UNIVERSITY OF CALIFORNIA

Los Angeles

Floating Phonotactics:

Variability in Infixation and Reduplication of Tagalog Loanwords

A thesis submitted in partial satisfaction of the

requirements for the degree Master of Arts

in Linguistics

by

Kie Ross

1996

The thesis of Kie Ross is approved.

Edward Stabler

Bruce Hayes, Committee Co-Chair

Donca Steriade, Committee Co-Chair

University of California, Los Angeles

1996

CONTENTS

ACKNOWLEDGMENTS

Many thanks to Bruce Hayes, Margaret MacEachern, Edward Stabler, and Donca Steriade for discussion and guidance, to Leah S. Gordon for programming advice, to Bryan Zuraw for editorial help, and to Tania Azores-Gunter for teaching me Tagalog. This work was supported by a National Science Foundation Graduate Fellowship.

ABSTRACT OF THE THESIS

Floating Phonotactics:

Variability in Infixation and Reduplication of Tagalog Loanwords

by

Kie Ross

Master of Arts in Linguistics

University of California, Los Angeles, 1996

Professor Bruce Hayes, Co-Chair

Professor Donca Steriade, Co-Chair

When a phonological constraint is unviolated in a language, but no lexical items exist for which the constraint is relevant, the constraint's ranking cannot be determined. Language contact can introduce loanwords which, if faithfully parsed, violate the constraint, thus exposing its ranking. Tagalog has many English and Spanish loanwords which, in the loaning language, contain segments and sequences foreign to Tagalog. Field data presented here show that constraints against the foreign segments and sequences ("new" constraints) are variably ranked with respect to established constraints (phonotactic constraints which were active in Tagalog prior to the loanwords' introduction, and the constraints which enforce uniformity of allomorphs and faithfulness to the lexical entry). That is, the new constraints may be obeyed despite violations of established constraints, or may be violated in order to obey the established constraints.

Constraints relevant to Spanish loanwords, which have been in Tagalog longer, tend to be ranked lower, suggesting they have drifted downward. Variable ranking is natural for a constraint whose ranking was, until recently, indeterminate. Older loanwords from Chinese and Sanskrit demonstrate that similar variability existed at the time of their adoption into Tagalog, although individual words have since become fossilized in the absence of strong, continuing language contact. Part II of the thesis presents an algorithm for imperfect learning of variable data, which results in a speech community's tolerating lower and lower ranking of a variably-ranked constraint over time.

Part I: Data and Analysis

0. Introduction

Prince and Smolensky 1993 propose a model of phonology based on constraint interaction known as *Optimality Theory*. In this model, the optimal utterance is selected from a set of candidate utterances according to how well it satisfies various phonological constraints. Since some of the constraints are in conflict with one another, they must be *ranked*, so that satisfaction of one constraint is more important than satisfaction of another. Most—ideally all—phonological constraints are presumed to be universal, with differences in constraint ranking producing language-to-language variation. Most constraints fall into two classes. The first class contains the *markedness* or *phonotactic* constraints, which reflect desirable surface properties of individual utterances, such as articulatory ease and acoustic clarity (Hayes 1995, Jun 1994, Silverman 1995, Flemming 1995, Steriade 1995a, 1995b). The second class contains constraints which enforce lexical contrast and morpheme identifiability—these are viewed by some as *faithfulness* constraints, requiring identity of inputs and outputs (Prince and Smolensky 1995) and by others as *paradigm uniformity* constraints, requiring identity of related surface forms (Steriade 1996, Benua 1995, Crosswhite 1995). The term *correspondence constraint* (McCarthy and Prince 1995) refers to any constraint that enforces identity, whether input-output (faithfulness) or output-output (paradigm uniformity, including reduplicative identity).

Gnanadesikan 1995 argues that the child language-learner comes to the task equipped with a large set of high-ranking universal phonotactic constraints, which are gradually suppressed (that is, outranked by faithfulness constraints) as the target language is acquired. The final, adult grammar is a partial order of constraints, with phonotactic and faithfulness constraints interleaved.

I know of no optimality-theoretic work that does not share this view that a grammar is a partial order of constraints (or, as a subcase, a total order), although some (e.g. Mester 1996,Hayes & MacEachern 1996) include mechanisms to explain variability, such as variable and stochastic constraint ranking. Although this thesis, ultimately, will share that view, it will come to do so only after considering a possible challenge to it.

The challenge is *invisible constraints*, constraint whose status in unclear from the surface forms of a language. There are two types of invisible constraint. The first is a phonotactic constraint C which is apparently ranked so low in a language L that its effects are never seen. It is freely violated, and never plays a role in candidate selection. The only reason for supposing that C is present at all in the grammar (since the results would be the same if it were not), is the assumption of constraint universality: if C is active in other languages, then C must play a role in the grammars of children acquiring L, although it may be discarded later in life. This first type of invisible constraint will not be considered here.

The second type of invisible constraint, which will be considered here, is a phonotactic constraint C which is always obeyed in L and is not in conflict with any other constraint—that is, satisfaction of C does not require violation of any other phonotactic constraint, and *it is not necessary to suppose any inputs in L for which satisfaction of C in the output would require violation of a faithfulness constraint.* A priori, there is no way to determine whether (i) a given speaker does assume at least some inputs which, if parsed faithfully, would violate C, so that C must be ranked very high to rule out the fully faithful parses, or (ii) the speaker assumes no such inputs, so that the ranking of C is irrelevant. We need not even suppose that C is present at all, except by indirect arguments from universality and from child language.

This paper probes the status of some such type-2 invisible constraints in Tagalog, by observing the behavior of the grammar when confronted with new inputs.

Tagalog, the language spoken by approximately one third of the population of the

Philippines and the basis for the national language, Pilipino (Grimes 1988), has a large number of loanwords from Spanish and, more recently, from English. In the foreign source language, many of these words contain violations of phonotactic constraints that were surface-true in Tagalog before the introduction of the loanwords. The variable behavior of these words under infixation and reduplication will be seen to obey a possibly general principle: *the ranking of unviolated invisible constraints is not fixed.*

0.1 A note on the data

Data presented without a source cited are uncontroversial facts about contemporary Manila Tagalog, which I have checked with Tagalog-speaking acquaintances, Ramos and Goulet 1981, Ramos and Cena 1990, and/or English 1986.

Soberano 1992, cited below, was a phonological experiment in which NP, a native speaker of Tagalog in her thirties who had recently come to Canada from Manila, was asked to "borrow" hypothetical English loanwords, some with no morphology, some with reduplication, and some with infixation. A copy of the wordlist was placed in front of NP, and each stimulus word was read aloud by Soberano, an English speaker. The data from Soberano presented here are my own transcriptions of the sessions, at which I was present. Items shown with no asterisk are either (i) NP's response(s) when asked to infix or reduplicate a hypothetical loanword, or (ii) positive grammaticality judgements of additional possibilities for that word proposed by Soberano. Items with an asterisk represent negative grammaticality judgements of additional possibilities by NP, either offered spontaneously or made in response to questioning by Soberano.

Also presented in this paper are data from my interviews with three Tagalog-English bilingual UCLA undergraduates, all of whom were born in the Philippines and had lived in the US for a few years. BH was the most fluently bilingual; he used Tagalog on a daily basis and his English was slightly accented. MC and NT, both female, had slight American accents when speaking Tagalog and used Tagalog mostly with their parents, who lived in other towns. Each of these speakers was given a list of English words and asked to apply first infixation, then reduplication, to each item on the list. NT had difficulty bringing reduplication and infixation under conscious control, and applied both simultaneously to all the items but one, which was infixed only.

All phonetic symbols used in this paper are IPA (International Phonetic Association 1989) unless otherwise noted.

1. Reduplication

Tagalog employs reduplication extensively in its derivational and inflectional morphology. The Tagalog reduplicant may be either partial or total, and precedes the base. The type of reduplication we will consider here involves a one-syllable reduplicant and marks incomplete aspect on verbs.

Reduplicants, which I will abbreviate R, are in general in stricter accordance with Tagalog phonology than are loanword bases, which I will abbreviate B.

A frequently discussed example is initial cluster simplification: complex onsets are not native to Tagalog (Quilis 1985), but are tolerated in loanwords. When reduplication is applied, the onset is simplified in R, the reduplicant (in boldface):

(1)	Tagalog	
	nag-tɾabahoh	‘X worked’
	nag-**ta**-tɾabahoh	‘X is working’ (French 1988, p. 28)
from	Spanish	
	trabaxo	‘I work’

This is explained by French 1988 with CV templates: in her view, reduplication is the result of associating a template consisting of consonant and vowel positions to a total copy of the base. Unassociated segmental and skeletal material is deleted or does not surface, and the associated Cs, Vs, and segments form the reduplicant. The template for one-syllable reduplication in Tagalog is CV, which associates to the first C and the first V in the base, skipping over the second C:

(2)
```
C V             CCVCVCVC
| |             ||||||||
tɾabahoh        tɾabahoh
```

The CV explanation does not refer directly to the foreign-ness of complex onsets in Tagalog, and thus does not generalize to the case of the foreign segments [f, θ, ʃ, dʒ][1], which are similarly nativized in R, the reduplicant (in boldface):

(3) (real and hypothetical loanwords compiled from Soberano 1992 and my own

[1] The other segments of Spanish and/or English not found in native Tagalog words are [v, ð, z, ʒ, r, ɹ]. Examples of these are not included for the following reasons:

