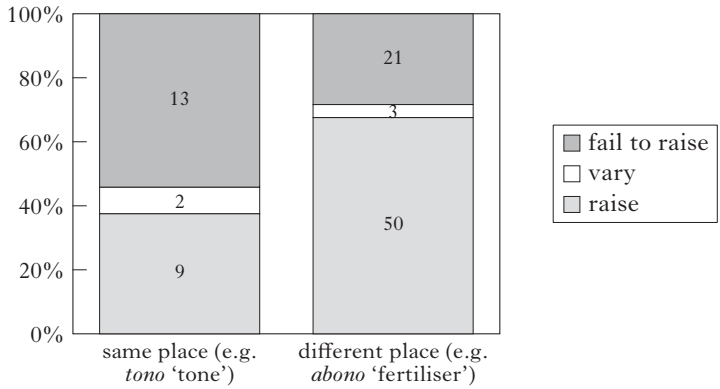


Figure 4
Effect of onset place on rate of raising.

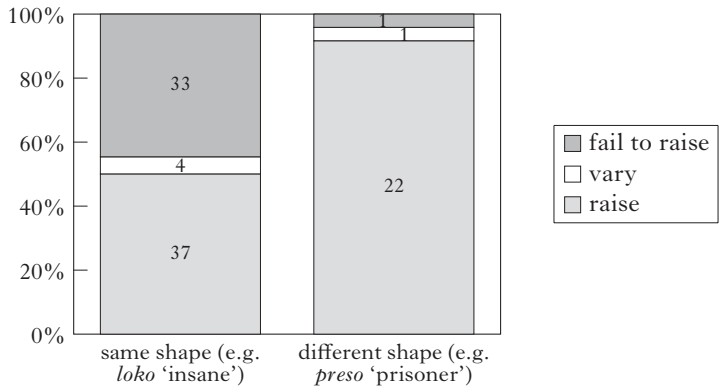


Figure 5
Effect of onset shape on rate of raising.

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 /tono/ must be identical in major place (labial *vs*. coronal *vs*. velar *vs*. glottal), 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.

As illustrated below, two other types of similarity – onset shape and rhyme shape – also significantly encourage non-raising. Two additional dimensions were considered that did not have any significant effect: onset manner and onset voicing.

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

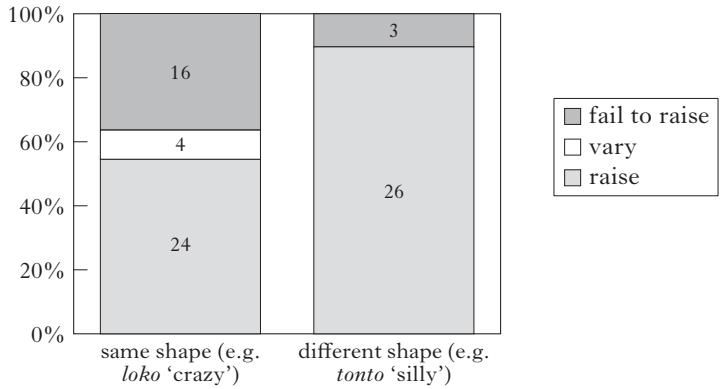


Figure 6
Effect of rhyme shape on rate of raising.

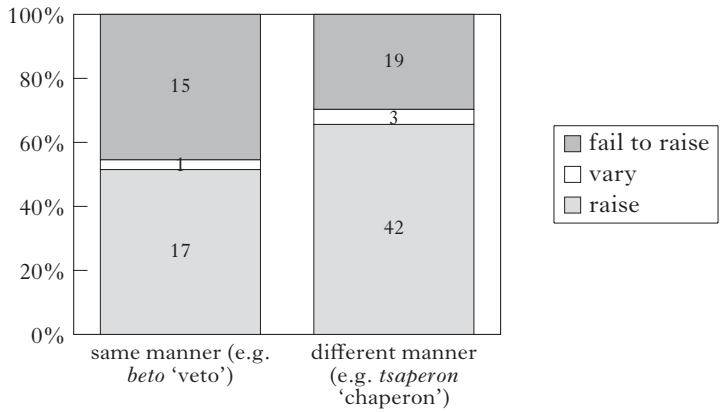


Figure 7
Effect of onset manner on rate of raising.

When onsets match in shape (simple *vs.* complex), non-raising is encouraged ($p<\cdot0001$), as shown in Fig. 5.[19] Here the crucial constraint is

[19] Because there are only two stems 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 in the discussion in §3.4 of markers of foreignness (the foreignness-marking analysis predicts the opposite, that a complex onset would discourage raising), the presence of a complex syllable margin in general in the dataset (that is, including stems where the penult is not mid and looking at all syllables, not just the penult and ultima) has no effect on raising.

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

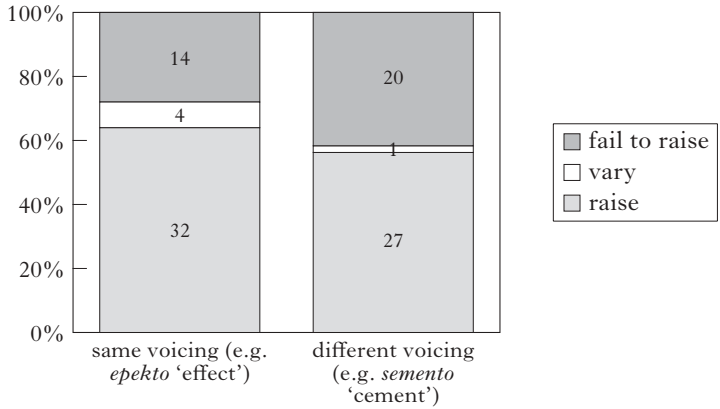


Figure 8
Effect of onset voicing on rate of raising.

MAX-*κκ*: 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 rhyme shape (open *vs.* closed), and again a match promotes non-raising ($p<\cdot01$), because coupling is possible without violating MAX-*κκ*. In $[\text{ton}]_\alpha[\text{to}]_\alpha$, with non-matching rhymes, there is no correspondent for the [n] of the first syllable, violating MAX-*κκ*. Another possibility is $[\text{to}]_\alpha\text{n}[\text{to}]_\alpha$, which would be penalised by constraints requiring adjacency of coupled strings, an issue that does not arise when the rhyme shapes match, as in $[\text{lo}]_\alpha[\text{ko}]_\alpha$.

