Indeterminate stress faithfulness in Tagalog
segmental vs. templatic, input vs. base

Kie Zuraw, kie@ucla.edu

8 September 2012, Phonology-Morphology Circle of Korea

(1) Overview/outline

- Stress correspondence can be segment-by-segment or to a template
- Tagalog has been proposed to show segment-by-segment stress correspondence in suffixed nouns, but template correspondence in verbs (Sabbagh 2004)
- This talk argues:
  - Tagalog derivational reduplication shows variation between segmental and templatic correspondence
  - B-R correspondence and Output-Output correspondence also both play roles
- **What does it mean?**
  - If speakers have little data to tell them how correspondence should work, all options may be used.

Data mostly from a dictionary, Ferrer 2006 ("Fe"), otherwise Schachter & Otanes 1972, English 1986, my observations.

A. Stress background

(2) Basic stress in Tagalog

- If open penult, stress on ultima or penult: ba**sáʔ** 'wet'
  bása 'reading'
- If closed penult, stress on ultima only: ban**tás** 'punctuation'
  - Except in some loans: **bén**da 'bandage' (Sp.)
    and there, it's often unstable: **bán**.jo 'bathroom' (Sp. báɲo)
    but ban.**jós** 'sponge bath' (Fe)

(3) Suffixed verbs

Stress shifts one to the right:

ba**sáʔ**	'wet'	basaʔ-**ín**	'be wetted'	→ always ultima
bása	'read'	ba**sá**-hin	'be read'	→ always penult
ban**tás**	'punctuation'	banta**s-án**	'be punctuated' (Fe)	→ always ultima

Except in those exceptional loans—there it shifts all the way to the end, sometimes with secondary stress retained:

bénda 'bandage' benda-**hán** ~ bènda-**hán** 'be bandaged' (Fe)

(4) What a basic analysis needs to do

- Allow lexical entry to determine whether a root is final-stressing or penult-stressing
- Carry this behavior over onto suffixed verbs
- (Treat stressed, closed penult as exceptional—when suffixed, stressed becomes final)

I won't talk about this. Doesn't arise much in data here.

There are many ways to do this

- Some lexical entries invoke constraint TROCHAIC, others IAMBIC
- Or, some lexical entries carry a piece of grid structure like `x . ]PWd`
- Or...

For the sake of brevity I won't spell these out, because the details won't matter to the main point here about reduplicated words.

(5) Suffixed nouns are different from suffixed verbs (Sabbagh 2004)

final stress:	no shift	pa**lít**	'exchange'	pa**lít**-an	'trade by exch. goods'
penultimate stress:	shifts, to *final*	**há**wak	'grasp'	hawa**k-án**	'handle'

Sabbagh's analysis

- Stress wants to stay in the same place (segment-by-segment O-O correspondence is ranked high for nouns): *pal**í**t ~pal**í**t-an*
- But for stems with penultimate stress, this would produce an unacceptable final stress lapse: *h**á**wak ~*h**á**wak-an*
    - So default final stress applies instead: *hawak-án*
    - (Default final stress further supported by verbs whose stem has stressed, closed penult.)

(6) Local summary: we've seen two kinds of faithfulness/correspondence:

- Templatic correspondence: Corr. strings should have same right-aligned stress pattern: verb suffixation like ***bá**sa ~ bas**á**-hin* (both penultimate stress)
- Segment-for-segment correspondence: Corresponding segments should have the same value for [stress]: noun suffixation like *pal**í**t ~ pal**í**t-an*.

Templatic correspondence has been invoked before:

- Sabbagh 2004: Tagalog stress shifts under suffixation
- Shryock 1993: Cebuano stress shifts under suffixation
- Chung 1983: Chamorro syllable-shape shifts under suffixation
- Bagemihl 1989, data from Hombert 1973: in a Bakwiri language game, syllable weight stays in place though segments move around.
- Ito, Kitagawa, & Mester 1996: in a Japanese language game, under certain circumstances vowel length remains in place although segments switch places.

There are also parallels to this idea in psycholinguistics.

- E.g., in Levelt 1999's model of speech production, a lexical entry's prosody is stored separately from its segments/melody.

Now let's turn to reduplication, where the two types of correspondence can compete.

B. Reduplication: two types of correspondence

(7) Two-syllable reduplication basics

- Happens in various derivational morphology in Tagalog: moderates or intensifies, creates pluractionals, 'fake X', among other uses (see glosses to get a feel for it).
- Copies full stem (except final /ʔ/) if disyllabic
    - otherwise copies two syllables' worth but with no coda (CONTIGUITY, ANCHOR-L >> ANCHOR-R >> NOCODA)
- Pseudoreduplicated stems can't undergo (*tatak*, *daldal*).