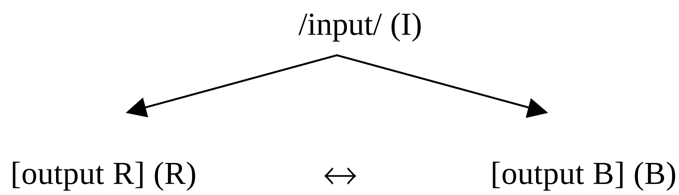
[v, z, r] always nativized in Tagalog to [b, s, |] respectively.
[ɹ] always nativized in Tagalog to [|] except sometimes in coda position, and there are no codas in Tagalog one-syllable reduplication.
[ð, ʒ] no loanwords beginning with these segments.

fieldwork, real loanwords from spontaneous speech marked with a †)

[f]	mag-**pa**fambʊl ‘to fumble’	
[θ]	mag-**tɛ**θæŋkju[1]	‘to say thank-you’†
	tuməθɹo, **t**inəθɹo	‘to throw’
	tumɪθɪŋk	‘to think’
[ʃ]	mag-**sa**ʃapiŋ	‘to shop’†
	mag-**sa**ʃɾab	‘to engage in a shrub-related activity’
[dʒ]	nag-**dja**dʒagiŋ ‘to jog’†	

McCarthy and Prince (1995) propose that three relationships are relevant in a reduplicated form, input-base, base-reduplicant, and, usually without effect, input-reduplicant:

(4)



This paper does not make use of the full machinery of Correspondence Theory, so I will only briefly summarize. I refer the reader to McCarthy and Prince 1995 for extensive development of Correspondence Theory as applied to reduplication. McCarthy and Prince’s Correspondence Theory involves first Correspondence constraints, which require coindexed—that is, corresponding—segments in the two representations involved, for instance input and base. There are then further Identity constraints, which enforce featural

[1] Note that there is also a lack of identity between the vowels in some of the reduplicants and the corresponding vowels in the base. It would be incautious to characterize this as an effect of reduplication, since the distribution of vowel allophones according to syllabic environment differs from speaker to speaker and is variable even within individuals (Stockwell 1957, Schacter&Otanes 1972).

similarity between coindexed segments.

Thus, Correspondence constraints may be obeyed even when Identity constraints are violated:

(5) s_1 a_2 - $ʃ_1$ a_2 p_3 i_4 $ŋ_5$

In (5), segments 1 and 2 of the base have correspondents in the reduplicant, so Correspondence is satisfied for this substring, but segment 1's reduplicant correspondent is not featurally identical, thus violating at least one Identity constraint, say Identity[anterior], while obeying others, such as Identity[continuant] and Identity[strident]).

Conversely, Correspondence constraints may be violated while Identity constraints are vacuously satisfied:

(6) t_1 a_3 - t_1 $ɾ_2$ a_3 b_4 a_5 h_6 o_7 h_8

In (6), segment 2 of the base lacks a correspondent in the reduplicant, thus violating one of the correspondence constraints, but there is no Identity violation, since those segments that are in correspondence are identical.

I will make use freely of both segmental Correspondence, which is relevant in cluster simplification, and featural Identity, which is relevant for the examples involving foreign segments and will refer to both types of constraint as Uniformity (I-B Uni, R-B Uni, and I-R Uni) in order to highlight parallels between the treatment of complex onsets and the treatment of foreign segments.

In (7), we see a tableau for [magtatɾabahoh], with [mag] omitted to save space. The Base-Reduplicant asymmetry exemplified here, where R is less marked than B, is known as the *emergence of the unmarked* (McCarthy & Prince 1994, 1995) and occurs because R-B

Uniformity and I-B Uniformity, which in this case disfavor the deletion of a segment, may be ranked differently with respect to a phonotactic or markedness constraint, such as *ComplexOns, the prohibition against complex onsets such as [tɾ] in [tɾabahoh]. Thus, [**ta**tɾabahoh] and [**tɾa**tɾabahoh] both satisfy I-B Uniformity, since both have [tɾ] in the base, but of these two, [**tɾa**tɾabahoh] has one more instance of a complex onset, and so [**ta**tɾabahoh] is the winning candidate, despite its violation of both R-B and I-R Uniformity.

(7) I-B Uni >> phonotactic >> {R-B Uni, I-R Uni [1]}

/RED+ tɾabahoh /	I-B Uni	*ComplexOns	R-B Uni	I-R Uni
-> **ta** tɾabahoh		*	*	*
tɾa tɾabahoh		**!		
ta tabahoh	*!			*
tɾa tabahoh	*!	*	*	

A parallel ranking holds for the foreign-segment examples. In this case, the Phono constraint[2] is a prohibition on the foreign segment, and the Uniformity constraints are those that require the correspondent of, for example, [θ], to be [θ]. Please note that I assume other, higher-ranking Uniformity constraints requiring that, for instance, the correspondent of a coronal be a coronal. This ensures that [θ] is nativized to [t], not, say [k].

In the tableau in (8), just as in (7), only [mag**tɛ**θænkju] and [mag**θɛ**θænkju] survive I-B Identity, since the base has [θ], and of these two, [mag**tɛ**θænkju] wins, since it has only one [θ]. Again, we will ignore the difference in vowels, since it may not be attributable to reduplication (see footnote 3).

(8) I-B Uni >> Phono >> {R-B Uni, I-R Uni}

[1] I-R Uniformity, although included in the tableaux for completeness, is ranked so low in Tagalog that it never has any effect. Sometimes that low ranking is crucial, sometimes the ranking is irrelevant.

[2] I use the term "Phono" to refer to phonotactic and markedness-type constraints, as opposed to correspondence constraints.

/mag+RED+θaŋkju/	I-BUni		R-B Uni	I-R Uni
-> mag**tɛ**θæŋkju		*	*	*
mag**θɛ**θæŋkju		**!		
mag**θɛ**tæŋkju	*!	*	*	
mag**tɛ**tæŋkju	*!			*

1.1 Variability in reduplication

There is, however, variability in reduplication of loanwords. [mag**tɛ**θæŋkju], [mag**θɛ**θæŋkju], and [mag**tɛ**tæŋkju] are all possible, as are [mag**ta**tɾabahoh] and [ma**gtɾa**tɾabahoh], although not *[mag**ta**tabahoh]. There is variability even within a speaker. This means that two other constraint rankings are also attested, at least for the phonotactic constraints which prohibit foreign segments.

First, the ranking that produces [mag**tɛ**tæŋkju]:

(9) Phono >> {I-B Uni, I-R Uni}
Note that the ranking of R-B Uniformity is irrelevant, since the winning candidate does not violate it, and the two candidates that do incur violations would be ruled out by *[θ] regardless of R-B Uniformity's ranking.

/mag+RED+θæŋkju/	*[θ]	I-B Uni	R-BUni	I-R Uni
mag**tɛ**θæŋkju	*!		*	*
mag**θɛ**θæŋkju	*!*			
mag**θɛ**tæŋkju	*!	*	*	
-> mag**tɛ**tæŋkju		*		*

In (9), the phonotactic ranks highest. Thus, only [mag**tɛ**tæŋkju], which never violates *[θ], survives, despite its Uniformity violations.

In (10), the phonotactic is outranked by I-B and R-B Uniformity. Note that here, the ranking of I-R Uniformity is completely irrelevant, since both candidates that earn violations of I-R Uniformity are ruled out by one of the other two Uniformity constraints. In this tableau, the candidate that best satisfies the Uniformity constraints (the most faithful parse), [mag**θɛ**θæŋkju], is the winner.

(10) {I-B Uni, R-B Uni} >> Phono

/mag+RED+θæŋkju/	I-BUni	R-B Uni	*[θ]	I-R Uni
mag**tɛ**θæŋkju		*!	*	*
-> mag**θɛ**θæŋkju			**	
mag**θɛ**tæŋkju	*(!)	*(!)	*	
mag**tɛ**tæŋkju	*!			*

One way to describe these three rankings is that *the Phono constraint is mobile.* That is, since speakers had no evidence prior to the introduction of the loanwords on which to decide where the constraint should be ranked, it's ranking is not fixed. In Case 1, [mag**tɛ**tæŋkju], the phonotactic is ranked high, in Case 2, [mag**tɛ**θæŋkju], it is somewhere in the middle, and in case 3, [mag**θɛ**θæŋkju], it is at the bottom.

(11)

	1: [mag**tɛ**tæŋkju]	2: [mag**tɛ**θæŋkju]	3: [mag**θɛ**θæŋkju]
	phonotactic	I-B	I-B, R-B, (I-R)
	I-B, (R-B), I-R	**phonotactic**	**phonotactic**
		R-B, I-R	

This is the only way to produce all three rankings by moving only one constraint. We can move the phonotactic without changing the one crucial relative ranking of Uniformity constraints: I-B Uniformity outranks R-B Uniformity. If we try to move any of the three Uniformity constraints, we will need to make other adjustments in addition:

It is not possible to move only I-B Uniformity, because in 1 and 2, the phonotactic crucially outranks R-B Uniformity, and in 3, this is reversed. Similarly, we cannot move either R-B or I-R Uniformity alone, since in case 1 the phonotactic outranks I-B Uniformity, and is cases 2 and 3 this is reversed.

Forms such as *[mag**θɛ**tæŋkju] and *[mag**tɾa**tabahoh], the third candidates in the tableaux, where R is more marked than B, are unattested. In order for such a form to be possible, I-R Uniformity would have to outrank the other three constraints[1]. The absence of this case is explained by my assumption that only the phonotactic constraint is mobile. Since I-B Uniformity crucially dominates I-R Uniformity in case 2, [mag**tɛ**θæŋkju], that ranking may not be reversed to yield the unattested *[mag**θɛ**tæŋkju].

At least for young, urban speakers, ranking 1 in (11), where the phonotactic is supreme, is impossible when the phonotactic is *ComplexOns. That is, as mentioned above, *[mag**ta**tabahoh] is impossible. So, the mobility of *ComplexOns is restricted. This is what we might expect: as large numbers of loanwords with the relevant property are incorporated into the lexicon, the ranking of the phonotactic constraint may stabilize, at a lower position, just as it is argued by Gnanadesikan to do in child language acquisition.