As discussed in more detail in §4.1, suffixation opens a final closed syllable, which alters the similarity properties of the final two syllables' rhymes. Therefore, the data in Fig. 6 are restricted to words with an open ultima, so that the characterisation of a stem's penult and ultima as 'same' or 'different' with respect to rhyme shape does not change under suffixation. Full data are given in Table II.

As shown in Fig. 7, identical onset manner[20] does not significantly encourage non-raising (though the effect is in that direction): $p = \cdot17$, while Fig. 8 shows that voicing is the one property examined that does not pattern in the predicted direction ($p = \cdot28$). 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 to do so.

Table I summarises the effects of each dimension of similarity.

[20] 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.

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

Dimension	% non-raising stems when similar	% non-raising stems when dissimilar	p	
matching backness	65%	11%	<·0001	
matching onset place	57%	33%	<·05	simple onsets only
matching onset manner	50%	36%	=·17	simple onsets only
matching onset voicing	40%	42%	=·28	simple onsets only
matching onset shape	40%	12%	<·0001	
matching rhyme shape	40%	22%	<·05	

Table I
Effect on raising of similarity between penult and ultima of stem.

The aggressive reduplication analysis can be summarised thus: if there is a mid vowel in the penult, coupling tends to discourage raising because raising would violate IDENT-*KK*[high]. Coupling is more likely when the final syllable is similar to the penult, because few CORR-*KK* constraints are violated.

Are there other opportunities for similarity enhancement or preservation? In addition to vowel raising, suffixation can induce tapping in /d/-final stems. Loans might be expected to undergo tapping more often when it would increase self-similarity, and less often when it would decrease self-similarity. I have found few loan-stems ending in /d/ that have a suffixed form, and no exceptions to tapping under suffixation, however, so we cannot examine the distribution of tapping.

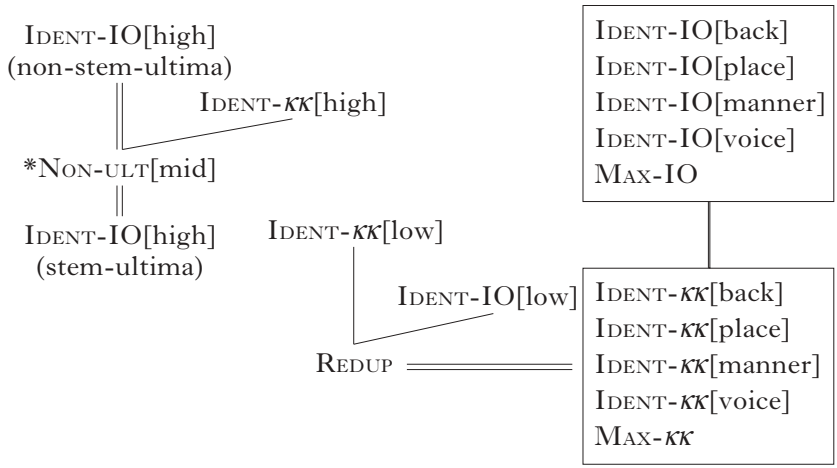
3.3 Accounting for variation

The distribution of non-raising, which I take to reflect the distribution of coupling, is probabilistic. No one dimension of similarity acts as a rigid cut-off. I assume, therefore, that certain constraints are variably ranked. This section details those rankings and addresses the distinction between free variation – where every word's pronunciation is variable – and lexical variation, where most individual words' pronunciations are fixed.

The constraints used are showed in (21). Single lines indicate that the upper constraint (or group of constraints shown in one of the two boxes) dominates the lower constraint (or group). Double lines indicate variable ranking. The double lines between REDUP and the box of five CORR-*KK* constraints should be interpreted to mean that REDUP is variably ranked with respect to any of those CORR-*KK* constraints, not that the CORR-*KK* constraints can only be ranked either above or below REDUP as a block.

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

(21)



First, we can establish the variable ranking between IDENT-IO[high] (stem-ultima) and *NON-ULTIMA[mid]. Even when the penult vowel is high or low, raising behaviour varies. Especially in the case of a high penult vowel, non-raising cannot be attributed to coupling, so it must be attributed to IDENT-IO[high](stem-ultima) (jagged lines indicate variable ranking). In (22), the only way for candidate (c) to win is for IDENT-IO[high](stem-ultima) to outrank *NON-ULTIMA[mid].

(22)

/CuCe+in/	ID-ʁʁ[hi]	REDUP	ID-ʁʁ[bk]	*NON-ULT [mid]	ID-IO[hi] (stem-ult)
a. $[Cu]_{\alpha}[Ce]_{\alpha}$hin	*!		*	*	
☞ b. $[Cu]_{\alpha}[Ci]_{\alpha}$hin			*		*
☞ c. CuCehin		*		*	
☞ d. CuCihin		*			*

Having a low vowel in the penult does not inhibit raising, although we might expect it to because of the shared [−high] feature of low and mid vowels. There are two solutions we could adopt: (i) use a three-height scale of unary features [low], [mid] and [high], so that low and mid vowels do not share any height feature that an IDENT-ʁʁ constraint could refer to, or (ii) let IDENT-ʁʁ[low] and IDENT-IO[low] outrank REDUP, so that low and mid vowels cannot correspond. I arbitrarily adopt the second solution, illustrated in (23). Because only non-reduplicated candidates (c) and (d) survive the top two constraints, raising is determined by the same variable ranking between IDENT-IO[high](stem-ultima) and *NON-ULTIMA[mid] as seen with high penults (dashed line indicates unknown or irrelevant ranking).