You don't have to hold these in your head—just interesting to note.

(8) Simplest case: disyllabic base, no suffix

(All counts based on data from Ferrer 2006)

- σσ́: If base has final stress, 214/224 = 95% of cases have final stress in reduplicant too:
  pa**tíd**-pa**tíd** 'disjointed' (pa**tíd** 'break')
- σ́σ: If base has penultimate stress, 170/179 = 95% of cases have penult. stress in redup. too:
  jákap-**já**kap 'lovingly embraced' (**já**kap 'embrace')

⇨ Reduplicant is faithful to base's stress both segmentally and templatically

(9) Second-simplest case: longer base (suffixed or not) with final stress

- σσ́: In unsuffixed base with final stress, all 18 have final stress in reduplicant too:
  da**lá**-dala**wá** 'two by two' (dala**wá** 'two')
  sa**lí**-salimu**ʔót** 'tangled' (salimu**ʔót** 'entanglement')

- σσ-σ́: In suffixed base with final stress, all 39 have final stress in reduplicant too:
  ta**mís**-tami**s-án** 'to sweeten' (ta**mís** 'sweetness')
  ba**lá**-balansa**ŋ-ín** 'to upset'

⇨ Reduplicant is faithful to base's stress templatically, and/or reduplicant shows default final stress because it's impossible to be faithful to base's stress segmentally.

⇨ This is already 76% of the words in the dictionary, and they don't tell the learner which type of faithfulness matters.

(10) The interesting case: longer base (suffixed or not) with penultimate stress

- **σσ́σ**: In unsuffixed base with penultimate stress...
  - 15/32=47% have penultimate stress in reduplicant:
    dóse-do**sé**na 'by the dozen' (do**sé**na 'dozen')
  - 16/32=50% have final stress in reduplicant:
    bi**hí**-bi**hí**raʔ 'very rarely' (bi**hí**raʔ 'rarely')

  (1 item varies)

- **(σ)σσ́-σ**: In suffixed base with penultimate stress...
  - 69/111=62% have penultimate stress in reduplicant:
    ʔísip-ʔi**sí**p-in 'to consider' (**ʔí**sip 'thought')
    bútu-butu**ní**s-an 'plant sp.' (bu**tó**nes 'button') [only 4 like this]
  - 40/111=36% have final stress in reduplicant:
    di**lí**-di**lí**-hin 'to reflect'
    ta**bá**-taba**kú**-han 'plant sp.' (ta**bá**ko 'tobacco') [only 4 like this]

  (2 items do something else)

These cases are crucial: without templatic correspondence, there's no reason for [**dó**se-do**sé**na], [**bú**tu-butu**ní**s-an]

⇨ There is variation resulting from competition between reduplicant's being faithful to base's stress segmentally (final stress) or templatically (penultimate stress).

Now let's move on to the question of what form the reduplicant is faithful to: just the base, or other forms of the stem too?

(11) Lexical conservatism? (Steriade 1999; Steriade 1997)
Steriade proposes constraints that require an alternation to be faithful to *some* listed allomorph.

In the cases above, where there can be variation, does it matter what stress allomorphs are available in the paradigm to be faithful to?

- Suffixed **(σ)σσ́-σ** bases
  - Can have unsuffixed form σσ́ (most common), σσ́, or both
    How is it that we can get two stress patterns in unsuffixed forms? As in English, there are some part-of-speech stress alternations:

típon	'N: gathering'	ti**pón**	'Adj: gathered'
ʔútos	'N: command'	pa-ʔu**tós**	'Adj: imperative'

  - It seems to make a difference (total only 97 because some bases lack unsuffixed form):

		Redup. stress		
base is suffixed (σ)σσ́-σ: faithfulness conflict ▪ templatic correspondence → penult stress in redup. ▪ segmental correspondence → final stress in redup		σσ́-	σσ́-	% σσ- (penult stress)
only unsuffixed form is σσ	Verb	27	4	
▪ unsuffixed form further supports penult stress in redup	Noun	20	3	88%
	Adj	4	0	
σσ and σσ́ both exist	Verb	5	3	
	Noun	5	4	55%
	Adj	1	2	
only unsuffixed form is σσ́	Verb	2	5	
▪ unsuffixed form further supports final stress in redup	Noun	1	8	16%
	Adj	0	3	