2. Infixation

McCarthy and Prince 1993 argue that the proper placement of the infixes /-um-/ and /-in-/ in Tagalog results from the avoidance of codas: the infix is as far to the left as possible given that no extra codas are created: No-Coda >> Leftmost:

(12) No-Coda: a violation is assessed for every syllbable having a coda[2]. (For clarity in

---

[1] McCarthy and Prince 1995 suggest a universal metaconstraint Root-faith >> Affix-faith, that is, Correspondence of roots to the input must always outrank Correspondence of affixes (including reduplicants) to the input. McCarthy and Prince propose that the metaconstraint is responsible for a universal ranking I-B Faithfulness >> I-R Faithfulness (in the terms used here, I-B Uni >> I-R Uni).

[2] Steriade 1995 provides phonetic motivation for both No-Coda and *ComplexOnset in terms of the importance of consonantal place cues in release bursts and in transitions to and from adjacent vowels.

these examples, only the violations involving the infix are shown)

Leftmost: a violation is assessed for every segment separating the left edge of the infix from the left edge of the word. (The segments themselves are listed in the tableaux.)

The following tableau illustrates how this ranking works for a native Tagalog root:

(13) No-Coda >> Leftmost

/**um**+sulat/ 'to write'	No-Coda	Leftmost
um.sulat	*!	
-> s**u.m**ulat		s
sul**u.m**at		su!l

The first candidate, [**um**.sulat], where the infix is perfectly leftmost, creates an extra coda and so is ruled out. The winning candidate has the infix as far to the left as possible without creating such an extra coda.

(14) repeats McCarthy and Prince's (222) (page 121) showing how /um/ is infixed into the English loan root /gɾadwet/, 'graduate'. Simply applying the same ranking as for the native root, we see that the infix is placed after the entire onset, rather than immediately after the the first consonant.

(14) No-Coda >> Leftmost

/**um**+gɾadwet/	No-Coda	Leftmost
um.gɾadwet	*!	
g**um**.ɾadwet	*!	g
-> gɾ**u.m**adwet		gɾ
gɾadw**u.m**et		gɾa!dw

The two candidates with the infix nearest the left edge are ruled out because an additional coda is created by the /m/ of the infix, and so the infix must be inserted after the entire onset.

In order to simplify matters, (15) repeats the tableau for a word in which the input root contains no additional internal clusters , /tɾip/, 'trip':

(15) No-Coda >> Leftmost

/**in**+tɾip/	No-Coda	Leftmost
in.tɾip	*!	
t**in.**ɾip	*!	t
-> tɾ**i.n**ip		tɾ
tɾi.p**in**		tɾi!p

The constraint *ComplexOnset, which was involved in reduplication, plays a role here, too, since some of the candidates in (14) and (15) violate it and some do not. No-Coda would have to outrank *ComplexOnset in order to prevent the second candidates, [g**um.**ɾadwet] and [t**in.**ɾip], from winning:

(16) No-Coda >> {*ComplexOns, Leftmost}

/**in**+tɾip/	No-Coda	*ComplexOns	Leftmost
in.tɾip	*!	*	
t**in.**ɾip	*!		t
-> tɾ**i.n**ip		*	tɾ
tɾi.p**in**		*	tɾi!p

Now note some characteristics of these constraints in Tagalog. Leftmost is violated in Tagalog regardless of the presence of loanwords, as we saw in [s**um**ulat]. Similarly, No-Coda, in addition to playing a role in forcing violations of Leftmost, is violated in many native Tagalog words, such as /lagnat/ 'fever', /saktan/ 'beat'; /kidlat/ 'lightning', which all contain both intervocalic CC clusters and final consonants. Thus, before the introduction of either Spanish or English loanwords, the ranking No-Coda >> Leftmost was thoroughly established, as was the ranking 'DoNotDeleteSegements'[1] >> No-Coda, which allows codas to appear when they could be eliminated by deletion. **ComplexOnset, on the other hand, was not violated in Tagalog prior to the introduction of Spanish loanwords*.

Given that *ComplexOnset played no role in pre-Spanish Tagalog phonology, since there were no inputs which, if faithfully parsed, would yield complex onsets, and there were no other constraints forcing complex onsets to occur, how did Tagalog speakers know that *ComplexOnset should be outranked by No-Coda? The data on variation in the next section show that they *did not* know this.

In the case of reduplication, we observed that the phonotactic which was unviolated in the native vocabulary was mobile—that is, that its ranking varied. We would then predict for infixation that the ranking of *ComplexOnset will vary, as indeed it does in the data in Section 2.1.

2.1 Variability in infixation

There is variability in infixation[2]. [gɾ**um**adwet]/[tɾ**in**ip] and [g**um**ɾadwet]/[t**in**ɾip] both occur:

---

1 That is, the Max of Correspondence Theory, which requires a segment of the input to correspond to a segment of the output.

2 For some onsets, some speakers, including speaker TA, who was not interviewed like the other speakers, but whom I frequently consult, express a preference to split the cluster rather than infix after it (ʃ**umɹ**iŋk, ʃ**umɹ**ibɛl, θ**umɹ**o, f**um**lo, t**inɹ**ip). This suggests that in addition to *ComplexOnset, there are feature-sensitive constraints which particularly disprefer certain clusters.

(16)

(a) Soberano 1992's NP:

	#C-C-infix	#C-infix-C	ʔ-inifix-isC[1]
plagiarize:	pl**in**ædʒɛɾajs	p**in**lædʒɛɾajs	
predispose:	pɾ**in**idɪspos	p**in**ɾidɪspos	
blaspheme:	bl**in**æspim	b**in**læspim	
brutalize:	bɾ**in**utalajz	b**in**ɾutalajz	
transmigrate:	tɾ**in**ansmajgɾɛjt		
twitter:	tw**um**ɪtəɾ	t**um**wɪtəɾ	
dramatize:	dɾ**in**amatajs	d**in**ɾamatajs	
climatize:	kl**um**imatajs	k**um**limatajs	
	klinimatajs	kinlimatajs	
criticize:	kɾinitisajs	kinɾitisajs	
quantify:	kwinantipaj	kinwantipaj	
glamorize:	glinamoɾajs	ginlamoɾajs	
graduate:	gɾumadʒuejt	gumɾadʒuejt	
flagellate:	plinadʒɛlɛjt	pinladʒɛlɛjt	
fraternize:	pɾumatɛɾnajs	pumɾatɛɾnajs	
throttle:	tɾinatɛl		
thwart:	twinaɾt	tinwaɾt	
specify:	spinɛsipaj	sinpɛsipaj	ʔinispɛsipaj
stereotype:	stɪnɛɾiotajp	* sintɛɾiotajp	ʔinistɛɾiotajp
sketch:	?/* skinɛtʃ	* sinkɛtʃ	ʔiniskɛtʃ
smuggle:	* sminagəl	* sinmagəl	ʔinismagəl
sneeze[2]:	snumis	* sumnis	
snatch:	* sninætʃ	* sinatʃ	ʔinisnætʃ
slenderize:	slinɛndɛɾajs	sinlɛndɛɾajs	
swindle:	swinindəl	sinwindəl	* ʔɪŋɪswindəl
shrivel:	ʃumɾibɛl		

[1] As noted below in Section 2.2, Tagalog often has a Spanish-like prothetic vowel in /s/C-initial loanwords.

[2] English *sneeze* is a real loanword in Tagalog, but always has the verbal prefix /mag-/, not the infixes.

shrink:	ʃɾɪnink	ʃɪnɾink	

(b) Speaker BH[1]:

flow:		fumlo	
float:		fumlot	
flock:		fulɔk	
fry:		finɾaj	
frisk:		finɾɪsk	
throw:		tinɹo	

(c) Speaker MC:

throw:		θumɹo	
three:		θumɹi	
flow:	flumo		
float:	flumot		
flock:	flumok		
fry:		fumɹaj	
frisk:		fumɹisk	
spark:	spumaɹk		
spin:	spumin		
stand:			ʔumistæn
stare:	stumeɹ		
start:			ʔumistaɹt
scale:			ʔumɪskel
skin:			ʔumɪskɪn
smile:	smumajl		
smear:			ʔumɪsmiɹ
sneer:			ʔumɪsniɹ
snoop:			ʔumɪsnup
snare:			ʔumɪsneɹ
slink:		smulɪŋk (sic)	
slide:			ʔumɪslajd

[1] No /s/C clusters are shown for BH because he always prothecized.

slip: sumlɪp
shrink: ʃumɹɪŋk
shrivel: ʃumɹɪbəl

Thus, in addition to the ranking No-Coda>>{*ComplexOns, Leftmost}, the ranking *ComplexOns>>{No-Coda, Leftmost} is also attested:

(17) *ComplexOns >> {No-Coda, Leftmost}

/**in**+tɾip/	*ComplexOns	No-Coda	Leftmost
in.tɾip	*!	*	
t**in.**ɾip		*	t
-> tɾ**i.n**ip	*!		tɾ
tɾi.p**in**	*!		tɾip

The difference between the two rankings is the relative ranking of No-Coda and *ComplexOns.

2.2 /sCL/ clusters

There are Tagalog loanwords from English beginning in /s/-stop-liquid clusters. As with /s/-stop clusters, these are optionally preceded by prothetic /ʔi/, which in some well-established loanwords is required:

(18) [ski] ~ [ʔiski] 'ski'
[stap] ~ [ʔistap] 'stop'
[ʔislajs] 'slice'
[ʔisketiŋ] 'skate'
[ʔisplɪt] 'split (as in banana)'
[skɾæts] ~ [ʔiskɾæts] 'scratch'
[ʔistɾakʔawt] 'strike out'

I will assume that there are thus two types of base forms available for infixation, the protheticized and the unprotheticized. The protheticized bases infix in the same way as native Tagalog /ʔV/-initial roots (e.g. [ʔ**um**islajs], [ʔ**um**isplɪt]), and so are not of interest in this section. This section will consider instead the unprotheticized forms only[1].