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

(23)

/CaCe+in/	Id-IO[lo]	Id-ĸĸ[lo]	Redup	*Non-ult [mid]	Id-IO[hi] (stem-ult)
a. [Ca]$_{\alpha}$[Ce]$_{\alpha}$hin		*!		*	
b. [Ca]$_{\alpha}$[Ci]$_{\alpha}$hin		*!			*
☞ c. CaCehin			*	*	
☞ d. CaCihin			*		*
e. CeCihin	*!		*	*	*

Although there are only two examples, one native and one borrowed, where the last two syllables of the stem are identical, so that no Corr-ĸĸ constraints can interfere with coupling, raising is blocked in both of them. Therefore, I tentatively assume that coupling when there is a mid vowel in the penult implies non-raising, and I attribute this to the fixed ranking Ident-ĸĸ[high], Redup ≫ *Non-ultima[mid].

(24)

/saklolo+in/	Id-ĸĸ[hi]	Redup	*Non-ult [mid]	Id-IO[hi] (stem-ult)
☞ a. sak[lo]$_{\alpha}$[lo]$_{\alpha}$hin			**	
b. sak[lo]$_{\alpha}$[lu]$_{\alpha}$hin	*!		*	*
c. saklolohin		*!	**	
d. sakloluhin		*!	*	*

The relatively high rate of non-raising seen with a mid penult is not attributable solely to the variable ranking of Ident-IO[high](stem-ultima) and *Non-ultima[mid], because we have seen that the variable ranking between those two constraints produces only a low rate of non-raising with non-mid penults (see Fig. 1). Raising with a mid penult is mainly driven instead by the availability of coupling. This availability is determined by the variable ranking of Redup and the Corr-ĸĸ constraints that are boxed in (21): Ident-ĸĸ[back], Ident-ĸĸ[place], Ident-ĸĸ[manner], Ident-ĸĸ[voice] and Max-ĸĸ, where, without making any claims about feature geometry, I use Ident-ĸĸ[place] and Ident-ĸĸ[manner] as shorthand for various Ident-ĸĸ[F] constraints.

In tableau (25), the same candidate types are potentially optimal ((i), (iii) and (iv)). So what makes (a.i) likelier than (b.i)? Under Anttila's (1997) conception, a variable ranking is a partially specified ordering of constraints that must be 'spelled out' at the time of utterance into a completely specified linear ordering. Any spell-out consistent with the fixed parts of the ranking is equally probable. Thus, for the grammar A ≫ {B,C,D} ≫ E, the spell-outs ABCDE, ABDCE, ACBDE, ACDBE, ADBCE and ADCBE should be used equally often by speakers.

There are 6! = 720 possible linear rankings of Redup and the five Corr-ĸĸ constraints with which it is variably ranked. Of these, the 360 in which

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

REDUP outranks IDENT-ĸĸ[voice] choose (25a.i). But only the 180 in which REDUP outranks IDENT-ĸĸ[place], IDENT-ĸĸ[voice] and MAX-ĸĸ choose (25b.i). Therefore, (a.i) is expected to be chosen more often than (b.i). In general, a ranking that allows coupling is more likely when fewer CORR-ĸĸ constraints would be violated.

(25)

		ID-ĸĸ [hi]	RE-DUP	ID-ĸĸ [pl]	ID-ĸĸ [vce]	ID-ĸĸ [man]	ID-ĸĸ [bk]	MAX-ĸĸ	*NON-ULT [mid]	ID-IO[hi] (stem-ult)
a.	/todo+in/									
	☞ i. [to]$_\alpha$[do]$_\alpha$hin				*				**	
	ii. [to]$_\alpha$[du]$_\alpha$hin	*!			*				*	*
	☞ iii. toduhin		*						*	*
	☞ iv. todohin		*						**	
b.	/estorbo+in/									
	☞ i. es[tor]$_\alpha$[bo]$_\alpha$hin			*	*			*	**	
	ii. es[tor]$_\alpha$[bo]$_\alpha$hin	*!		*	*			*	*	*
	☞ iii. estorbuhin		*						*	*
	☞ iv. estorbohin		*						**	

Anttila's model makes a two-way distinction between fixed and free rankings. There do seem to be some intermediate cases, however. Although IDENT-IO[high](stem-ultima) and *NON-ULTIMA[mid] are variably ranked, *NON-ULTIMA[mid] ⩥ IDENT-IO[high](stem-ultima) is more frequent. Also, although none of the five CORR-ĸĸ constraints can consistently rule out coupling – and thus they must all be variably ranked with respect to REDUP – IDENT-ĸĸ[back], IDENT-ĸĸ[place] and MAX-ĸĸ seem to be higher-ranked than IDENT-ĸĸ[voice] and IDENT-ĸĸ[manner]. A probabilistic-ranking model like Boersma's (1998) or Hayes & MacEachern's (1998), which encodes ranking tendencies within variable rankings, is therefore more appropriate for Tagalog vowel raising. In a probabilistic-ranking model, the explanation for the greater likelihood of (25a.i) compared to (25b.i) is the same as in an Anttilan model: because the ranking requirements for (b.i) are stricter (REDUP must outrank three CORR-ĸĸ constraints instead of just one), a randomly chosen linear ranking is less likely to satisfy them.

Why are the CORR-ĸĸ constraints ranked differently from the correspondence constraints seen in morphological reduplication? For example, in morphological reduplication the reduplicant and base consonants always match in nasality (driving overapplication of nasal coalescence), but we have seen that a mismatch in manner between penult and ultima onsets does not significantly affect rates of raising. As mentioned above in §2, correspondence constraints between morphological reduplicants and their

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

bases can be thought of as morpheme-specific instances of CORR-$\kappa\kappa$ constraints, which we could notate CORR-$\kappa_B\kappa_R$. A language with multiple reduplication patterns often requires multiple CORR-$\kappa_B\kappa_R$ series, to deal with differences in reduplicant size and behaviour. For example, in Tagalog complex onsets are often simplified with CV reduplicants, but never in σσ reduplicants. This could be accounted for by variable ranking between MAX-$\kappa_{B\text{-}CV}\kappa_{R\text{-}CV}$ and *COMPLEX, but a fixed ranking MAX-$\kappa_{B\text{-}\sigma\sigma}\kappa_{R\text{-}\sigma\sigma} \gg$ *COMPLEX. If different rankings of CORR-$\kappa_B\kappa_R$ constraints can occur for different reduplicative morphemes, it is not surprising that the rankings of non-morpheme-specific CORR-$\kappa\kappa$ constraints can also diverge.