⇒ Reduplicant looks not only to base, but to other unsuffixed forms too

- Unsuffixed σσσ cases—not relevant (only 32 cases anyway)
  - Nearly all have just unsuffixed σσσ, possibly suffixed σσσ́-σ in paradigm (and sometimes a σσσ-σ́ noun)—i.e., no σσσ

(12) So do we still need B-R correspondence?
Yes. Recall that with σσ-σσ words, base and reduplicant match 95% of the time:

patíd-patíd	'disjointed'	(patíd 'break')
jákap-jákap	'lovingly embraced'	(jákap 'embrace')

And yet, their unsuffixed forms would often support a different stress pattern (sample of 1/3 of these words):

	unsuffixed forms all σσ́	σσ and σσ́ both exist	unsuffixed forms are all σσ
σσ́ base e.g., patíd-patíd	75%	11%	13%
σσ base e.g., jákap-jákap	31%	11%	58%

C. Model

(13) Logistic regression

To check that these factors actually matter.

- Because some factor combinations predict the outcome perfectly, I'm using Gelman & al.'s (2012) bayesGLM(arm) function for R (R Development Core Team 2012).
- Dependent variable is whether reduplicant has penultimate stress
- "sw" = "strong-weak", or όσ; "ws" = "weak-strong", or σό

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	4.50746	1.05807	4.260	0.000020	***
sw_is_seg_unfaith_to_base	**-4.20815**	**1.02977**	**-4.087**	**0.000044**	***
sw_is_templ_unfaith_to_base	**-2.04350**	**0.51577**	**-3.962**	**0.000074**	***
sw_is_seg_unfaith_to_all_other_forms	**-1.33782**	**0.45307**	**-2.953**	**0.003149**	**
sw_is_templ_unfaith_to_all_other_forms	-0.01552	0.61430	-0.025	0.979842	
ws_is_seg_unfaith_to_base	-0.17511	0.57528	-0.304	0.760831	
ws_is_seg_unfaith_to_all_other_forms	**-1.10625**	**0.49943**	**-2.215**	**0.026758**	*
ws_is_templ_unfaith_to_all_other_forms	**1.59854**	**0.41438**	**3.858**	**0.000114**	***
sw_violates_*CV´C.CV(C)	-0.24986	0.54716	-0.457	0.647915	

mystery why negative

(ws_is_templ_unfaith_to_base was 100% redundant with sw_is_templ_unfaith_to_base)
(I tried including whether reduplicant would cause a stress clash or lapse, but this was almost entirely redundant with the other factors)

(14) How to translate into a probability

To calculate probability of όσ reduplicant for a base like *–doséna*, whose paradigm also contains unaffixed *doséna* and suffixed *dosená-hin*:

- Add up intercept (all words get this) and coefficients for all applicable factors
  - intercept: 4.50746
  - sw_is_seg_unfaith_to_base: -4.20815
  - sw_is_seg_unfaith_to_all_other_forms: -1.33782
  - ws_is_templ_unfaith_to_all_other_forms: 1.59854
  - sum = 0.56003
- Take $1/(1+e^{-0.56003}) = 64\%$

(15) What do the results mean?

- Significant positive coefficients: factors that *encourage* a όσ reduplicant
- Significant negative coefficients: factors that *discourage* a όσ reduplicant (encourage σό)

A lot of these factors are correlated, so it's not surprising that they're not all significant, but...

- segmental and templatic correspondence both matter, both to the base and to the rest of the paradigm

(16) MaxEnt grammar
Translating the factors into OT-style constraints

logistic regression model	OT constraint
`Intercept`	FINALSTRESS, PENULTSTRESS
`sw_is_seg_unfaith_to_base`, `ws_is_seg_unfaith_to_base`	MATCHSEGMENTAL-BR
`sw_is_templ_unfaith_to_base`	MATCHTEMPLATE-BR
`sw_is_seg_unfaith_to_all_other_forms`, **`ws_is_seg_unfaith_to_all_other_forms`**	MATCHSEGMENTAL-OR
`sw_is_templ_unfaith_to_all_other_forms`, **`ws_is_templ_unfaith_to_all_other_forms`**	MATCHTEMPLATE-OR
`sw_violates_*CV´C.CV(C)`	*CV́C.CV(C)

(17) Fitting MaxEnt grammar to data (used Hayes & al.'s 2009 MaxEnt grammar tool.)