There should be three possible sites for infixation, but only two are attested:

(19)	s-C-C-infix	s-C-infix-C	s-infix-C-C
(a) Soberano 1992's NP:			
splutter:	splinatəɾ	spinlatəɾ	* sinplatəɾ
sprinkle:	spɾinɪŋkəl	spinɾiŋkɛl	
strangulate:	stɾɪnaŋgjulɛjt	stɪnɾangjulɛjt	* sɪntɾaŋgjulɛjt
scrutinize:	skɾinutinajs	skinɾutinajs	
squander:	skwinandɛɾ	skinwandɛɾ	* sinkwandəɾ
(b) Speaker MC			
splotch:		spumlotʃ	
split:		spumlɪt	
sprout:		spumɾawt	
spray:		spumɾe	
stripe:		stumɹajp	
stride:		stumɹajd	
scrimp:		skumɹɪmp	

Let us consider the constraint rankings that would be responsible for each variant. First, though, we must explain why [sC**um.**LV...] should ever be preferred over *[s**um.**CLV...]. The two have the same number of codas and the same number of complex onsets, but the unattested candidate has the infix further to the left, which should count in

[1] Speaker BH always prothecizes, so there are no data from him in (19)

its favour. I propose, following Steriade 1995a and 1995b, that an obstruent without an adjacent vowel is disfavored, because of the lack of transitional place cues. A strident such as /s/ without an adjacent vowel is also disfavored, but to a lesser degree, since the internal place cues of a strident are strong.

(20)

(a) *Non-V-Adjacent [-strident] Obstruent: One violation is incurred for every [-strident] without an adjacent vowel—that is, in the contexts C_C, #_C, or C_# (abbreviated C=V-Adj).

(b) *Non-V-Adjacent [+strident] Obstruent: One violation is incurred for every [+strident] without an adjacent vowel—that is, in the contexts C_C, #_C, or C_# (abbreviated s=V-Adj).

Again following Steriade, I assume a universal ranking C=V-Adj >> s=V-Adj.

C=V-Adj and s=V-Adj, like *ComplexOnset and the constraints against foreign segments, was irrelevant in Tagalog prior to the introduction of Spanish and English loanwords. There were no morpheme-internal CCC, #CC, or CC# sequences. Therefore, we should expect that the rankings of C=V-Adj and s=V-Adj, like the rankings of *ComplexOnset and the constraints against foreign segments, should be uncertain. This is borne out by the attested variants.

First, [sCL**in**...]/[sCL**um**...]. For concreteness, /-um-/ is illustrated. For this candidate to prevail, avoidance of Codas must be the first priority. Note that in all the tableaux for this input, the ranking of *ComplexOns is irrelevant, since each of the candidates violates it once. *ComplexOns will therefore be omitted from the tableaux for clarity.

(21) No-Coda >> { C=V-Adj, s=V-Adj, Leftmost }

/sCLV.../	No-Coda	C=V-Adj	s=V-Adj	Leftmost
um.sCLV...	*!	C	s	
s**um.**CLV...	*!	C		s
sC**um.**LV...	*!		s	sC
-> sCL**u.m**V...		C	s	sCS

Second, [sC**um**LV...]/[sC**in**LV...]. For this candidate to win, the first priority must be to avoid a consonant without an adjacent vowel.

(22) C=V-Adj >> { No-Coda, s=V-Adj, Leftmost }

/sCLV.../	C=V-Adj	No-Coda	s=V-Adj	Leftmost
um.sCLV...	C!	*	s	
s**um.**CLV...	C!	*		s
->sC**um.**LV...		*	s	sC
sCL**u.m**V...	C!		s	sCS

The difference between the two attested rankings is the relative ranking of C=V-Adj and No-Coda.

If Leftmost were top-ranking, the unattested *[ʔ**um.**sCLV...]/[ʔ**in.**sCLV...] would be selected. Notice that for unattested *[s**um.**CLV...]/[s**in.**CLV...] to be selected, s=V-Adj would have to outrank C=V-Adj, which, as mentioned above, is ruled out universally.

To summarize the attested rankings for infixation from this and the previous section, note first that there is no variation in the position of infixation in native roots. In the example (13), only [s**um**ulat] is attested. This means that No-Coda must outrank Leftmost, and this ranking was well-established in Tagalog prior to the introduction of cluster-initial loanwords. Given No-Coda>>Leftmost, we have the following attested rankings:

(23) ***ComplexOns**
|
No-Coda
|
Leftmost

No-Coda
/ \
***ComplexOns** Leftmost

(24) **C=V-Adj**
/ |\
No-Coda , s=V-Adj
|
Leftmost

No-Coda
/ \
C=V-Adj Leftmost
|
s=V-Adj

As before in (11), in both (23) and (24), if we wish to characterize the range of possible rankings by allowing just one constraint to be freely ranked, it must be the constraint whose ranking has only recently become relevant in the language (*ComplexOns in (23) and C=V-Adj in (24)). Although s=V-Adj should also be freely ranked, as long as C=V-Adj >> s=V-Adj, the ranking of s=V-Adj relative to No-Coda and Leftmost is indeterminate in the infixation cases, so its ranking is not discussed here. If we allowed No-Coda to be freely ranked, we would obtain *[**um**sulat]-type outputs for native roots when No-Coda was outranked by Leftmost , and if we allowed Leftmost to be freely ranked, we would obtain *[**um**sulat], *[**um**gɾadwet], and *[**in**splatɛɾ] when Leftmost was ranked topmost.

3. Diachrony

We have now seen cases of variability in Tagalog loanwords which can be explained by allowing previously unviolated constraints to be mobile. At an earlier stage of the language, these constraints were never put to the test; there was no input that challenged them, and so their ranking was unknowable. With the introduction of loanword inputs that, if faithfully parsed, would violate these constraints, their rankings are exposed and found to be variable.

What is the future of this variable-ranking situation? Tagalog has many loanwords from Chinese and Sanskrit, which have been in the language for much longer than the English and Spanish loanwords, and we can find some clues in the current state of these Sanskrit and Chinese loanwords in Tagalog. I will focus on two foreign elements introduced by these words: first, what is transcribed as [c] in Yap 1980 and in Manuel 1948 from Chinese. I assume this to be either a palatal stop or an alveolar affricate.

Chinese [c] was generally borrowed into Tagalog as [ts] intervocalically, and [s] word-initially, as in /**s**akjaʔ/, /twa**ts**at/, and /**s**u**ts**wa/. It would seem that [ts] is being treated as two segments: a CC cluster, which is allowed intervocalically, but not word-initially:

(25) [c] from Hokkien Chinese (Yap 1980, Manuel 1948)

sakjaʔ	'wooden clogs'	> **c**$^{\mathbf{h}}$a k^{h}iaʔ
twa**ts**at	'deceive'	> tua **c**$^{\mathbf{h}}$at
su**ts**wah	'medicinal/straw paper'	> **c**$^{\mathbf{h}}$o **c**ua

This is not always the case, though. We see in (34) that sometimes [c] is borrowed as [ts] word-initially, as in /**ts**ambwah/, suggesting it is being treated as a single segment. Sometimes, [c] is borrowed as [ts] after [n], producing a nasal-C-C cluster as in /gin**ts**am/. But sometimes, it is borrowed as [s] after [n], as in /kin**s**eh/.

(26)

tsambwah	'gold/silver shavings'	> **c**am bua	
gin**ts**am	'chisel for gold'	> gin **c**am	versus
kin**s**eh	'cut of beef for soup'	> kien **c**i	

Chinese aspirated stops are usually deaspirated in Tagalog, as in /wayu**k**ak/ and /**t**uwah/, below in (27). They are sometimes, however, C-h clusters intervocalically (where ‘vocalic’ includes glides), as in /bi**th**ay/ and /li**th**aw/.

(27)

wayu**k**ak	'container used in welding'	> oã iu $\mathbf{k^h}$ak
tuwah	'smiths' tool drawer'	> $\mathbf{t^h}$uaʔ
bi**th**ay	‘flat sieve’	> bi $\mathbf{t^h}$ai
li**th**aw	‘plow’	> le $\mathbf{t^h}$au

In Sanskrit loanwords in Tagalog, there are many C-h clusters intervocalically, as in /kasu**bh**ah/ and /mu**kh**aʔ/. Word-initially, aspiration either is lost, as in /**d**upa/, or appears elsewhere in the word as [h], as in /ba**th**alah/:

(28) from Francisco 1973

kasu**bh**ah	'safflower'	> kusum**bh**a
mu**kh**a	'face'	> mu**kh**a
dupah	'incense'	> **dh**ūpa
bat**h**alah	‘God’	> **bh**attāra

Some new elements, [c] and aspirated stops, were introduced as [ts] and stop-h cluster respectively, but not consistently, and they were particularly dispreferred in certain positions (nonintervocalic, in this case). Treatment of the foreign elements varies lexically: in some words, the foreign element is present, in others, it is absent.

What is the status of these elements in the grammar? Clearly, they are now both fully tolerated intervocalically, although in some loans where we might expect them, they are now simply not present in the input. Word-initially, stop-h clusters are not possible, and speaker-to-speaker variability remains in the treatment of [ts], as seen in these Spanish loanwords; /tsitsaɾon/ or /sitsaɾon/ from Spanish *chicharron*, and /tsinelas/ or /sinelas/ from Spanish *chinelas*.

