Variation in Tagalog tapping: word structure and frequency

Kie Zuraw, UCLA
kie@ucla.edu

0. Introduction

(1) [d] vs. [ɾ] in Tagalog

The tap [ɾ] (with perhaps an occasional trilled variant pronunciation) is spelled *r* in Tagalog, and since there is no separate trill phoneme [r], I'll use *r* to represent the tap throughout this talk.

Schachter & Otanes 1972:

- Spanish and English loans have introduced a contrast
  - ***d**isko* 'disc' ***r**isko* 'risk'
  - *kantó**d*** 'limp' *kantó**r*** 'singer'
  - *se**d**a* 'silk' *se**r**a* 'wax'
- But in the rest of the vocabulary, there is largely complementary distribution, and a corresponding alternation, with *r* intervocalically and *d* elsewhere: d→ɾ/V__V
  - *tawa**d*** 'forgiveness' *pa-tawa**r**-in* 'forgive'

What is the domain of this tapping rule?

- obligatory at stem-suffix boundary:
  *tawa**d*** 'forgiveness' *pa-tawa**r**-in* 'forgive'
- varies at prefix-stem boundary: ***d**alitâ* 'poverty' *ma-**r**alitâ* 'poor'
  ***d**ilím* 'darkness' *ma-**d**ilím* 'dark'
- within compounds and compounding reduplication?
- may occur at beginning of enclitic... *ako **d**in ~ ako **r**in* 'me too'
- (...though S&O note that *r*-initial clitic allomorphs occur, less commonly, even after consonant-final words *pangit **d**aw ~ pangit **r**aw* 'ugly they say')

(2) Overview of talk

Goal: explore the distribution of the tapping rule in these various contexts using a **written corpus**

Preview of conclusions:

- tapping applies within the p-word
- there are two possible prosodic structures for prefixed words; the choice is conditioned in part by word/root frequency ratio

Outline:

- About the corpus
- Prosodic assumptions
- Obligatory tapping at the stem-suffix boundary
- Variation at the prefix-stem boundary—effect of word/root frequency ratio
- No tapping at the stem-stem boundary

1. About the corpus

(3) Phonology in a written corpus?

Normally, we get primary phonological data face-to-face or from audio recordings or introspection (or from a linguist's transcription of one of those sources). But some phonological data can be gleaned from writing.

Abby Cohn & Lisa Lavoie (p.c.): choice of *X-er* vs. *more X* comparative formation in English as evidence for syllable count in ambiguous cases (e.g., *foul*, *fair*).

Tagalog is rich in phonological (/morphological) phenomena that are reflected (perhaps to varying degrees of reliability) in writing—all are optional in some way, so their distribution in a corpus may be of interest:

- intervocalic tapping
- vowel-height alternations: *hal**o*** 'mix' *hal**u**-in* 'to mix (OF)' (Zuraw 2002);
- nasal assimilation: *pa**m**-butas* 'borer' *pa**n**-damot* 'picker-upper' *pa**ng**-gamas* 'trowel'
- nasal substitution: ***p**ili* 'choosing' *ma-**m**ili* 'to choose (AF)' (Zuraw 2000, among many others)
- syncope, and changes in resulting consonant cluster: *gawa* 'act' *gaw_-in* 'to do (OF)', *ti**ngin*** 'look' *ti**g_n**-an* 'to look (OF)'
- partial reduplication (tricky—stem's spelling may not reflect pronunciation): *nag-**pi-frie**ndster ~ nag-**fri-frie**ndster* 'using Friendster' (Ross 1996)
- infix location: *g-**um-**raduate ~ gr-**um-**aduate* 'graduated' (Zuraw 2005)
- infix *in* vs. prefix *ni*: ***ni-**luto ~ l-**in-**uto* 'to cook (OF)'
- location of reduplication: ***pa-**pag-lagy-an ~ pag-**la-**lagy-an* 'will place' (Rackowski 1999)

(4) Web-based corpus construction

For a few years now, linguists have been pulling text from the web in order to create corpora for languages that don't already have them: de Schryver 2002; Ghani, Jones & Mladenić 2004; Kevin Scannell's *An Crúbadán* project (borel.slu.edu/crubadan).

With the help of Ivan Tam, a student research assistant, I started compiling a corpus of Tagalog in summer 2004.

- Ivan's software sends daily queries like *kami pangulo* or *+at salita oo* to the Google Web APIs service.
- The queries are not phrases, just lists of words. They are created automatically using word-frequency data from a smaller Tagalog corpus, kindly supplied by Rosie Jones.
- Google returns a list of URLs (web addresses) containing the target words, and Ivan's software retrieves the text of those pages. (Existing software such as *wget* can also be used.)
- Corpus has about 20 million words of Tagalog so far.
- The resulting body of text can be processed into a list of words with their corpus frequencies:

...

magbabala	33
magbabalak	21
magbabalance	2
magbabalangibog	2
magbabalangkas	4
mag-babalangkas	1
magbabalanse	2
magbabalaod	10
magbabalat	2
magbabalatkayo	7
magbabalaud	5
magbabalay	2
magba-balebol	1

...