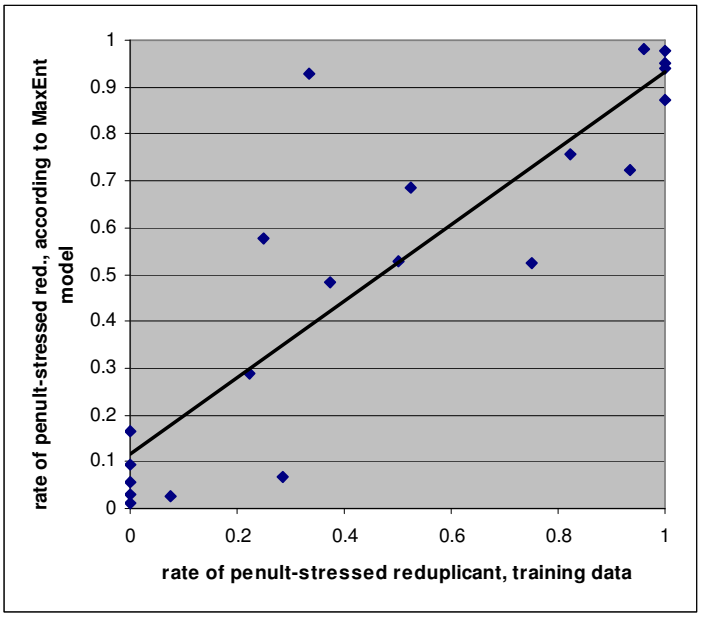
What training data looks like, plus weights learned:

		weights	**0.00**	**0.16**	**1.42**	**1.38**	**0.18**	**0.84**	**0.67**
ex. (info about rest of paradigm is encoded in constraint violations)	cands.	count	Final Stress	Penult Stress	Match Seg. -BR	Match Templ. -BR	Match Seg. -OR	Match Templ. -OR	*CVC.CV(C)
abut-abot	όσ-	10	1	0	0	0	0	0	0
	σό-	0	0	1	1	1	0	0	0
karaka-raka	όσ-	9	1	0	0	0	0	0	0
	σό-	0	0	1	1	1	0	1	0
solo-solo	όσ-	1	1	0	0	0	0	0	0
	σό-	0	0	1	1	1	1	0	0
turú-turò	όσ-	24	1	0	0	0	0	0	0
	σό-	1	0	1	1	1	1	1	0
sumirko-sirko	όσ-	1	1	0	0	0	0	0	1
	σό-	0	0	1	1	1	1	1	0
pabalik-balik	όσ-	3	1	0	0	0	1	0	0
	σό-	0	0	1	1	1	0	0	0
linggu-linggo	όσ-	1	1	0	0	0	1	0	1
	σό-	0	0	1	1	1	0	0	0
mag-aliw-aliw	όσ-	16	1	0	0	0	1	1	0
	σό-	0	0	1	1	1	0	0	0
kamulá-mulaan	όσ-	13	1	0	1	0	0	0	0
	σό-	13	0	1	0	1	0	0	0
magtinda-tindahan	όσ-	1	1	0	1	0	0	0	1
	σό-	0	0	1	0	1	0	0	0
pag-ipún-ipunin	όσ-	28	1	0	1	0	0	0	0
	σό-	2	0	1	0	1	0	1	0
pagpantig-pantigin	όσ-	1	1	0	1	0	0	0	1
	σό-	0	0	1	0	1	0	1	0

... and 33 other violation profiles

(18) Results of fitting MaxEnt grammar to data: pretty good fit

Excluding violation profiles with fewer than 3 words:



(19) What would learner extrapolate from basic data?

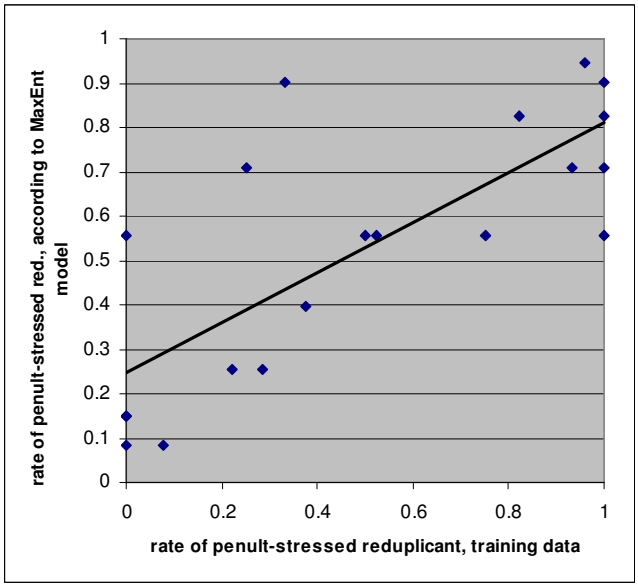
Inspired by Ryan 2010: in a situation where several candidates compete, he trains a learner on just the most-frequent candidates and observes the pattern of generalization to other candidates.