(29)

tsi**ts**aɾon ~ **s**i**ts**aɾon	'pork rind'	> **tʃ**i**tʃ**aron
tsinelas ~ **s**inelas	'slippers'	> **tʃ**inelas

This, then, is my prediction for the future of complex onsets, CCC clusters, and new phonemes from Spanish and English, which we have seen to behave variably: In certain lexical items, the input will be fossilized without the foreign segment or sequence, if input from foreign speakers and fluent bilinguals ceases while the relevant Phono constraint still ranks very high, so that monolingual speakers hearing the word second-hand (with Phono unviolated) have no reason to believe that the input contains the foreign segment or strucutre. In other lexical items, widespread foreign input will continue at least until after the Phono constraint has sunk low enough so monolinguals hearing the word second-hand encounter the foreign segment or structure and include it in their input representation. For this second group of lexical items, the ranking of the Phono constraint will vary, and its upper range may become restricted, as we saw with *ComplexOnset, which has fallen low enough that *[magtatabahoh] is impossible for most speakers.

Part II: A Mechanism for Constraint Drift

Although empirical verification of detailed diachronic predictions is difficult, some support can be drawn from a computational illustration. In this section, I will propose a simple algorithm for evaluating a set of possible constraint rankings in the face of variable evidence. I will show that repeated application of this algorithm to a grammar with an initially top-ranked variable constraint results, over time, in an increasingly stable grammar, with the variable constraint tending to be ranked near the center. In addition, I will provide a mechanism—confusion between loanwords and code-mixing—whereby the probability that a variable constraint is top-ranked can be reduced drastically.

4. Assumptions

The algorithm operates on the following assumptions:

(30)

(a) Speakers differentiate between fixed constraints and floating constraints.

(b) Speakers are aware of the percentages of tokens consistent with different constraint rankings in the ambient speech community.

(c) The relative rankings of fixed constraints are absolute.

(d) A speaker chooses, in a given utterance, among all the overall constraint rankings which meet the minimum requirement of the fixed constraints' being correctly ranked.

(e) A value is associated with each pairwise subranking of two floating constraints or a fixed and a floating constraint. This value is based on the percentage occurrence in the ambient language of tokens which obey the pairwise subranking.

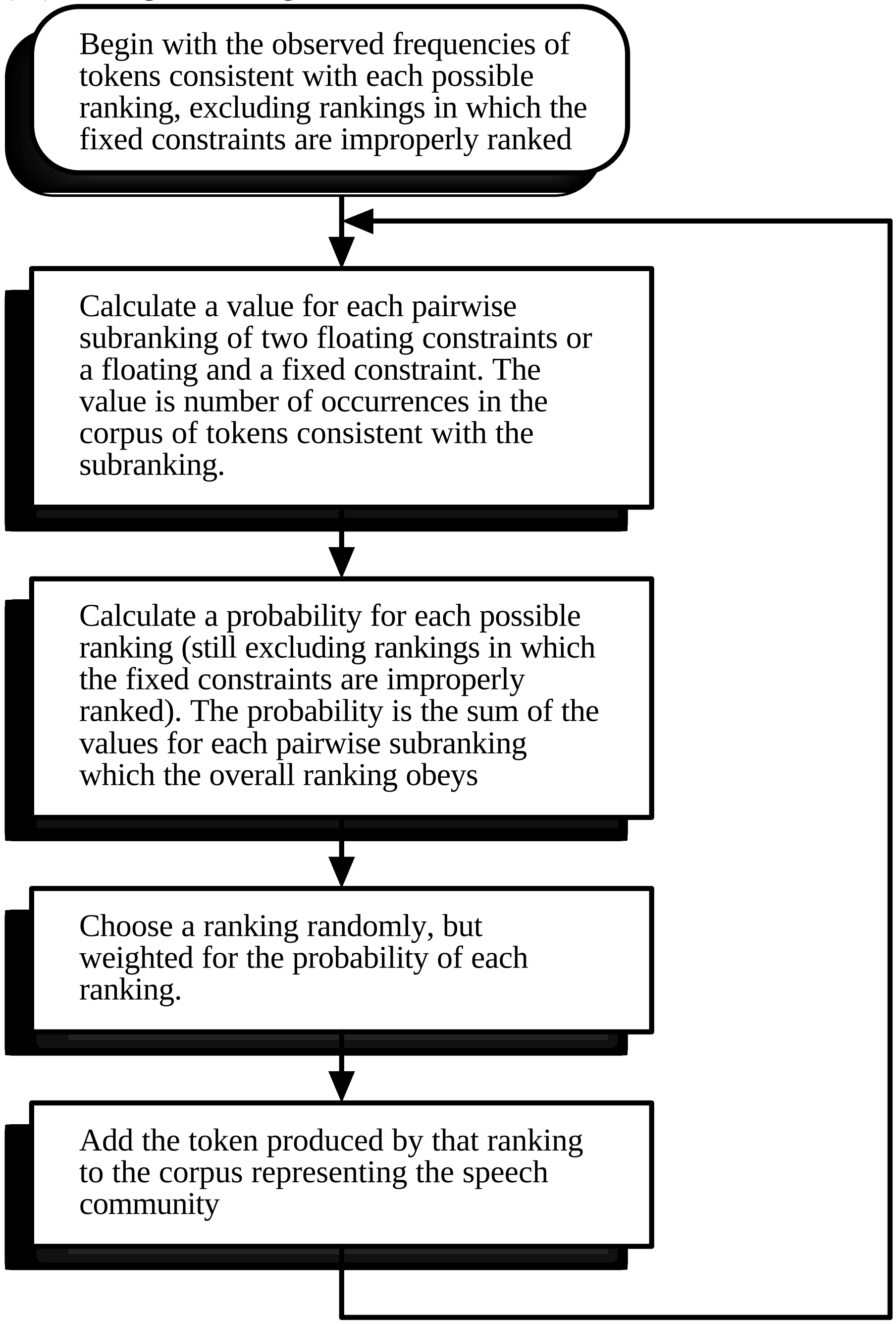
(f) The probability of an overall constraint ranking's being used in a given utterance is based on the values associated with each pairwise subranking which the overall ranking contains.

(g) When a speaker uses a ranking and utters a token, this obviously contributes to the distribution of tokens consistent with various constraint rankings in the speech community.

5. Algorithm

The following flowchart illustrates the mechanism by which change over time in the distribution of token types in the speech community changes over time. Each iteration of the looped portion represents an utterance by a speaker, not necessarily the same speaker every time. Formal definitions are given in Appendix B.

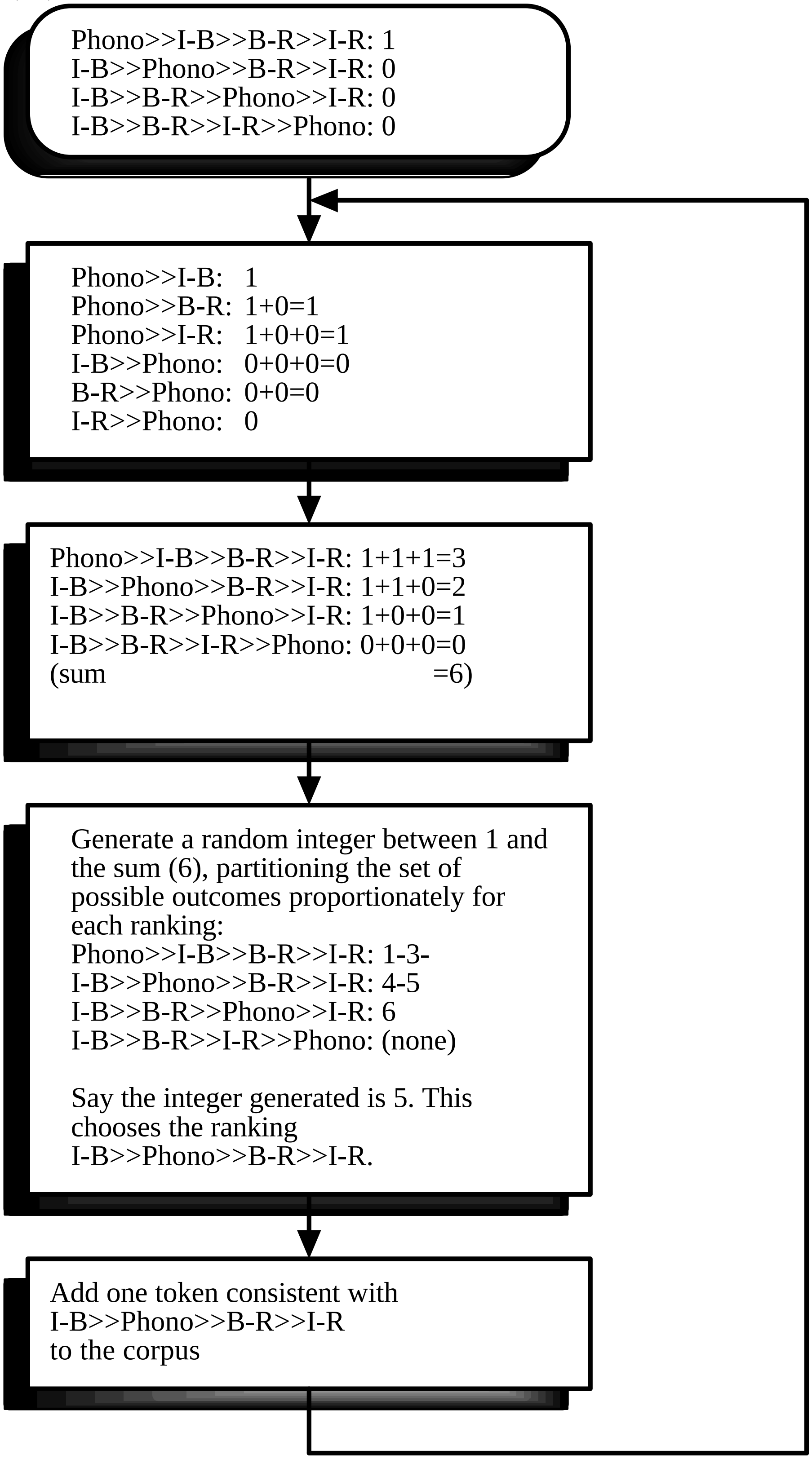
(31) The algorithm in general



6. Numerical results

This algorithm can be applied to the Tagalog cases. First, the reduplication cases, in which a Phono constraint is variably ranked with respect to the Correspondence constraints: Assuming that the initial state is one existing token consistent only with the Phono constraint being top-ranked, (32) illustrates the first pass through the loop:

