Class 1: Free variation and lexical variation

To do for next time
- Study questions for Friday (sorry! they'll be due Tuesdays after this): Anttila 1997
- To get the reading, you'll need to log in to ccle.ucla.edu (and be enrolled in the course!)
- To get the study questions and related files, go to the main course web page, on www.linguistics.ucla.edu/people/zuraw
- To do the questions, you'll need to run OTSoft, on your Windows computer or one of the departmental ones

Overview: Variation permeates real phonological data. It will be convenient for the rest of the course if we spend this week on variation and how to handle it.

1. Idealized free variation

Suppose a language has an optional vowel harmony process:

/álkat+i/ → [álkat-i] *or* [álket-i]

In true free variation...
- The same speaker can produce both variants, for any word.
- There is no meaning difference between the variants, though they may represent different degrees of formality, etc.
- One variant may be more frequent, but the rate is the same for all target morphemes, and for all triggering morphemes (if target and trigger are in different morphemes):

[álkat-i]	70%	[álket-i]	30%
[móbak-im]	70%	[móbek-im]	30%
[sélab-ik]	70%	[séleb-ik]	30%

- Exception to the above: there might be other phonological factors that affect the rate of variation, but words with the same phonological properties will behave alike:

e.g., suppose that stress matters

[semát-i]	90%	[semét-i]	10%
[lukár-im]	90%	[lukér-im]	10%
[sikáb-ik]	90%	[sikéb-ik]	10%

Why did I use an imaginary language? Because it's really hard (impossible?) to fine a real example!

2. Idealized lexical variation

Suppose a language has two different ways to ensure that adjacent obstruents match in voicing ($*\begin{bmatrix}-\text{sonorant} \\ \alpha\ \text{voice}\end{bmatrix}\begin{bmatrix}-\text{sonorant} \\ -\alpha\ \text{voice}\end{bmatrix}$):

/sif+z/ → [sif-s]
/wof+z/ → [wov-z]

In the simplest form of lexical variation...
- Each word has just one behavior—the variation is across items, not within items

3. Free variation as stylistic variation

Classic work in sociolinguistics focused on how "variable rules" are affected by social factors. Classic graph, Labov 1972, showing how New York City English speakers pronounced /θ/:

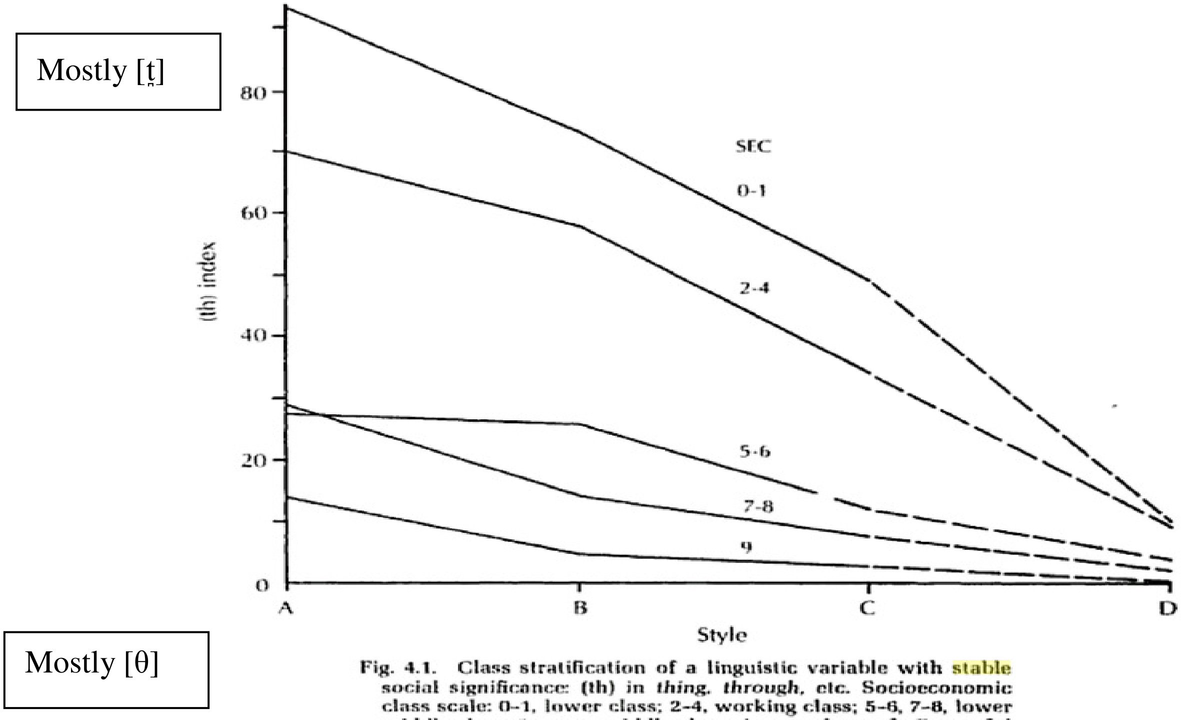


Fig. 4.1. Class stratification of a linguistic variable with stable social significance: (th) in *thing, through*, etc. Socioeconomic class scale: 0-1, lower class; 2-4, working class; 5-6, 7-8, lower middle class; 9, upper middle class. A, casual speech; B, careful speech; C, reading style; D, word lists.