There are two parts of the ranking in (21) still to be accounted for. The box of CORR-IO constraints outranks the box of CORR-$\kappa\kappa$ constraints because outright enhancements of similarity are rare. Any candidates that, like (20b.ii), violate a CORR-IO constraint to satisfy a CORR-$\kappa\kappa$ constraint (there IDENT-IO[back] and IDENT-$\kappa\kappa$[back]) are ruled out.

The one remaining part of the ranking to be accounted for is the place of IDENT-IO[high](non-stem-ultima). IDENT-IO[high](non-stem-ultima) is variably ranked with *NON-ULTIMA[mid], because, as mentioned above, a few stems undergo 'double raising' (obligatorily or optionally), as in /relebo/ 'relief', /relebu-han/ ~ /relibu-han/ 'to relieve'. There are also a few stems that have varying penult height when unsuffixed, but only a high penult when suffixed: /polbos/ ~ /pulbos/ 'powder', /pulbus-in/ 'to powder'. As shown in the partial tableau in (26), among the REDUP-satisfying candidates either no raising or double raising is possible. This is another case where probabilistic ranking would be useful: there seems to be a preference for IDENT-IO[high](non-stem-ultima) $\gg$ *NON-ULTIMA [mid] $\gg$ IDENT-IO[high](stem-ultima), despite the variable ranking of the three constraints.

(26)

/CoCo + in/	ID-$\kappa\kappa$[hi]	ID-IO[hi] (non-stem ult)	*NON-ULT[mid]	ID-IO[hi] (stem-ult)
☞ a. $[Co]_\alpha[Co]_\alpha$hin			**	
b. $[Co]_\alpha[Cu]_\alpha$hin	*!		*	*
☞ c. $[Cu]_\alpha[Cu]_\alpha$hin		*		*

3.3.1 *Lexical variation.* The rankings given above predict free variation. That is, almost every loan-stem should have both raised and non-raised variants, though the frequency of the variants will be affected by whether the penult has a mid vowel and if so, how similar the penult is to the ultima. In reality, the variation is lexical: the behaviour of most stems is fixed, with either raising or non-raising. It is the distribution of raising across stems, rather than within stems, that is probabilistic.

I adopt here the solution I proposed in Zuraw (2000) to the general problem of how speakers can learn, use and maintain patterns of lexical variation. Existing words' properties – in this case, whether or not a stem

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

undergoes raising – are encoded in their lexical entries in some form (in this case, perhaps by listing the suffixal allomorph of the stem).[21] High-ranked faithfulness constraints ensure that this listed information is expressed. Constraints like REDUP that do not play a core role in the language are ranked lower and variably, because the learner is exposed to limited and perhaps conflicting information about where to rank them. But when the speech community is still in the process of establishing what a loan-stem's new suffixed form should be, faithfulness to a listed suffixal allomorph is vacuously satisfied – there is no listed allomorph to be faithful to – and the lower-ranked constraints come into play.

The fixed and variable rankings argued for above determine, probabilistically, whether a stem without a listed suffixal allomorph undergoes raising. When the stem has a mid vowel in the penult, REDUP and IDENT-*ĸĸ*[high] favour non-raising, but *NON-ULTIMA[mid] and other IDENT-*ĸĸ* constraints favour raising. As seen above, the more dissimilar the penult and the ultima are, the less likely coupling is in these early stages. On any given occasion, a speaker is more likely to assign reduplicated structure to a newly coined suffixed form of /todo/ than to a newly coined suffixed form of /ʔestorbo/. The outcomes of these coinages form the input to other speakers' lexical learning – that is, a speaker's coinage determines whether a listener adds a raised or a non-raised suffixal allomorph to her lexical entry for that stem. Probabilities of raising in new coinages make it more likely that /todo/ will become lexicalised as non-raising, and /ʔestorbo/ as raising. From there, faithfulness to the listed allomorph (abbreviated in (27) as FAITH) takes over. In (27), /todo/'s lexical entry contains a free and a suffixed allomorph. Candidates (b) and (c) are ruled out because they are unfaithful to the suffixal allomorph, leaving the homophonous candidates (a) and (d). The rest of the variable ranking is irrelevant to how the word is pronounced; regardless of whether coupling occurs, raising does not.

(27)

/todo$_{\text{free}}$, todo$_{\text{suffixed}}$+in/	FAITH	ID-*ĸĸ* [hi]	REDUP	ID-*ĸĸ* [vce]	*NON-ULT [mid]	ID-IO[hi] (stem-ult)
☞ a. [to]$_{\alpha}$[do]$_{\alpha}$hin				*	**	
b. [to]$_{\alpha}$[du]$_{\alpha}$hin	*!	*		*	* *	
c. toduhin	*!		*		*	*
☞ d. todohin			*		**	

The reason why only some pseudoreduplicated roots exhibit reduplicative effects is similar: because of high-ranking faithfulness, information

[21] 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.

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

in an individual word's lexical entry determines whether it displays such effects.[22] 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 the direction of aggressive reduplication.

Listed suffixal allomorphs could help explain a difference between loans and native stems: among loan-stems with a non-mid penult vowel, the rate of non-raising is 7%, but among native stems with a non-mid penult vowel, it is 0%. In addition, there is a difference between English and Spanish loans. English loans are more recent, and although there are only 17 English loan-stems in the database, their overall rate of non-raising (65%) is significantly higher than that of the Spanish stems (13%; $p<\cdot0001$). One possible solution is to divide the lexicon into etymological strata, to which different constraints or rankings apply (see e.g. Itô & Mester 1995). This solution seems somewhat unattractive, though, given that cues to foreignness that speakers could use to assign loans to the correct stratum seem to have no effect on rates of raising (see §3.4 below). Another possibility is that, because of the influx of loans with mid penult vowels over the past few centuries, the ranking of *NON-ULTIMA[mid] has declined slightly. The higher rate of raising of native stems can be attributed to their having a listed, raised form that was established when *NON-ULTIMA[mid] was ranked higher.

3.4 Other explanations of vowel raising

In light of the data presented above in §3.2, three other possible explanations for why a mid vowel in the penult inhibits raising can be rejected in favour of aggressive reduplication. The first is that the presence of a non-ultima mid vowel, which is unexpectedly [−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 would have to assume that a non-ultima mid vowel is more likely to mark adjacent syllables as contrastive for [high].