(32)


Phono>>I-B>>B-R>>I-R: 1
I-B>>Phono>>B-R>>I-R: 0
I-B>>B-R>>Phono>>I-R: 0
I-B>>B-R>>I-R>>Phono: 0
Phono>>I-B: 1
Phono>>B-R: 1+0=1
Phono>>I-R: 1+0+0=1
I-B>>Phono: 0+0+0=0
B-R>>Phono: 0+0=0
I-R>>Phono: 0
Phono>>I-B>>B-R>>I-R: 1+1+1=3
I-B>>Phono>>B-R>>I-R: 1+1+0=2
I-B>>B-R>>Phono>>I-R: 1+0+0=1
I-B>>B-R>>I-R>>Phono: 0+0+0=0
(sum =6)
Generate a random integer between 1 and the sum (6), partitioning the set of possible outcomes proportionately for each ranking:
Phono>>I-B>>B-R>>I-R: 1-3-
I-B>>Phono>>B-R>>I-R: 4-5
I-B>>B-R>>Phono>>I-R: 6
I-B>>B-R>>I-R>>Phono: (none)
Say the integer generated is 5. This chooses the ranking I-B>>Phono>>B-R>>I-R.
Add one token consistent with I-B>>Phono>>B-R>>I-R to the corpus

In detail, begin by assuming that there has been one utterance in the speech community so far, and it is consistent with the ranking where Phono is topmost, and not with any of the other three possible total rankings (e.g. [mag**tɛ**tæŋkju]). In the first pass of the loop, values are calculated for each pairwise subranking as shown in the second box, where the pairs consistent with the one observed token each receive a score of 1, and the other pairs receive a score of 0. Then, raw probabilities (raw because they are not scaled to sum to 1) are calculated. Each total ranking is assigned the sum of the values for each pair it obeys, as in the 3rd box. Next, one total ranking is chosen, randomly but weighted for probability. If the ranking chosen is I-B>>Phono>>B-R>>I-R (the second most likely ranking on this pass), then one token consistent with that ranking is added to the corpus.

On the second pass, we have

(33)

	# of tokens:
Phono>>I-B>>B-R>>I-R	1
I-B>>Phono>>B-R>>I-R	1
I-B>>B-R>>Phono>>I-R	0
I-B>>B-R>>I-R>>Phono	0

	value
Phono>>I-B	1
Phono>>B-R	1+1=2
Phono>>I-R	1+1+0=2
I-B>>Phono	1+0+0=1
B-R>>Phono	0+0=0
I-R>>Phono	0

	probability (unscaled)
Phono>>I-B>>B-R>>I-R	1+2+2=5
I-B>>Phono>>B-R>>I-R	2+2+1=5
I-B>>B-R>>Phono>>I-R	2+1+0=3
I-B>>B-R>>I-R>>Phono	1+0+0=1

Now, the first two total rankings are equally likely, and the fourth is now possible, although it is the least likely. Suppose that the second is chosen again, adding one more token:

(34)

	# of tokens:
Phono>>I-B>>B-R>>I-R	1
I-B>>Phono>>B-R>>I-R	2
I-B>>B-R>>Phono>>I-R	0
I-B>>B-R>>I-R>>Phono	0

	value
Phono>>I-B	1
Phono>>B-R	1+2=3
Phono>>I-R	1+2+0=3
I-B>>Phono	2+0+0=2
B-R>>Phono	0+0=0
I-R>>Phono	0

	probability (unscaled)
Phono>>I-B>>B-R>>I-R	1+3+3=7
I-B>>Phono>>B-R>>I-R	3+3+2=8
I-B>>B-R>>Phono>>I-R	3+2+0=5
I-B>>B-R>>I-R>>Phono	2+0+0=2

Now, the second total ranking is the most likely. To illustrate one more pass, suppose that the first ranking is chosen this time. Then, we have:

(35)

	# of tokens:
Phono>>I-B>>B-R>>I-R	2
I-B>>Phono>>B-R>>I-R	2
I-B>>B-R>>Phono>>I-R	0
I-B>>B-R>>I-R>>Phono	0

	value
Phono>>I-B	2
Phono>>B-R	2+2=4
Phono>>I-R	2+2+0=4
I-B>>Phono	2+0+0=2
B-R>>Phono	0+0=0
I-R>>Phono	0

	probability (unscaled)
Phono>>I-B>>B-R>>I-R	2+4+4=10
I-B>>Phono>>B-R>>I-R	4+4+2=10
I-B>>B-R>>Phono>>I-R	4+2+0=6
I-B>>B-R>>I-R>>Phono	2+0+0=2

A simulation using a simple C program (Appendix A) runs the algorithm and gives the following results, shown for every tenth pass of the loop at first, and for every hundredth pass later:

(36)

Rankings	# of tokens at Stage...						
	Stage 0	10	20	30	40	50	60
Phono>>I-B Uni>>R-B Uni>>I-R Uni	1	6	8	9	10	12	15
I-B Uni>>Phono>>R-B Uni>>I-R Uni	0	3	6	10	12	15	18
I-B Uni>>R-B Uni>>Phono>>I-R Uni	0	1	5	9	12	15	17
I-B Uni>>R-B Uni>>I-R Uni>>Phono	0	1	2	3	7	9	11
	Stage 70	80	90	100	200	300	400
Phono>>I-B Uni>>R-B Uni>>I-R Uni	16	19	21	21	41	72	101
I-B Uni>>Phono>>R-B Uni>>I-R Uni	24	29	34	36	68	94	121
I-B Uni>>R-B Uni>>Phono>>I-R Uni	19	20	22	26	60	89	112
I-B Uni>>R-B Uni>>I-R Uni>>Phono	12	13	14	18	32	46	67
	Stage 500	600	700	800	900	1000	1100
Phono>>I-B Uni>>R-B Uni>>I-R Uni	122	139	159	174	190	209	240
I-B Uni>>Phono>>R-B Uni>>I-R Uni	153	185	210	234	264	294	321
I-B Uni>>R-B Uni>>Phono>>I-R Uni	140	167	200	237	264	299	327
I-B Uni>>R-B Uni>>I-R Uni>>Phono	86	110	132	156	183	199	213
	Stage 1200	1300	1400	1500	1600	1700	1800
Phono>>I-B Uni>>R-B Uni>>I-R Uni	261	282	308	331	352	376	394
I-B Uni>>Phono>>R-B Uni>>I-R Uni	247	380	412	438	474	497	525
I-B Uni>>R-B Uni>>Phono>>I-R Uni	360	388	413	444	467	497	527
I-B Uni>>R-B Uni>>I-R Uni>>Phono	233	251	268	288	308	331	355
	Stage						

	1900	2000	2100	2200	2300	2400	2500
Phono>>I-B Uni>>R-B Uni>>I-R Uni	420	440	463	483	505	543	564
I-B Uni>>Phono>>R-B Uni>>I-R Uni	552	586	617	650	680	709	742
I-B Uni>>R-B Uni>>Phono>>I-R Uni	555	584	606	630	655	683	707
I-B Uni>>R-B Uni>>I-R Uni>>Phono	374	391	415	438	461	486	509

	Stage 2600	2700	2800	2900	3000	3100	3200
Phono>>I-B Uni>>R-B Uni>>I-R Uni	587	613	638	656	676	697	720
I-B Uni>>Phono>>R-B Uni>>I-R Uni	773	806	861	889	916	950	974
I-B Uni>>R-B Uni>>Phono>>I-R Uni	733	761	818	853	892	918	948
I-B Uni>>R-B Uni>>I-R Uni>>Phono	531	547	584	603	617	636	659

	Stage 3300	3400	3500	3600	3700	3800	3900
Phono>>I-B Uni>>R-B Uni>>I-R Uni	738	763	783	805	828	850	863
I-B Uni>>Phono>>R-B Uni>>I-R Uni	993	1018	1046	1075	1102	1133	1162
I-B Uni>>R-B Uni>>Phono>>I-R Uni	992	1023	1059	1085	1119	1142	1178
I-B Uni>>R-B Uni>>I-R Uni>>Phono	678	697	713	736	752	776	798

	Stage 4000	4100	4200	4300	(variable size exceeded)
Phono>>I-B Uni>>R-B Uni>>I-R Uni	886	915	933	951	
I-B Uni>>Phono>>R-B Uni>>I-R Uni	1195	1215	1240	1273	
I-B Uni>>R-B Uni>>Phono>>I-R Uni	1196	1221	1257	1283	
I-B Uni>>R-B Uni>>I-R Uni>>Phono	824	850	871	894	

We can see from these results that, as an algorithm for faithfully learning the distribution of free variants in the ambient language, (31) is quite imperfect. From an initial state where 100% of tokens reflect just one ranking, running the algorithm results in a state where just 21.6% (951/4401) of tokens now reflect that ranking. But an imperfect algorithm is exactly what is required in order to achieve downward drift of the phonotactic constraint.