(5) Drawbacks of the corpus

- It needs more processing to filter out pages from other Philippine languages (e.g., *magbabalaud*=Cebuano?)
- We don't know who the authors are—some may not be native speakers of Tagalog.
- Typographical errors may misrepresent the writer's intended pronunciation.

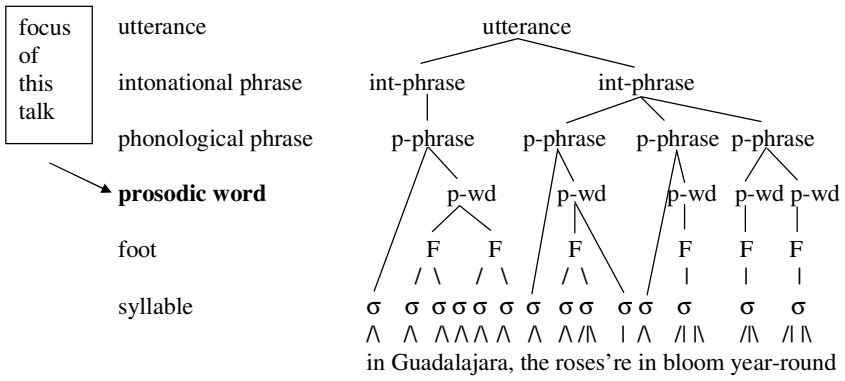
(6) Advantages of the corpus

- It exists!
- Creating it was relatively easy and not at all expensive.
- Very informal writing, as in blogs and web forums, is a great source of unusual coinages that apply Tagalog morphology to new loanwords.

2. Prosodic assumptions

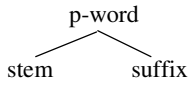
(7) Selkirk's (1978) prosodic hierarchy

[authors vary in how they would treat the function words in this example]



The main motivation for this hierarchy is that different rules and constraints within a language can be characterized as applying to different levels of the hierarchy.

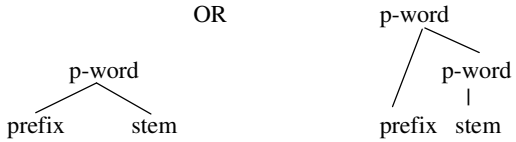
(8) Stem and suffix form a single p-word



No p-word boundary intervenes between the two morphemes.

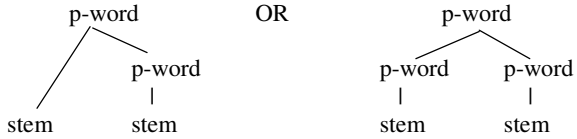
(9) Two structures possible for prefixed words

I'll propose that Tagalog employs two structures for prefixed words, with the choice determined by the degree to which the word is treated as a single unit (see below for clarification).



Depending on the structure, a p-word boundary may intervene between the two morphemes or not.

(10) Two structures possible for compounds and 2-syllable reduplication



Again, the choice depends on unithood—but in either case, a p-word boundary intervenes between the two morphemes.

(11) Why these structures?

Because they work out well in a fairly simple OT analysis (see Appendix B—the analysis there draws inspiration from Peperkamp's 1997 analysis of Italian and Spanish).

(12) Domain of tapping

I'll propose that it's the p-word: tapping applies to a *VdV* sequence if no p-word boundary intervenes.

3. Obligatory tapping at the stem-suffix boundary

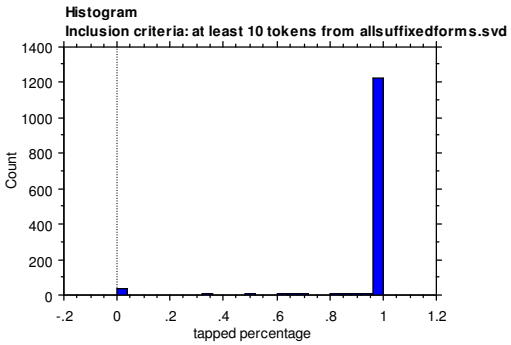
(13) Prediction

We expect tapping to apply to *stem+suffix*, because no p-word boundary intervenes between the two morphemes: (stem+suffix)$_{PWd}$.

(14) Method

All 160 (non-Spanish, non-English etymology) roots ending with *d* were extracted from English's (1986) dictionary, and the corpus was searched for any suffixed forms of each root (with or without additional prefixes and infixes).