The second explanation 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 & Mester 1995) that

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

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

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 and found that they did not. Stress or length on a non-final closed syllable and pre-penultimate stress or length are rare or non-existent 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 the difference is not significant ($p > \cdot 5$). 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 ($p > \cdot 25$). 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.

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.

Although these three accounts could possibly be made to capture the matching-backness and proximity effects, they cannot capture the effects of similarity between the penult and the ultima along other dimensions. If the penult's and ultima's onsets share place of articulation, for example, that should not encourage the penult vowel's contrastiveness for [high] to be extended to the ultima vowel, or the penult vowel's foreignness to be extended to the ultima vowel, or the penult vowel to trigger harmony on the ultima vowel. I therefore reject these other explanations for the distribution of exceptions to vowel raising.

4 Generation or learning?

The above sections have argued that, because of the interaction of REDUP with correspondence constraints, self-similarity within a word can cause it to have internal coupling *during generation*, leading to lexical entries that display enhancement or preservation of self-similarity.

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

An alternative to the generation account is that reduplicative construals are imposed during lexical learning itself.[23] The resulting lexical entries with morphosyntactically unmotivated reduplicative structure would be similar to those proposed by Buckley (1998), Gafos (1998) and Golston & Thurgood (in press) (see §1). The Tagalog vowel raising case differs from the cases examined by those authors in that the reduplication is imperfect in the Tagalog case, and segments must be placed in correspondence that have varying amounts of non-shared underlying material (e.g. [t] and [d], [k] and [d]).

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 & 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 (28). 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. If the grammar is made up solely of markedness and faithfulness constraints, then, putting aside allomorphy, lexicon optimisation will cause all lexical forms to be learned as identical to surface forms. Prince & Smolensky suggest, however, a constraint *Spec (or perhaps a family *Spec[V], *Spec[round], etc.), which penalises phonological material in the underlying form, much as *Struc (Zoll 1993) penalises material in the surface form. The extent to which the underlying form matches the surface form in lexicon optimisation is determined, then, by the ranking of faithfulness constraints with respect to *Spec. 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 (28), candidate (a), with no aspiration in the underlying form, is preferred because *Spec ≫ Ident-IO[spread glottis].[24]

(28) *Lexicon optimisation*

[kʰæt]	Dep(C)	*Spec	Id-IO[spr gl]	*[+spr gl]	NoCoda
☞ a. /kæt/ → [kʰæt]		k,æ,t	*	*	*
b. /kʰæt/ → [kʰæt]		k,ʰ!,æ,t		*	*
c. /æt/ → [kʰæt]	*!	æ,t		*	*

A reduplicative construal of the underlying form may save violations of *Spec, by representing repeated material (segments or features) once

[23] I'm indebted to an anonymous reviewer and to Dominique Sportiche and Donca Steriade for suggesting and pressing this line of explanation.

[24] How the child would learn the ranking of *Spec constraints is uncertain. Perhaps *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.

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

instead of twice. Without committing to any particular model of imperfectly reduplicated lexical representations, (29) 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, (29a) is given as /RED[−voice] do/, with the non-repeated material of the first 'copy' in parentheses after RED, and RED is treated by the constraints as a normal reduplicative morpheme, so that CORR-BR constraints apply.

(29) a. *imperfectly reduplicated* b. *unreduplicated*



The lexical learner also must determine 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 tableau in (30a) illustrate the four candidate types that the learner must consider, given an overt form (shown in quotes): the underlying form may be reduplicated (i, ii) or not (iii, iv), and the surface form may have correspondence between the non-identical segments (i, iii) or not (ii, iv). The ranking of *SPEC with respect to CORR-BR constraints determines how much deviation from perfect identity is to be tolerated in a reduplicative construal.[25] 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.[26]

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

[26] Although the choice between candidates (i) and (ii) that MAX-BR makes is irrelevant to lexical learning itself ((i) and (ii) 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.

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

(30) *Reduplicative construal in lexicon optimisation*

		MAX-BR	ID-BR [pl]	*SPEC	ID-BR [vce]
a.	'todo'				
☞	i. /RED[−vce] do/ → [[to]$_R$[do]$_B$]			[−vce]do	*
	ii. /RED[−vce] do/ → [t[o]$_R$[do]$_B$]	*!		[−vce]do	
	iii. /todo/ → [[to]$_R$[do]$_B$]			todo !	*
	iv. /todo/ → [todo]			todo !	
b.	'kodo'				
	i. /RED[−vce, dors] do/ → [[ko]$_R$[do]$_B$]		*!	[−vce] [dors]do	*
	ii. /RED[−vce, dors] do/ → [k[o]$_R$[do]$_B$]	*!		[−vce] [dors]do	
	iii. /kodo/ → [[ko]$_R$[do]$_B$]		*!	kodo	*
☞	iv. /kodo/ → [kodo]			kodo	

As with the generation account, we must allow for variable constraint ranking, so that a word that is self-similar along more dimensions has a higher probability of being learned with a reduplicated underlying representation. 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 suffixed 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.

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 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 is implausible – mishearing influenced by top-down

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

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 their treatment under the lexical learning account as a topic for future research.

4.1 Distinguishing between learning and generation

How can we determine whether aggressive reduplication is caused by *SPEC's activity in lexical learning or by REDUP's activity in generation? To distinguish between the learning and generation accounts on empirical grounds, we need cases in which a stem has one set of self-similarity properties in the form that would be used for lexical learning, and a different set in the form where reduplicative structure would have a detectable effect. In the Tagalog vowel-raising case, lexical learning should take place based on the unsuffixed form, because (i) that is the form in which the loan-stems enter the language,[27] 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 of one generation treat a word consistently, the next generation is not in a position to decide whether the word should be construed reduplicatively). The learner's decision about the structure of the bare stem would carry over to the suffixed stem, even if its self-similarity properties were different. The self-similarity properties of the unsuffixed form, not of the suffixed form, should determine raising. By contrast, in the generation account, although unsuffixed forms are subject to REDUP, what matters for raising is whether coupling is imposed during generation of suffixed forms. 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? 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 rhyme shape. Because suffixation always opens the final syllable of the stem (a final consonant becomes an onset), identical rhyme 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: similarity of rhyme 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 Table II). Because stems from English have higher rates of non-raising than stems from Spanish, and are all consonant-final, counts in Table II are shown separated by language of

[27] Thanks to Michael Wagner for raising this point. The suffixes in question are derivational, not inflectional, so it is not necessary for a new loan-stem to be suffixed right away. Moreover, the majority of loan-stems do not have a suffixed form in the dictionary, whereas only a handful of loan-stems fail to occur unsuffixed.