Similarly, modifying the program in Appendix A, we can run the algorithm for infixation:

(37)

	Stage							
	0	10	20	30	40	50	60	70
Phono>>No-Coda>>Leftmost	1	9	14	18	21	23	29	32
No-Coda>>Phono>>Leftmost	0	2	6	11	14	20	22	28
No-Coda>>Leftmost>>Phono	0	0	1	2	6	8	10	11

	Stage							
	80	90	100	200	300	400	500	600
Phono>>No-Coda>>Leftmost	36	42	43	78	118	156	189	214
No-Coda>>Phono>>Leftmost	33	36	41	87	127	167	212	256
No-Coda>>Leftmost>>Phono	12	13	17	36	56	78	100	131

	Stage							
	700	800	900	1000	1100	1200	1300	1400
Phono>>No-Coda>>Leftmost	245	300	327	368	398	429	467	502
No-Coda>>Phono>>Leftmost	295	337	378	431	469	512	559	592
No-Coda>>Leftmost>>Phono	161	192	223	243	264	291	316	342

	Stage							
	1500	1600	1700	1800	1900	2000	2100	2200
Phono>>No-Coda>>Leftmost	530	562	591	629	658	692	725	756
No-Coda>>Phono>>Leftmost	680	720	759	798	847	883	923	960
No-Coda>>Leftmost>>Phono	391	419	51	474	496	526	553	585

	Stage							
	2300	2400	2500	2600	2700	2800	2900	3000
Phono>>No-Coda>>Leftmost	781	814	846	880	916	948	976	1010
No-Coda>>Phono>>Leftmost	999	1038	1078	1119	1154	1200	1245	1289
No-Coda>>Leftmost>>Phono	621	649	677	702	731	753	780	802

	Stage						
	4000	5000	6000	7000	8000	9000	9400

Phono>>No-Coda>>Leftmost	1295	1620	1926	2242	2546	3776	2945
No-Coda>>Phono>>Leftmost	1724	2113	2523	2935	3362	3776	3952
No-Coda>>Leftmost>>Phono	1082	1368	1652	1924	2193	2489	2604

Again, we see that gradually, Phono becomes most likely to be ranked somewhere in the middle.

7. The algorithm in contact situations

7.1 Hearer confusion

Given that the system will tend to stabilize with a center-ranking of the Phono constraint, how does top-ranking of the Phono constraint ever become impossible? I propose that speakers are not able to distinguish reliably between, on the one hand, the use of an unaffixed loanword by a fellow member of the speech community with a high ranking of Correspondence constraints and, on the other hand, code-mixing by that fellow member of the speech community. Both kinds of token would sound the same: they would sound like the foreign word. Frequent and sustained mistaking by individual speakers of code-mixing for loanword use results in an artificial inflation of the perceived incidence of tokens obeying a low ranking of Phono, which in turn results in an artificial inflation of the values associated with pairwise rankings of the form Uni>>Phono and deflation of the values for Phono>>Uni, which finally results in an artificial skewing of probabilities towards rankings with Phono on the bottom and away from rankings with Phono on top.

7.2 Insufficient Borrowings

Bruce Hayes (p.c.) points out that despite the large number of French words and phrases introduced into English, many of them continue to be perceived by English speakers as foreign (e.g. *joie de vivre*, *gendarmes*), and the English Phono constraints that they violate remain high-ranking in English.

The reason for this is that the language contact situation of English and French is very different from that of Tagalog and English. In the case of French loans in English, the algorithm simply has not iterated enough to affect English phonology. These foreign-seeming French loans are usually encountered in print by most speakers, so that the input must be constructed by those speakers based only or mostly on orthography. When the loans are heard in speech, they are rare enough that they may be construed by hearers as simply code-mixing (rather than loans), especially since no English affixes are applied to them.

In Tagalog, by contrast, because of the presence of many English speakers and English-Tagalog bilinguals and the pervasiveness of English-language movies, radio, television, and music, there are ample spoken examples of English loans which speakers can use to construct input representations. In addition, because of the rich morphology of Tagalog, a high percentage of loanword tokens will be affixed, so that they cannot be mistaken by hearers for mere code-mixing, and thus the tokens will be added to the corpus used by the algorithm, rather than ignored.

8. Conclusion

This paper shows in Part I that there is variation in the reduplication and infixation of loanwords in contemporary Tagalog. This variation can be described in terms of indeterminate ranking of certain Phono constraints. The Phono constraints whose rankings are variable are just those for whose ranking the speech community has only recently had evidence. Part II proposes a mechanism by which such variability stabilizes and by which certain rankings can cease to occur.

The evidence necessary to refine this model would come from careful sociolinguistic study, to determine whether there are consistent correlations of constraint ranking with such factors as age, socioeconomic status, and degree of bilingualism of the speaker, as well as formality of the speech situation. If such correlations exist, then (31), which is indifferent to the prestige or group-identification value of any ranking, would require modification to reflect the relevant sociolinguistic pressures.

Appendix A: C Program

```
#include	<stdlib.h>

/*********This is for the reduplication example (3 fixed constraints)********/

/**************RollOne*************Generates a random number between 1 and
increment*/
RollOne( int increment )
{
	long	rawResult;
	unsigned int		roll;

	rawResult = rand();

	roll = ( rawResult * increment ) / 32768;

	return( roll + 1 );
}

/*****************Flush***********Flushes the input stream*/
void	Flush()
{
	while( getchar() != '\n' )
		;
}

/***************GetCommand*******Allows the user to continue or stop*/
char	GetCommand()
{
	char command = 0;

	while ( (command != 'q') && (command != 'n') )
	{
		printf( "Enter command (q=quit, n=a hundred new rounds): ");
		scanf( "%c", &command );
		Flush();
	}

	printf( "\n------------------\n" );
	return( command );
}
```

```
/**************main*****************/
main()
{
	char			command;
	int			i, j, result;
	unsigned int		tokens[ 4 ], rankings[ 4 ], pairs[ 6 ], increment, round,
random;

	round = 0;

	/**Set initial occurrences of tokens**/
	tokens[ 0 ] = 1;
	tokens[ 1 ] = 0;
	tokens[ 2 ] = 0;
	tokens[ 3 ] = 0;

	/**Print initial occurrences of tokens**/
	printf( " Incidence of A123 is %d	\n", tokens[ 0 ] );
	printf( " Incidence of 1A23 is %d	\n", tokens[ 1 ] );
	printf( " Incidence of 12A3 is %d	\n", tokens[ 2 ] );
	printf( " Incidence of 123A is %d	\n", tokens[ 3 ] );

	/**the loop**/
	while( (command = GetCommand() ) != 'q' )
	{
		for( j = 1; j <= 100; j++ )
		{
			round++;

			/**calculate value for each pairwise subranking**/
			pairs[ 0 ] = tokens[ 0 ];
			pairs[ 1 ] = tokens[ 0 ] + tokens[ 1 ];
			pairs[ 2 ] = tokens[ 0 ] + tokens[ 1 ] + tokens[ 2 ];
			pairs[ 3 ] = tokens[ 1 ] + tokens[ 2 ] + tokens[ 3 ];
			pairs[ 4 ] = tokens[ 2 ] + tokens[ 3 ];
			pairs[ 5 ] = tokens[ 3 ];

			/**calculate probability for each ranking**/
			rankings[ 0 ] = pairs[ 0 ] + pairs[ 1 ] + pairs[ 2 ];
			rankings[ 1 ] = pairs[ 1 ] + pairs[ 2 ] + pairs[ 3 ];
			rankings[ 2 ] = pairs[ 2 ] + pairs[ 3 ] + pairs[ 4 ];
			rankings[ 3 ] = pairs[ 3 ] + pairs[ 4 ] + pairs[ 5 ];
```

```
			/**calculate sum of probabilities*/
			increment = ( rankings[ 0 ] + rankings[ 1 ] + rankings[ 2 ] +
rankings[ 3 ] );

			/**generate a random number between 1 and increment**/
			random = RollOne( increment );

			/**choose a ranking based on the random number and the
probabilities**/
			if ( random <= rankings[ 0 ] )
				result = 0;
			else
			{
				if  ( random <= ( rankings[ 0 ] + rankings[ 1 ] ) )
					result = 1;
				else
				{
					if ( random <= ( rankings[0] + rankings[1] +
rankings[2] ) )
						result = 2;
					else
						result = 3;
				}
			}

			/**add one token of the appropriate type to the corpus**/
			tokens[ result ]++;
		}

		/**output every 100 rounds**/
		printf( "This is round %d. \n", round );

		printf( " Incidence of A123 is %d	\n", tokens[ 0 ] );
		printf( " Incidence of 1A23 is %d	\n", tokens[ 1 ] );
		printf( " Incidence of 12A3 is %d	\n", tokens[ 2 ] );
	}

	printf( "Goodbye..." );
}
```

Appendix B: Formalisms[1]

We begin with two finite and disjoint sets, E and N_0. *E* represents the set of existing (fixed) constraints, and N_0 represents the set of novel (free) constraints at the starting point of the algorithm. Since more novel constraints can be added at any time t, we can stipulate

$$N_t \supseteq N_{t-1} \qquad \text{for } t \geq 1$$

Constraint rankings will be defined in terms of pairwise subrankings. There are $\binom{|E|+|N_t|}{2} \cdot 2! = \frac{(|E|+|N_t|)!}{(|E|+|N_t|-2)!} = (|E|+|N_t|) \cdot (|E|+|N_t|-1)$ ordered pairs of distinct ($C_i \neq C_j$) constraints $<C_i,C_j>$ such that $C_i, C_j \in E \cup N_t$, but we are not interested in pairs of fixed

constraints, since their respective rankings must remain constant. So, we will consider only the $\left(\binom{|E|+|N_t|}{2} - \binom{|E|}{2}\right) \cdot 2!$ or, equivalently, $\left(\binom{|N_t|}{2} + |N_t| \cdot |E|\right) \cdot 2!$ ordered pairs[2] of constraints $<C_i,C_j>$ such that $C_i \in N_t$ or $C_j \in N_t$ and $C_i \neq C_j$.