(15) Result: quasi-obligatory tapping, as expected[1]



Explanation of histogram:

- For each item with a corpus frequency of at least 10, calculate its proportion of tapping.
- E.g., *ka-pa-tawa**r**-an* (frequency: 603) / *ka-pa-tawa**d**-an* (frequency: 1) has a tapping rate of 99.8%, or 0.998
- The histogram above divides the items into 20 bins, spread out along the horizontal axis: tapping rate 0% to 5%, 5% to 10%, etc., and counts how many items fall into each bin. The count is shown, for each bin, on the vertical axis.
- In this case, nearly all the items fall into the highest bin, with tapping rates of 95% to 100%.

4. Variable tapping at the prefix-stem boundary

(16) Prediction

Of the two prosodic structures proposed above for prefixed words, one should produce tapping—(prefix+stem)$_{PWd}$—and the other shouldn't—(prefix+(stem)$_{PWd}$)$_{PWd}$.

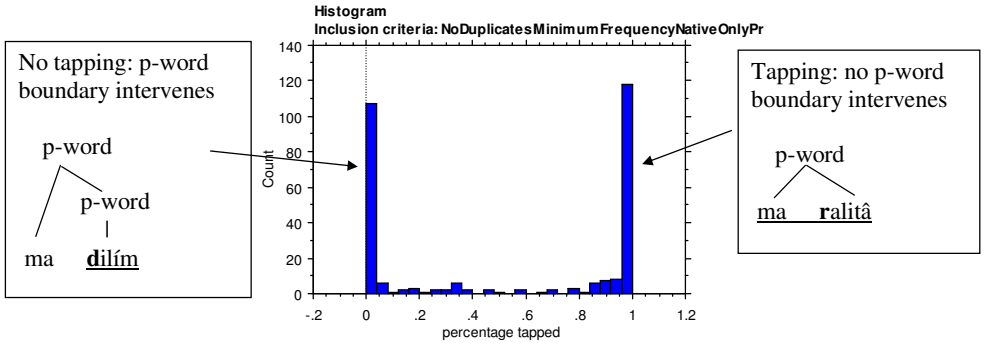
[1] All graphs and statistical results from Statview.

(17) Method

- To avoid having to hand-check every item, the corpus was searched only for the 592 orthographically distinct prefixed *d*-stem words that appear in a dictionary (English 1987). Thanks to R.A. Nikki Foster for extracting these words from the dictionary and entering them into a spreadsheet.
- No other morphology was allowed, except the "linker" *ng* (this results in lower counts over all than for the suffixed-word data).

(18) Result I: variation exists

This histogram is constructed the same way as the one above; again, only words with a total corpus frequency of at least 10 are included:



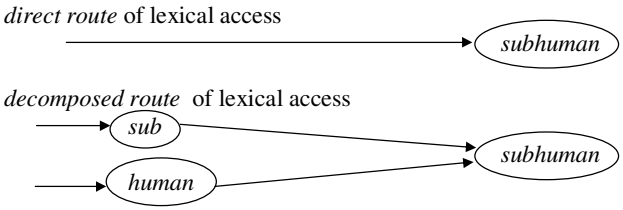
For prefixed words, we see a polarization: most fall into either the lowest-tapping-rate bin (0% to 5% tapping) or the highest (95% to 100%).

(19) What conditions the variation?

When a *prefix+stem* combination forms a simple p-word, it seems, intuitively, that the prefix is "better integrated" into the word than in the adjoined structure—that the word is treated more as a unit than as a concatenation of two units.

(20) Hay's (2003) proposal

Hay (2003) proposes that in English, an affixed word is more likely to be treated as a unit (as diagnosed by subjects' judgments of word complexity, application of phonological processes, and semantic divergence from the stem) if it is usually accessed through a *direct route*.



Hay assumes that the two routes always compete, and the speed of each route is strongly affected by the frequency of the units being accessed.

(21) Summary of Hay's prediction

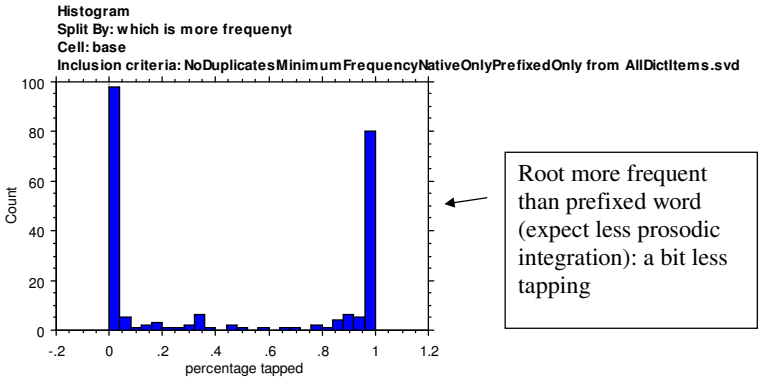
- if affixed word is more frequent than stem → direct route is favored → word gets treated more like a unit (in our case, more tapping).
- if stem is more frequent than affixed word → decomposed route is favored → word gets treated less like a unit (in our case, less tapping).

(Of course, there are probably other factors that affect which route wins...)