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

origin and totalled. As predicted by both the generation and learning accounts, the 'same-same' (a) cells have significantly more non-raising than the 'different-different' (c) cells, both for Spanish origin ($p<\cdot05$) 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 ($p<\cdot05$), 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 ($p<\cdot001$) or Spanish counts alone ($p<\cdot005$). Which has a stronger effect, similarity in the unsuffixed form or similarity in the suffixed form? The overall rate of non-raising 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. Rhyme shape fails, therefore, to distinguish between the lexical-learning and generation accounts, and suggests that both play a role.

	unsuffixed shapes	suffixed shapes	non-raising (Spanish origin)	non-raising (English origin)	% non-raising (total)
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**.to**ɾ**)	different (d**ok**.t**o**.ɾ-in)	2 / 8	4 / 4	50%

Table II

Effect of penult–ultima rhyme shape similarity in suffixed *vs.* unsuffixed forms.

The evidence from other properties is also inconclusive. For stress and length, which can shift to the right under suffixation, there is a non-significant effect ($p = \cdot052$) in the direction predicted by the generation account, but there are insufficient tokens to test the prediction of the lexical-learning account. Prefixation with a vowel-final prefix such as /ka-/ can induce tapping of [d] to [ɾ] (native [dáːmot] 'stinginess', [ka-ɾamúːt-an] 'stinginess'), which, in disyllable loan-stems could change 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

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

at tapping under suffixation of stem-final [d]'s, but there are none in the database. Finally, certain nasal-coalescing 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 (native [**k**upás] 'faded', [ma-**ŋ**úːpas] 'to become faded'). There are, however, no two-syllable stems in the database that take a potentially nasal-coalescing prefix.

There are some sporadic examples from another Western Austronesian language, Palauan, that suggest that aggressive reduplication can be conditioned solely by similarity properties of the affixed form, implying 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 3rd person possessive suffix [-ɛl] is attached to [kləkədáll] 'parts', the result has variants [kləkədəll-él], with reduction of the stressed [a] to schwa, and [kdəkdəll-él], 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.[28] No other instances of [l] changing to [d] were found in a database of 1019 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 [kCəkCə…]), which is then enhanced (by turning the first C into a [d] to match the second).

Diphthongs in Palauan reduce to their fronter or higher member when unstressed (in a conflict, the frontness preference prevails): [orʂóməl] 'forest', [orɛməl-él] 'her/his/its forest'. There are three cases of an unstressed [ə] 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).

(31)	məŋərʂóməl	'preserve'	məŋɛrɛməl-él	'its captain'
	ʔəllέo̯ʔ	'spoiling'	ʔɛllɛʔ-él	'her/his spoiling (of a child)'
	bləkέu̯	'bravery'	blɛkɛŋ-él	'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; 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]).

[28] 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.

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

5 Other approaches to word-internal similarity

This section reviews some related proposals of constraints or conditions that promote word-internal similarity and concludes that correspondence between individual segments is insufficient to deal with all cases. A relation between strings, like the coupling that REDUP imposes, is necessary.

5.1 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[29] 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^{h}usp^{h}u] 'boiled beans'). MacEachern finds that in languages with such an 'escape clause' for identical segments, laryngeally dissimilar homorganic consonants are forbidden or rare (*[t^{h}ata], *[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],[30] then homorganic consonants will always be identical, and non-homorganic consonants must not be laryngeally similar:

(32)

	/k'at'a/	ID-IO[pl]	BEIDENT	*LARSIM	ID-IO[lar]
a.	☞ i. k'ata		*		*
	ii. k'at'a		*	*!	
	iii. t'at'a	*!		*	
b.	/t'ata/				
	☞ i. t'at'a			*	*
	ii. t'ata		*!		

In terms of aggressive reduplication, this is equivalent to the ranking IDENT-IO[place], IDENT-ʁʁ[place], IDENT-ʁʁ[laryngeal] ≫ REDUP ≫ *LARYNGEALSIMILARITY ≫ IDENT-IO[laryngeal]. That is, consonants that already have the same place of articulation (high-ranked IDENT-IO[place]

[29] 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.

[30] 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.

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

and IDENT-ĸĸ[place] require this) are forced to have the same laryngeal features also:

(33)

	/k'at'a/	ID-IO [pl]	ID-ĸĸ [pl]	ID-ĸĸ [lar]	REDUP	*LARSIM	ID-IO [lar]
a.	i. k'at'a				*	*!	
	☞ ii. k'ata				*		*
	iii. [k'a]$_\alpha$[ta]$_\alpha$		*!	*!			*
	iv. [k'a]$_\alpha$[t'a]$_\alpha$		*!			*	
	v. [t'a]$_\alpha$[t'a]$_\alpha$	*!				*	
b.	/t'ata/						
	i. t'ata				*!		
	ii. [t'a]$_\alpha$[ta]$_\alpha$			*!			
	☞ iii. [t'a]$_\alpha$[t'a]$_\alpha$					*	*

How can we test whether the escape clause is due to BEIDENTICAL, which requires that single consonants be identical, or to aggressive reduplication, which requires maximal similarity between corresponding strings? The aggressive reduplication analysis predicts that, if IDENT-ĸĸ 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^{h}usp^{h}u] and few words like [k'ink'u]. MacEachern (personal communication) 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$^h{}_1$V$_2$(C)C$^h{}_1$V$_3$ … roots (84%), and in 18 of 23 C'$_1$V$_2$(C)C'$_1$V$_3$ … 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 ($p<\cdot002$ for comparing the first two root types with the controls, scaled to match the number of test cases, 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$^h{}_1$V$_2$(C)C$^h{}_1$V$_3$ … roots (82%) and 18 of 18 C'$_1$V$_2$(C)C'$_1$V$_3$ … roots (100%), but only 54 of 114 of controls (47%) ($p<\cdot0001$).