Now we can define constraint rankings. A constraint ranking will be defined as a subset of the set $S_t = \{<C_i,C_j> \mid C_i \text{ or } C_j \in N_t \text{ and } C_i \neq C_j\}$:

R is a ranking iff $R \subseteq S_t$ for some t, and

1) if $<C_i,C_j> \in R$ then $<C_j,C_i> \notin R$ (asymmetry)

2) if $<C_i,C_j> \in R$ and $<C_j,C_k> \in R$ then $<C_i,C_k> \in R$ (transitivity)

Keep in mind that there are finitely many rankings $R \subseteq S_t$ for any t (since E and N_t are finite in size). Thus, we will be able to refer to the rankings $R_{t,1}$, $R_{t,2}$, etc.

This subset of pairwise subrankings should be understood as the set of such subrankings which are obeyed in the overall constraint ranking R, where the first member of an ordered pair outranks the second.

---

1 The definitions in this appendix follow the algorithm of (31) closely, except that probabilities are scaled to sum to 1.

2 Both expressions are equal to $|N|^2 - |N| + 2|N| \cdot |E|$

For example, an overall ranking R

C_i>> ... >>C_1>> ... >>C_j>> ... >>C_2>> ... >>C_k

would obey the pairwise subranking C_1>>C_2, and so <C_1,C_2> R.
Thus, the constraint ranking

1>>A>>2>>3>>B>>4>>C

where A,B, and C are free constraints and 1,2,3,4 are fixed, can be represented as

{<1,A>,<1,B>,<1,C>,<A,2>,<A,3>,<A,B>,<A,4>,<A,C>,
<2,B>,<2,C>,<3,B>,<3,C>,<B,4>,<B,C>,<4,C>}

Working from the intuitive idea of constraint rankings, there are

$$\binom{|N_t|+|E|}{|N_t|} \cdot |N_t|! = \frac{(|N_t|+|E|)!}{|E|!\,|N_t|!} \cdot |N_t|! = \frac{(|N_t|+|E|)!}{|E|!}$$

possible rankings R for any time t.
That is, out of the total number of positions for available for constraints (E+N_t), choose the positions that will be occupied by free constraints. For each such selection of positions, there are then N_t! possible permutations of the free constraints.

Before going further, let us reassure ourselves that for any t 1, {R| R is a ranking and R ΔS_t }Γ{R| R is a ranking and R ΔS_{t-1}}. That is, that the set of rankings at time t is a superset of the set of rankings at time t-1. For exposition, let’s call those two sets A and B respectively:

(To show: if R B then R A)
We know, by definition of a Ranking, that if R B
1) R ΔS_{t-1}={<C_i,C_j> | C_i or C_j N_{t-1} and C_idC_j}
2) if <C_i,C_j> R then <C_j ,C_i>®R (asymmetry)

3) if <C_i,C_j> R and <C_j,C_k> R then <C_i,C_k> R (transitivity)

Since $N_t Γ N_{t-1}$ for t 1, if C $_i$ N_{t-1} then C_i N_t and if C_j N_{t-1} then C_j N_t so if C_i or C_j N_{t-1} then C_i or C_j $N_{t.}$ This fact, combined with clauses (2) and (3) yields R A.

We can now recursively define the two-place *occur*(probability) function, which returns an occurrence value for any constraint $RΔS_0$ at any time t 0.

occur(R,0)=$\frac{1}{|N_0|!}$ if {<C_i,C_j> R (C_j N_0 C_i N_0)

occur(R,0)=0 otherwise

That is, we begin by evenly dividing a total of 1 among the $|N_0|$! rankings in which no fixed constraint (member of E) outranks any free constraint (member of N_0). All other rankings are assigned the value 0.

For t>0, we must introduce another two-place function, *val* (value), defined for any pairwise subranking P S_t, at any time t 1, t ¢.

$$\text{val}(P,t) = \sum_{P \, 0 \, R_i \, d \, S_{t-1}} \text{occur}(R_i, t-1)$$

Val assigns a value to every P, based on the occurrences of the rankings $RΔS_t$ which contain P.

Note that, for t-1=0, we have given by stipulation the values of the *occur* function for every ranking $RΔS_0$. So, we have also thus defined the *val* function for t= 1. *Occur* for t 1 will depned on the value of *prob*, the probability function.

We will now define *prob* for t 1in terms of *val*:

First we define a function *rawprob* (raw probability), for t 1, t ¢. Let T_t be the set of all rankings $RΔS_{t.}$

For any R T_t

$$\text{rawprob}(R,t) = \sum_{P_i \, 0 \, R} \text{val}(P_i, t)$$

Rawprob sums the values of the Ps that R obeys, and then *prob* divides the *rawprob* number by the sum of all *rawprobs* at that *t*, so that the probabilities will add up to 1.

For t 1, t ¢ and for any R T_t

$$prob(R,t) = \frac{rawprob(R,t)}{\sum_{R_i \in T_i} rawprob(R_i,t)}$$

Thus, *prob* is defined for any positive integer time *t*.

Occur for t 1 depends on *prob*, but not deterministically. Suppose that there is a function *rand*(t), which returns a random number in the interval (0, 1] for any time t. Then

$$occur(R_{t,n}, t+1) = occur(R_{t,n}, t) + \frac{1}{\sum occur(R, t)} \text{ if } \sum_{i=0}^{n-1} prob(R_{t,i}) < rand(t) \leq prob(R_{t,n})$$

$$occur(R_{t,n}, t) \text{ otherwise}$$

That is, the occurrence value of a given ranking is incremented iff that ranking was ‘chosen’ at the previous t, by having rand(t) fall within the interval assigned to that ranking, based on its probability value.

References

Benua, Laura (1995) "Identity effects in Morphological Truncation" in J. Beckman, S. Urbanczyk & L. Walsh, eds., *University of massachusetts Occasional Papers in Linguistics* 18: Papers in Optimality Theory, Graduate Linguistic Student Association, Amherst, Mass.

Crosswhite, Katherine (1995) "Base-Derivative Correspondence in Chamorro" UCLA MA thesis.

Flemming, Edward (1995) "Auditory Representations in Phonology" UCLA PhD dissertation.

French, Koleen Matsuda (1988) *Insights into Tagalog Reduplication, Infixation, and Stress from Nonlinear Phonology* SIL & University of Texas at Arlington, Arlington, Tex.

Francisco, Juan R. (1973) "Sanskrit in the Philippine Languages" In Andrew B. Gonzalez, Teodoro Llamzon, and Fe Otanes, eds. *Readings in Philippine Linguistics* Linguistic Society of the Philippines, Manila.

Gnanadesikan, Amalia E. (1995) "Markedness and Faithfulness Constraints in Child Phonology" ms., Rutgers University and Rutgers Optimality Archive.

Grimes, Barbara F., ed. (1988). *Ethnologue: Languages of the World.* 11th edition. SIL, Dallas, Texas.

Hayes, Bruce (1995) *Phonetically Driven Optimality Theoretic Phonology* UCLA proseminar.

Hayes, Bruce and Margaret MacEachern (1995) "Folk Verse Form in English," ms., UCLA, submitted to *Language*.

International Phonetic Association (1989) "Report on the 1989 Kiel Convention" *Journal* *of the International Phonetic Association*, pp. 67-80.

Jun, Jongho (1995) “Perceptual and Articulatory Factors in Place Assimilation: An Optimality Theoretic Approach.” UCLA PhD dissertation.

Manuel, Arsenio E. (1948) *Chinese Elements in the Tagalog Language* Filipiniana Publications, Manila.

McCarthy, John J. and Alan Prince (1993) “Prosodic Morphology I,” ms, University of Massachusetts, Amherst and Brandeis University.

McCarthy, John J. and Alan Prince (1994) “The Emergence of the Unmarked: Optimality in Prosodic Morphology” ms., UMass, Amherst and Rutgers.

McCarthy, John J. and Alan Prince (1995) “Faithfulness and Reduplicative Identity” In Jill Beckman, Laura Walsh Dickey & Suzanne Urbanczyk, eds. *University of Massachusetts Occasional Papers in Linguistics* 18. GLSA, Amherst, Mass, pp. 249-383.

Mester, Armin (1996) “Correspondence, Compositionality, and Free Ranking” ms., UC Santa Cruz.

Quilis, Antonio (1985) “A Comparison of the Phonemic Systems of Spanish and Tagalog” *Scientific and Humanistic Dimensions of Language: Festschrift for Robert Lado* Kurt R. Jankowsky, ed. Benjamins.

Silverman, Daniel (1995) “Phasing and Recoverability.” UCLA PhD dissertation

Schachter, Paul and Fe Otanes (1972) *Tagalog Reference Grammar*: University of California Press, Berkeley.

Soberano, Marinette (1992) student elicitation session, Lisa Travis, instructor, Natividad del Pilar, consultant, McGill University.

Steriade, Donca (1995a) “Positional Neutralization” ms., UCLA.

Steriade, Donca (1995b) “The Syllable” UCLA proseminar.

Steriade, Donca (1996) “Paradigm Uniformity and the Phonetics-Phonology Boundary” presented at the 5th Conference in Laboratory Phonology, Evanston, Ill.

Stockwell, Robert (1957) *A Contrastive Analysis of English and Tagalog.* UCLA Working Papers in Applied Linguistics.

Yap, Gloria Chan (1980) *Hokkien Chinese Borrowings in Tagalog* Pacific Linguistics Series B No. 17, Department of Linguistics, School of Pacific Studies, Australian National University, Canberra.