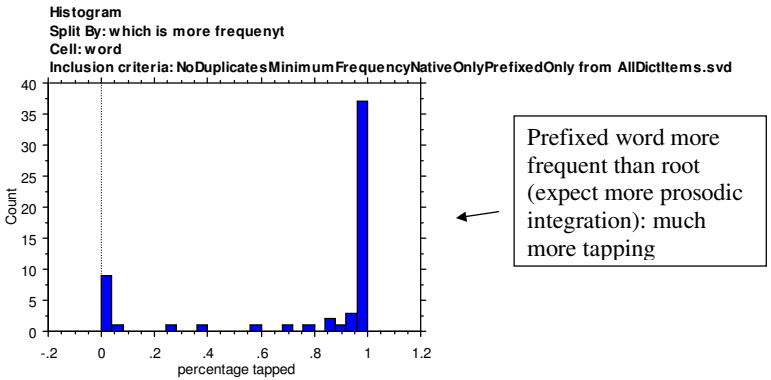
(22) Result II: frequency effect

The corpus was used to obtain the frequency of each prefixed word's root. (In the case of words with complex morphology, a desirable refinement would be to identify the word's immediate morphological predecessor.)

We can re-do the histogram from (18), splitting the words into two groups:



Root more frequent than prefixed word (expect less prosodic integration): a bit less tapping



Prefixed word more frequent than root (expect more prosodic integration): much more tapping

As predicted, when the word is more frequent than its root, tapping—which I'm attributing to closer prosodic structure—is favored.

(23) One glimpse at the OT analysis

The constraint responsible for the difference between the two structures requires each accessed unit to initiate a p-word.

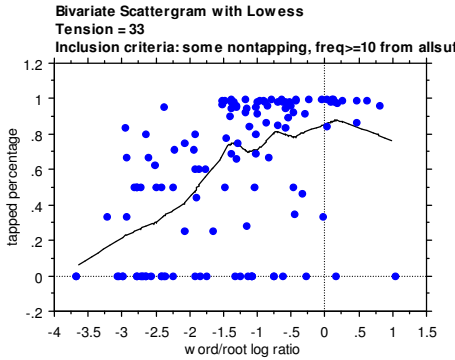
- Decomposed route (prefix and stem both accessed): (prefix (stem)$_{PWd}$)$_{PWd}$
- Direct route (only whole word accessed): (prefix stem)$_{PWd}$

(24) Back to suffixed words

There were 124 words that exhibited some degree of non-tapping. This was a small minority, but could still be worth investigating—are they just typos, or is there some systematicity to them?

Instead of dividing words into two groups—word more frequent than root or vice-versa—we can see if there is a correlation between the tapping rate and the (log) ratio of word frequency to root frequency.

The correlation turns out to be significant: Spearman's rho=.534, p<.0001.



Upward slope of trendline: the more frequent the word as compared to the root, the more tapping.

(25) Puzzling result

Non-tapping in suffixed words is very rare, but when it occurs there is a frequency effect.

Possible interpretations:

- the *stem+suffix* always form a single p-word, but paradigm uniformity effects (Steriade 1998) can, if sufficiently strong, occasionally block tapping even within a p-word
- the requirement that a suffix be integrated into the prosodic word can be overridden, occasionally, by frequency effects (whatever the mechanism is)
- the "nontapping" here is all typos, but there is a frequency effect on typos that preserve the spelling of the root intact

5. (Almost) no tapping at stem-stem boundary

(26) Prediction

There should be no tapping in either structure, because a p-word boundary intervenes between the two stems: (stem (stem)$_{PWd}$)$_{PWd}$ ((stem)$_{PWd}$ (stem)$_{PWd}$)$_{PWd}$

(Why two structures? This allows an account of variation in vowel alternations at the stem-stem boundary—not discussed here.
I assume a constraint requiring each accessed unit to end with a p-word end:

- decomposed route (each stem accessed): ((stem)$_{PWd}$ (stem)$_{PWd}$)$_{PWd}$
- direct route (only whole word accessed): (stem (stem)$_{PWd}$)$_{PWd}$)

(27) Stem+stem morphology occurs...

- in compounding: *basag-ulo* 'fight' (lit. *breaking-head*)
- and, I'll assume, in two-syllable reduplication: *pa-balik-balik* 'recurrent', from *balík* 'return'

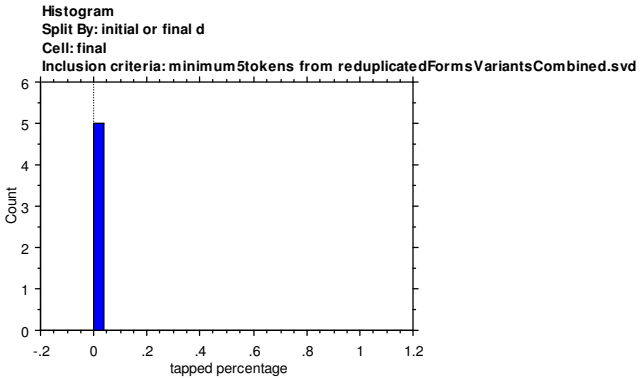
I assume that two-syllable reduplication has the same prosodic structure as compounding, because

- it helps explain the tapping data! (also nasal assimilation, vowel alternations)
- the copy is the same size as the minimal p-word in Tagalog (two syllables)

Because of practical issues in looking at compounds (when spelled with a space, impossible to identify through automatic means; when spelled with a hyphen, more doable but not yet done), I'm restricting my attention here to 2-syllable reduplication.