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, although 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.

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

5.2 Suzuki's IDENTONSET

Suzuki (1999), in reanalysing data from Cohn (1992), proposes a constraint family IDENTONSET 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 *l* … *r* is blocked in Sundanese when the two /r/'s are onsets of adjacent syllables. Cohn explains the phenomenon, which occurs both under *-ar-* infixation and within roots, by linking two adjacent-onset /r/'s to a single [−lateral] specification. Suzuki's proposal differs from MacEachern's in requiring consonants to be identical only if they are in a particular prosodic arrangement (as Suzuki points out, in MacEachern's data, ejectives and aspirated consonants are restricted to onset position anyway).

An 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 two of 22 infixed /CVrCV …/ words, both speakers she consulted unexpectedly infix *-al-*, and one speaker finds both *-ar-* and *-al-* acceptable for an additional five 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 strings that are similar in other ways.

5.3 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.

5.4 Struijke's IDENT[F]$_{\Sigma\Sigma}$

Struijke (2000) proposes that when two output segments derive from a single input segment, whether in reduplication or in other types of fission, they are required to be similar not just indirectly (through their shared correspondence to the same underlying segment) but also directly, because of a constraint family $\text{IDENT[F]}_{\Sigma\Sigma}$ that requires them to bear identical feature values. Struijke's proposal would not apply to the Tagalog raising cases, because there the segments that would have to correspond (the penult and ultima vowels) do not derive from a common underlying segment.

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

5.5 Walker and Rose's Consonantal Correspondence

Walker and Rose (Walker 2000, Rose & Walker 2001) propose a family of constraints that require consonants to enter into correspondence if they already share certain feature values. For example, CORRIFIDENT[voice][31] requires two consonants that are identical on the surface in [voice] to correspond. The constraint family is similar to REDUP in that perfect identity is not required – only a correspondence relation is required, and it is left to the IDENT-CC[F] family to enforce similarity (partial or total) along other dimensions. IDENT-CC[F] constraints require identical feature values for consonants that are in correspondence within an output. Rose & Walker'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 penalises non-coupling regardless of prior similarity, whereas consonantal correspondence constraints apply only if 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-$\kappa\kappa$[F] high. Therefore, these two types of correspondence-requiring constraint make largely overlapping empirical predictions. There is, however, one exception.[32] 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 (34), /daba/ $\rightarrow$[$d_\alpha ad_\alpha a$] (matching Greek-letter subscripts indicate correspondence between the two consonants), and /data/$\rightarrow$[$d_\alpha ad_\alpha a$] (because the input consonants are already identical in either voicing or place), but /dapa/ $\rightarrow$[$d_\alpha ad_\beta a$], with no change (mismatched subscripts indicate lack of correspondence).

[31] Rose & Walker's notation is CORR-$C_1 \leftrightarrow C_2$. CORRIFIDENT[voice] = CORR-$T_1 \leftrightarrow D_2$ and CORRIFIDENT[place] = CORR-$T_1 \leftrightarrow K_2$.

[32] Factorial typologies for the two approaches were calculated using Hayes *et al.* (2000).

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

(34)

	/daba/	Id-CC [pl]	Id-CC [vce]	CorrIf Id[pl]	CorrIf Id[vce]	Id-IO [pl]	Id-IO [vce]
a.	i. $d_\alpha ab_\beta a$				*!		
	ii. $d_\alpha ab_\alpha a$	*!					
	☞ iii. $d_\alpha ad_\alpha a$					*	
b.	/data/						
	i. $d_\alpha at_\beta a$			*!			
	ii. $d_\alpha at_\alpha a$		*!				
	☞ iii. $d_\alpha ad_\alpha a$						*
c.	/dapa/						
	☞ i. $d_\alpha ap_\beta a$						
	ii. $d_\alpha ab_\beta a$				*!		*
	iii. $d_\alpha at_\beta a$			*!		*	
	iv. $d_\alpha ap_\alpha a$	*!	*!				
	v. $d_\alpha ab_\alpha a$	*!					*
	vi. $d_\alpha at_\alpha a$		*!			*	
	vii. $d_\alpha ad_\alpha a$					*!	*!

In aggressive reduplication, by contrast, if Redup and the Ident-κκ[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]$_\alpha$].

(35)

	/daba/	Id-κκ[pl]	Id-κκ[vce]	Redup	Id-IO[pl]	Id-IO[vce]
a.	i. daba			*!		
	ii. [da]$_\alpha$[ba]$_\alpha$	*!				
	☞ iii. [da]$_\alpha$[da]$_\alpha$				*	
b.	/data/					
	i. data			*!		
	ii. [da]$_\alpha$[ta]$_\alpha$		*!			
	☞ iii. [da]$_\alpha$[da]$_\alpha$					*
c.	/dapa/					
	i. dapa			*!		
	ii. [da]$_\alpha$[pa]$_\alpha$	*!	*!			
	iii. [da]$_\alpha$[ba]$_\alpha$	*!				*
	iv. [da]$_\alpha$[ta]$_\alpha$		*!		*	
	☞ v. [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

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

laryngeal features and place, or differ in both (MacEachern 1999), but in those cases there are independently attested constraints against laryngeal similarity. Those cases can be analysed with either consonantal correspondence or aggressive reduplication.

6 Conclusion

I have proposed that there is a purely phonological drive for words to be treated as though they were reduplicated, through a relation between strings (coupling) that invokes correspondence constraints between segments of an output. 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 coupling, and coupling blocks raising.

Are there alternations in other languages that could be systematically affected by aggressive reduplication? One candidate is laxing in Quebec French (e.g. Dumas 1976). High vowels in closed final syllables are usually laxed ([kɔl**ɪ**n] 'hill', as opposed to European French [kɔl**i**n]). If the penult also has a high vowel, it optionally undergoes laxing harmony ([m**i**nʏt] ~ [m**ɪ**nʏt] 'minute'). Aggressive reduplication predicts that similarity between the penult and the ultima should promote laxing harmony. To test this hypothesis, it would be necessary to examine a large corpus of casual, spoken Quebec French.