- Here, training data are just the basic (and most numerous) cases, σό-σό (rest of paradigm = σό) and όσ-όσ (rest of paradigm = όσ)
- Moderate prior (Gaussian prior with sigma^2 = 2.5)
  - in compromise between keeping weights close to zero vs. fitting the data, learner will give moderate priority to both
  - learner will prefer to give several constraints a small weight than to give one a large weight (see Martin 2007)
- Test resulting grammar on all violation profiles

constraint	Final Stress	Penult Stress	Match Seg.-BR	Match Templ.-BR	Match Seg.-OR	Match Templ.-OR	*CVC.CV(C)
weight learned	**0.00**	**0.23**	**0.66**	**0.66**	**0.66**	**0.66**	**0.00**

⇒ **responsibility is spread evenly over all 4 faithfulness constraints**

Still a pretty good fit:



D. Conclusions

(20) Summary

- It was already known that segmental and templatic stress correspondence both play systematic roles in Tagalog morphology (verb and noun suffixation).
- In two-syllable reduplication, both segmental and templatic stress correspondence contribute
  - and correspondence both to base of reduplication and to rest of the paradigm matters
- Do speakers need to learn this about reduplication?
  - not necessarily: a model that remains agnostic about which type of faithfulness is responsible for the basic cases produces a reasonably good fit to the rest of the cases.
- Like Ryan 2010, this is a case where we don't have to claim that speakers learn all the details
  - In this case, the data can be explained by saying that speakers base their grammar on the core data—if the core data are ambiguous, the grammar remains uncertain.

(21) Future work

- Dictionaries disagree about stress pattern of certain words, and there's probably quite a bit of variation within words
  - I need to elicit a large set of pronunciations from several speakers.

References

Bagemihl, Bruce. 1989. The Crossing Constraint and "backwards languages." *Natural Language and Linguistic Theory* 7. 481–549.

Chung, Sandra. 1983. Transderivational constraints in Chamorro phonology. Language 59. 35–66.

English, Leo. 1986. Tagalog-English dictionary.. Manila: Congregation of the Most Holy Redeemer; distributed by Philippine National Book Store.

Ferrer, Alicia S. 2006. Diksyunaryo Filipino-English. Millennium ed. Manila, Philippines: MECS Pub. House.

Gelman, Andrew, Yu-Sung Su, Masanao Yajima & Jennifer Hill. 2010. arm: data analysis using regression and multilevel/hierarchical models. http://CRAN.R-project.org/package=arm.

Hombert, Jean-Marie. 1973. Speaking backwards in Bakwiri. *Studies in African Linguistics* 4. 227–236.

Ito, Junko, Yoshihisa Kitagawa & Armin Mester. 1996. Prosodic faithfulness and correspondence: Evidence from a Japanese Argot. Journal of East Asian Linguistics 5. 217–294.

Levelt, Willem J. M. 1999. Models of word production. Trends in Cognitive Sciences 3(6). 223-232. doi:10.1016/S1364-6613(99)01319-4 (5 January, 2011).

Martin, Andrew. 2007. The evolving lexicon. Ph.D. dissertation, UCLA.

R Development Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/

Ryan, Kevin. 2010. Variable affix order: grammar and learning. *Language* 86(4): 758-791.

Sabbagh, Joseph. 2004. Stress shift and prosodic correspondence in Tagalog.. Paper presented at the Western Conference on Linguistics (WECOL), University of Southern California.

Schachter, Paul & Fe T Otanes. 1972. Tagalog Reference Grammar.. Berkeley, CA: University of California Press.

Shryock, Aaron. 1993. A metrical analysis of stress in Cebuano. Lingua 91(2–3). 103-148. doi:10.1016/0024-3841(93)90010-T (30 January, 2012).

Steriade, Donca. 1997. Lexical conservatism and its analysis.

Steriade, Donca. 1999. Lexical conservatism in French adjectival liaison.. In J. -Marc Authier, Barbara Bullock, & Lisa Reid (eds.), Formal Perspectives on Romance Linguistics, 243–270. Amsterdam: John Benjamins.