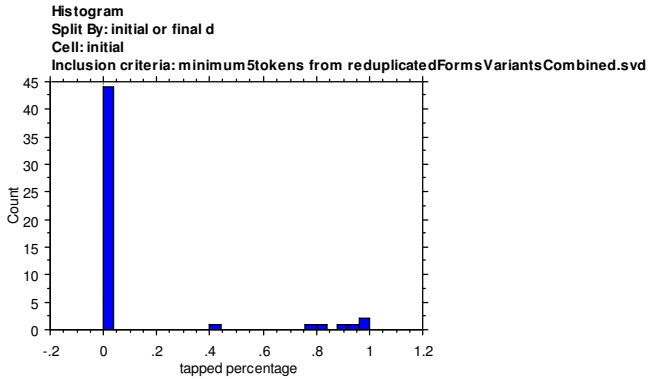
(28) Stem-final d*: few cases, but no tapping*

Tapping does not occur in words like *agá**d**-agád* 'immediately' (only words with frequency of at least 5 were used)—but this could be because the *d* is not truly intervocalic: *agá**d**-[ʔ]agád*



(29) Stem-initial d*: no tapping*

Tapping is very rare in words like *daglí-**d**aglî* 'right away':



(30) Positional faithfulness?

The difference between prefixation and suffixation might look like a positional faithfulness effect (e.g., Beckman 1996): stem-initial *d*s have some ability to resist tapping, while stem-final *d*s have no such ability.

If positional faithfulness were the only factor at work, then in stem+stem cases we would expect some variation: *daglí-**d**aglî ~ *daglí-**r**aglî*

(31) Is the lack of tapping a reduplicative identity effect?

(E.g., Wilbur 1973, McCarthy & Prince 1995)

Probably not, because tapping at the stem-stem boundary is blocked even when the other copy undergoes tapping: *ka-aga**d**-aga**r**-an, ka-**r**aga-**d**aga-n*

(32) Is the lack of tapping a frequency effect?

It's true that most of the reduplicated forms are less frequent than their roots, but in prefixed forms, this caused only a small preference for non-tapping, not the huge preference we see here.

➔ I think the prosodic effect is real.

6. Conclusion

(33) Summary

I. The proposed prosodic structures—see Appendix B for an analysis that derives them—account for the distribution of tapping, assuming that tapping's domain is the p-word:

- (stem+suffix)$_{PWd}$: obligatory tapping
- (prefix (stem)$_{PWd}$)$_{PWd}$ or (prefix+stem)$_{PWd}$: variation
- (stem (stem)$_{PWd}$)$_{PWd}$ or ((stem)$_{PWd}$ (stem)$_{PWd}$)$_{PWd}$: no tapping

II. The variation in prefixed cases is conditioned in part by how frequent a prefixed word is in comparison to its root.

III. Frequency ratio is weakly predictive even in governing the distribution of rare exceptions to a seeming requirement of the grammar (tapping at the stem-suffix boundary).

(34) Further issues (just a sample!)

- Word+clitic sequences (see Appendix A)
- Are the prosodic structures proposed here consistent with other phenomena, such as vowel alternations?
  - I'm currently thinking that lowering of high vowels to mid occurs before a p-word boundary, which is one of the reasons for having two structures for *stem+stem*.
- Other predictors of tapping in prefixed words—e.g., construction-specific differences (do some prefixes like to induce tapping more than others, independent of the frequency effect on individual items with those prefixes?)
- Do other variable phenomena—such as nasal substitution, which displays a similar frequency effect—show polarization or gradient variation?
- What causes polarization vs. gradient variation? Baroni (2001), looking at Northern Italian *s*-voicing (a similar case to Tagalog tapping) found gradient variation instead.
- Is frequency ratio a determinant of word structure (Hay proposes that frequency ratio affects a word's ability to undergo semantic drift, for example), or is frequency ratio just a proxy for underlying determinants, such as semantics.

References

Baroni, Marco (2001). The representation of prefixed forms in the Italian lexicon: Evidence from the distribution of intervocalic [s] and [z] in northern Italian. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology* 1999, Springer, Dordrecht: 121-152.