Although the proposal made here is not that speakers attribute an actual reduplicative morpheme to words with sufficient self-similarity, I do propose that they impose a structure that is the same as reduplication (coupling), but without tagging for any particular RED morpheme. The idea of purely phonological drives for inferring 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 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.[33] 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.

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

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

Appendix

In a database of 4390 disyllabic non-loan roots from English's (1986) dictionary (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. What is the probability, p, of obtaining this many pseudoreduplicated roots if phonemes were combined randomly? To answer the question directly, we should perform every possible reshuffling of the phonemes in the database and determine how many reshufflings yield as many pseudoreduplicated roots as were actually observed. 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 Brett Kessler for discussing his technique and its application to this case.) The distribution obtained was approximately normal, with medians well below the actually observed numbers. The result of interest is the number of reshufflings that yielded as many pseudoreduplicated roots as actually occur, or more. This number, divided by 10,000, gives the estimated value of p. Not even one of the 10,000 reshufflings had as many pseudoreduplicated roots as actually occur in any of the three pseudoreduplication categories, so the estimated p is less than ·0001. As shown in Table III below, the results were similar for shuffling onset and rhyme, onset+nucleus ('head') and coda, and whole syllables. The highest and lowest numbers obtained are also given; these would tend to move closer to the median if fewer shufflings were performed, and further from the median if more were performed.

		unit of shuffling				
		phoneme	onset/rhyme	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	

Table III
Number of pseudoreduplicated roots obtained.

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

REFERENCES

Anttila, Arto (1997). Deriving variation from grammar. In Frans Hinskens, Roeland van Hout & W. Leo Wetzels (eds.) *Variation, change and phonological theory*. Amsterdam: Benjamins. 35–68.
Ayala Loayza, Juan Luis (1988). *Diccionario español–aymara, aymara–español*. Lima: Editorial J. Mejia Baca.
Banksira, Degif Petros (2000). *Sound mutations: the morphophonology of Chaha*. Amsterdam: Benjamins.
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 (1998). 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: Graphos 100 Editores.
Dumas, Denis (1976). Quebec French high vowel harmony: the progression of a phonological rule. *CLS* **12**. 161–167.
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. *NLLT* **16**. 223–278.
Goldsmith, John (2001). Unsupervised learning of the morphology of a natural language. *Computational Linguistics* **27**. 153–198.
Golston, Chris & Elzbieta Thurgood (in press). Reduplication as echo: evidence from Bontok and Chumash. In Ritva Laury, Gerald McMenamin, Shigeko Okamoto, Vida Samiian & 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 73rd Annual Meeting of the Linguistic Society of America, Los Angeles.
Hayes, Bruce & Margaret MacEachern (1998). Quatrain form in English folk verse. *Lg* **74**. 473–507.
Hayes, Bruce, Bruce Tesar & Kie Zuraw (2000). OTSoft. Software package, http://www.linguistics.ucla.edu/people/hayes/otsoft/.
Itô, Junko & Armin Mester (1995). Japanese phonology. In John Goldsmith (ed.) *The handbook of phonological theory*. Cambridge, Mass. & Oxford: Blackwell. 817–838.
Josephs, Lewis (1975). *Palauan reference grammar*. Honolulu: Hawaii University Press.
Josephs, Lewis (1990). *New Palauan–English dictionary*. Honolulu: University Press of Hawaii.
Kaun, Abigail (1995). *The typology of rounding harmony: an Optimality Theoretic approach*. PhD dissertation, UCLA.
Kessler, Brett (2001). *The significance of word lists*. Stanford: 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.

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press

McCarthy, John (1981). A prosodic theory of nonconcatenative morphology. *LI* **12**. 373–418.
McCarthy, John (1983). Consonantal morphology in the Chaha verb. *WCCFL* **2**. 176–188.
McCarthy, John & Alan Prince (1995). Faithfulness and reduplicative identity. In Jill Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*. Amherst: GLSA. 249–384.
McCarthy, John & Alan Prince (1999). Faithfulness and identity in Prosodic Morphology. In René Kager, Harry van der Hulst & Wim Zonneveld (eds.) *The prosody–morphology interface*. Cambridge: Cambridge University Press. 218–309.
MacEachern, Margaret (1999). *Laryngeal cooccurrence restrictions*. New York: Garland.
Marantz, Alec (1982). Re reduplication. *LI* **13**. 435–482.
Mester, R. Armin (1988). *Studies in tier structure*. New York: Garland.
Nash, David (1980). *Topics in Warlpiri grammar*. PhD dissertation, MIT.
Prince, Alan & Paul Smolensky (1993). *Optimality Theory : constraint interaction in generative grammar*. Ms, Rutgers University & University of Colorado, Boulder.
Rose, Sharon & Rachel Walker (2001). A typology of consonant agreement as correspondence. Ms, UCSD & USC.
Schachter, Paul & Fe Otanes (1972). *Tagalog reference grammar*. Berkeley: University of California Press.
Struijke, Caro (2000). *Existential faithfulness : a study of reduplicative TETU, feature movement, and dissimilation*. PhD dissertation, University of Maryland, College Park.
Suzuki, Keiichiro (1999). Identity avoidance vs. identity preference: the case of Sundanese. Paper presented at the 73rd Annual Meeting of the Linguistic Society of America, 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. Technical Report TR-58, Department of Computer Science, University of Colorado, Boulder.
Walker, Rachel (2000). Long distance consonantal identity effects. *WCCFL* **19**. 532–545.
Wilbur, Ronnie (1973). *The phonology of reduplication*. 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 & Patrick Farrell (eds.) *Morphology and its relation to phonology and syntax*. Stanford: CSLI. 216–246.
Zoll, Cheryl (1993). Directionless syllabification and ghosts in Yawelmani. Transcript of paper presented at Rutgers Optimality Workshop 1, Rutgers University.
Zuraw, Kie (2000). *Patterned exceptions in phonology*. PhD dissertation, UCLA.

https://doi.org/10.1017/S095267570300441X Published online by Cambridge University Press