Beckman, Jill (1999) *Positional faithfulness: An Optimality Theoretic treatment of phonological asymmetries.* Garland, New York.

de Schryver, Gilles-Maurice (2002). Web for/as corpus: a perspective for the African languages. *Nordic Journal of African Studies* 11: 266-282.

English, Leo (1986). *Tagalog-English Dictionary*. Congregation of the Most Holy Redeemer, Manila. Distributed by (Philippine) National Book Store.

Ghani, Rayid, Rosie Jones & Dunja Mladenić (2004). Building minority language corpora by learning to generate Web search queries. *Knowledge and Information Systems* 7: 56-83.

Hay, Jennifer (2003). *Causes and Consequences of Word Structure*. Routledge, New York and London.

Hayes, Bruce, Bruce Tesar, and Kie Zuraw (2003). OTSoft 2.1. Software package, http://www.linguistics.ucla.edu/people/hayes/otsoft/.

McCarthy, John & Alan Prince (1995). Faithfulness and reduplicative identity. *Papers in Optimality Theory, UMass Occasional Papers in Linguistics* 18: 249-348

Nespor, Marina and Irene Vogel (1986). *Prosodic Phonology*. Foris, Dordrecht.

Peperkamp, Sharon (1997). *Prosodic Words*. Holland Academic Graphics, The Hague.

Rackowski, Andrea (1999). Morphological optionality in Tagalog aspectual reduplication. *Papers on Morphology and Syntax, Cycle Two, MIT Working Papers in Linguistics* 34: 107-136.

Ross, Kie (1996). Floating phonotactics: infixation and reduplication in Tagalog loanwords. UCLA M.A. thesis.

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

Selkirk, Elisabeth (1978). On prosodic structure and its relation to syntactic structure. In T. Fretheim (ed.), *Nordic prosody* Vol. 2. TAPIR, Trondheim: 111-140.

Steriade, Donca (2000). Paradigm Uniformity and the phonetics/phonology boundary. In Janet Pierrehumbert and Michael Broe (eds.), *Papers in Laboratory Phonology* vol. 6, Cambridge University Press, Cambridge.

Truckenbrodt, Hubert (1999). On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry* 30: 219–255

Wilbur, Ronnie Bring (1973). *The Phonology of Reduplication.* Indiana University Linguistics Club, Bloomington.

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

Zuraw, Kie (2002). Aggressive reduplication. *Phonology* 19: 395-439.

Zuraw, Kie (2005). The role of phonetic knowledge in phonological patterning: Corpus and survey evidence from Tagalog. Manuscript, UCLA.

Appendix A: Clitic-initial tapping

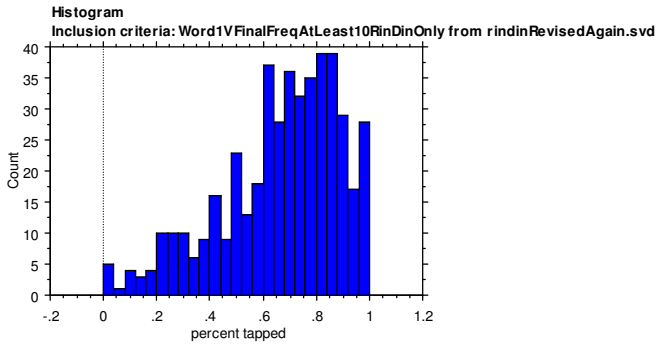
Tagalog has two *d/r*-initial enclitics, *din/rin* 'also' and *daw/raw* '(reported speech)'.

(35) Method

For now, I've looked only at *din/rin*.
All *X {d,r}in* bigrams (2-word sequences) were extracted from the corpus.

The data shown here are only for sequences where the first word (*X*) ends in a vowel (and only for sequences with frequency of at least 10).

(36) Results I: true gradience rather than item-by-item preference



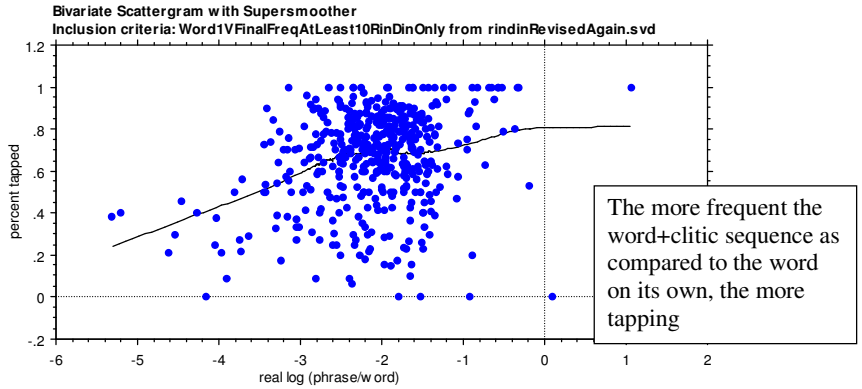
In the prefixed cases, it looked at though most words have a lexicalized form: with tapping or without.

Here, either...

- each word+clitic's behavior is decided on the fly.
- or, if there are lexical entries for (frequent) word+clitic combinations, the lexical entries are themselves variable.

(37) Results II: weak frequency effect

R^2=.072, but p<.0001



So, one of the factors that goes into this on-the-fly decision may be the frequency of the word+clitic sequence.

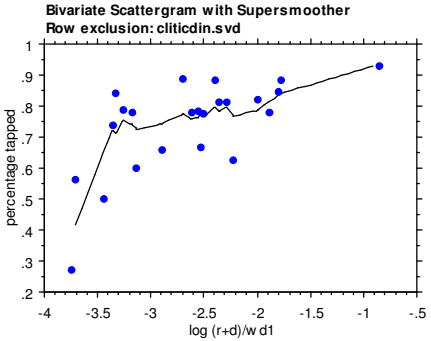
Possible interpretations

- word+clitic sequences can be stored in the lexicon
- frequency and tapping are both products of some underlying cause, such as discourse cohesion (whatever that is)

(38) Clitic+clitic sequences

(Because of ordering restrictions of clitics, not all sequences will be possible)

Here, the correlation is stronger: R^2=.424, p=.0005



Appendix B: OT analysis

I've schematized things somewhat by assuming the stem is always a disyllabic foot. Of course that's not always true. In order to allow the tableaux to fit on the page, only crucial constraints are shown in each tableau.

Thanks to OTSoft (Hayes, Tesar & Zuraw 2003) for checking my calculations.

1. Prefixed word, accessed through decomposed route

- Prefix can't form a p-word on its own (*a*,*b*) because monosyllabic. FOOTBINARITY rules out footing the prefix (*a*), and *n* DOM *n*–1 ("a prosodic unit of level *n* must dominate some prosodic unit of level *n-1*") rules out forming a p-word from an unfooted syllable (*b*) (every p-word must dominate some foot).
- Prefix and stem can't form a single p-word (*e*), because of ALIGN(AccessedUnit,L,PWd,L): the beginning of every accessed unit—in this case, the prefix, the stem, and the whole word—must coincide with the beginning of a p-word.
- The winner (*c*) adjoins the prefix to the stem's p-word to form a higher p-word.

	prefix+stem *decomposed route*	FTBIN	*n* DOM *n*-1	WRAP (LxWd,PWd)	ALIGN (AccUn,L,PWd,L)	*n* ImmDOM *n*-1
a	$[(\text{pref})_F]\omega\ [(\text{stem})_F]\omega$	*!		*!		
b	$[\text{pref}]\omega\ [(\text{stem})_F]\omega$		*!	*!		*
☞ *c*	$[\text{pref}\ [(\text{stem})_F]\omega]\omega$					*
d	$\text{pref}\ [(\text{stem})_F]\omega$			*!	**	
e	$[\text{pref}\ \ (\text{stem})_F]\omega$				*!	

2. Prefixed word, accessed through direct route

- Prefix still can't form a p-word on its own (*a*, *b*) because monosyllabic.
- Now it's OK for the prefix and stem to form a single p-word (*e*), because the prefix and stem aren't "accessed units" (only the whole word is) → no violation of ALIGN(AccessedUnit,L,PWd,L).
- The adjoined candidate (*c*) is ruled out by *n* ImmDOM *n*-1: a prosodic unit of level *n* must *immediately* dominate a unit of level *n*-1. Candidate (*c*) violates this constraint because the higher p-word immediately dominates a syllable and a p-word, but no foot.

	prefix+stem *direct route*	FTBIN	*n* DOM *n*-1	WRAP (LxWd,PWd)	ALIGN (AccUn,L,PWd,L)	*n* ImmDOM *n*-1
a	$[(\text{pref})_F]\omega\ [(\text{stem})_F]\omega$	*!		*!		
b	$[\text{pref}]\omega\ [(\text{stem})_F]\omega$		*!	*!		*
c	$[\text{pref}\ [(\text{stem})_F]\omega]\omega$					*!
d	$\text{pref}\ [(\text{stem})_F]\omega$			*!	*	
☛ *e*	$[\text{pref}\ \ (\text{stem})_F]\omega$					

3. Suffixed word, accessed through decomposed route

- Like a prefix, a suffix can't form a p-word on its own (*a*, *b*).
- Because the ALIGN constraint shown refers to left edges, in the suffixation case adjoining the suffix to the stem's p-word (*c*) does not satisfy the ALIGN constraint. Thus, (*e*), with stem and suffix forming a single p-word, is the winner even if the suffix is accessed as a unit.

	stem+suffix *decomposed route*	FTBIN	*n* DOM *n*-1	WRAP (LxWd,PWd)	ALIGN (AccUn,L,PWd,L)	*n* IMMDOM *n*-1
a	$[(\text{stem})_F]\omega\ [(\text{suffix})_F]\omega$	*!		*!		
b	$[(\text{stem})_F]\omega$ [suffix]ω		*!	*!		*
c	$[[(\text{stem})_F]\omega$ suffix]ω				*	*!
d	$[(\text{stem})_F]\omega$ suffix			*!	*	
☞ *e*	$[(\text{stem})_F$ suffix]ω[2]				*	

4. Suffixed word, accessed through direct route

- The suffix still can't form a p-word on its own (*a*, *b*).
- Now the ALIGN constraint is satisfied by both (*c*) and (*e*), because the suffix was not an accessed unit, so *n* IMMDOM *n*-1 makes the decision.
- Note that the winning structure is the same for suffixed words, whether accessed through decomposed route or direct route.

	stem+suffix *direct route*	FTBIN	*n* DOM *n*-1	WRAP (LxWd,PWd)	ALIGN (AccUn,L,PWd,L)	*n* IMMDOM *n*-1
a	$[(\text{stem})_F]\omega\ [(\text{suffix})_F]\omega$	*!		*!		
b	$[(\text{stem})_F]\omega$ [suffix]ω		*!	*!		*
c	$[[(\text{stem})_F]\omega$ suffix]ω					*!
d	$[(\text{stem})_F]\omega$ suffix			*!		
☞ *e*	$[(\text{stem})_F$ suffix]ω					

5. Compound word, accessed through decomposed route

- Each stem can be footed, so there's nothing wrong with making each a p-word.
- The constraint WRAP(LexicalWord,ProsodicWord), inspired by Peperkamp's adaptation of Truckenbrodt's (1999) WRAP(XP,PPhrase) approach to p-phrases, requires that each lexical word (in this case, each stem, plus the compound) be contained within a p-word.[3] This rules out (*a*), where the two stems' p-words are separate—though presumably part of the same p-phrase—and (*e*, *f*), where one stem is left out of the p-word.
- Candidate (*c*), with adjunction as in a prefixed word accessed through the direct rout, is ruled out by an ALIGN constraint requiring the *end* of every accessed unit to coincide with the end of a p-word; thus, each of the two stems must be followed by a p-word boundary.
- The winner, (*b*), joins two p-words into a higher p-word (in violation of low-ranked NOCOORD, which forbids recursion of prosodic levels except in an adjunction configuration).

[2] This is not the actual footing: there is presumably a foot at the right edge of the p-word in most cases. The point with this candidate is that there is a foot that the p-word dominates; it doesn't matter, for this tableau's purposes, whether the foot takes both its syllables from the stem or one from the stem and one from the syllable.

[3] Truckenbrodt applies the syntactic relation of segment-dominance to adjoined structures; I adopt a more primitive approach here.

	stem+stem *decomposed route*	WRAP (LxWd,PWd)	ALIGN (AccUn,L,PWd,L)	ALIGN (AccUn,R,PWd,R)	ALIGN (LxWd,L,PWd,L)	NOCOORD
a	$[(\text{stem})_F]\omega\ [(\text{stem})_F]\omega$	*!				
☞ *b*	$[[(\text{stem})_F]\omega\ [(\text{stem})_F]\omega]\omega$					*
c	$[(\text{stem})_F\ [(\text{stem})_F]\omega]\omega$			*!		
d	$[[(\text{stem})_F]\omega\ (\text{stem})_F]\omega$		*!		*	
e	$(\text{stem})_F\ [(\text{stem})_F]\omega$	*!	*	*	*	
f	$[(\text{stem})_F]\omega\ (\text{stem})_F$	*!	*	*	*	
g	$[(\text{stem})_F\ (\text{stem})_F]\omega$		*!	*	*	

6. Compound word, accessed through direct route

- The two alignment constraints on accessed units become irrelevant, because only the whole compound is an accessed unit; thus, (*c*) is not ruled out. (It does presumably violate a constraint against recursion of prosodic levels, ranked too low to be shown here).
- Candidates (*d*) and (*g*) are ruled out by a constraint requiring every lexical word (in this case, each stem and the whole word) to initiate a prosodic word.

	stem+stem *direct route*	WRAP (LxWd,PWd)	ALIGN (AccUn,L,PWd,L)	ALIGN (AccUn,R,PWd,R)	ALIGN (LxWd,L,PWd,L)	NOCOORD
a	$[(\text{stem})_F]\omega\ [(\text{stem})_F]\omega$	*!				
b	$[[(\text{stem})_F]\omega\ [(\text{stem})_F]\omega]\omega$					*!
☞ *c*	$[(\text{stem})_F\ [(\text{stem})_F]\omega]\omega$					
d	$[[(\text{stem})_F]\omega\ (\text{stem})_F]\omega$				*!	
e	$(\text{stem})_F\ [(\text{stem})_F]\omega$	*!	*		*	
f	$[(\text{stem})_F]\omega\ (\text{stem})_F$	*!		*	*	
g	$[(\text{stem})_F\ (\text{stem})_F]\omega$				*!	