p. 113

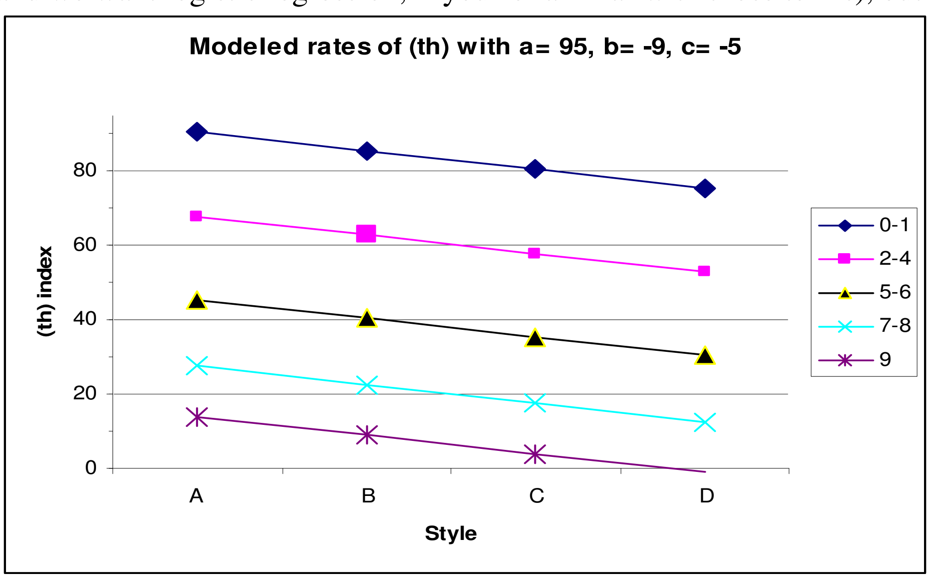
Labov's approach: model the frequency with which a rule applies

/θ/ → [–continuant]

rate of applying rule = *a* + *b**Class + *c**Style

- Different people have different baseline rates of applying rule (*a*+*b**Class)
- But they vary the same way in response to "style" (*c**Style, where A=0, B=1, etc.)

That's not quite the right model, because it predicts parallel lines (this is plain linear regression, and we want logistic regression, if you're familiar with those terms), but you get the idea:



4. Modeling idealized free variation: variable constraint ranking

Most of the work on modeling variation assumes idealized free variation, so we'll start with that, and then build in more realism gradually.

Jagged line (not standard notation): ranking of these two constraints *varies*
On some occasions, *θ >> IDENT(cont)
On other occasions, IDENT(cont) >> *θ

/θɪk/	*θ	IDENT(cont)
☞ *a* [θɪk]	*	
☞ *b* [t̪ɪk]		*

o How is the jagged line different from the dashed/dotted line you often see in tableaux?

5. Modeling idealized free variation II: probabilistic constraint ranking

The jagged line means both rankings are possible.
It doesn't say whether one ranking is more frequent than the other.

1st theory to quantify ranking preferences was **Stochastic OT** (Boersma 1997; Boersma & Hayes 2001) ("stochastic" just means "probabilistic", so various theories could be described as "stochastic OT". With a capital *S*, though, I mean specifically Boersma's theory)

In the grammar, each constraint has a "ranking value":
*θ 101
IDENT(cont) 99

Every time a person speaks, they add a little noise to each of these numbers, then rank the constraints according to these numbers.

⇒ Go to demo (I've prepared a demo in an Excel file to clarify)

o Researchers who use this model often acknowledge stylistic conditioning, but idealize away from it. How could we modify the model to add in the effect of style?

6. How to learn ranking values? Gradual Learning Algorithm

(Boersma 1997; Boersma & Hayes 2001)

1. Suppose you're a child. You start out with both constraints' ranking values at 100.
2. You hear an adult say something—suppose /θɪk/ → [t̪ɪk]
3. You use your current ranking values to produce an output. Suppose it's /θɪk/ → [θɪk].
4. Your grammar produced the wrong result! (If the result was right, repeat from Step 2)
5. Constraints that [t̪ɪk] violates are ranked too low; constraints that [θɪk] violates are too high.
6. So, promote and demote them, by some fixed amount (say 0.33 points)

/θɪk/	*θ	IDENT(cont)
the adult said this [θɪk]	* demote to 99.67	
your grammar produced this [t̪ɪk]		* promote to 100.33

7. Repeat.

⇒ Go to demo (same Excel file, different worksheet)

- Suppose, as in our demo, that adults produce [t̪ɪk] 90% of the time. Will your grammar ever stop making errors?
- What's the effect of the column labeled 'plasticity'?
- What if the adults actually don't vary, and the outcome is always [θɪk]. What will happen to the ranking values? (After discussing, let's try it in the spreadsheet.)
- In that case, will your grammar ever stop making errors?

7. Using the Gradual Learning Algorithm

Fortunately, you don't need to make an Excel file like this. Bruce Hayes's OTSoft (Hayes, Tesar, & Zuraw 2003) will do the work for you!

⇒ Go to OTSoft demo

8. Rival theory: Maximum Entropy OT

Logically, we should discuss this theory now. But I don't want to throw more than one quantitative model at you today, so we'll save it for next time.

9. Lexical variation

Recall: in lexical variation, each word has its own behavior.

Tagalog: Austronesian language from the Philippines with ~17 million native speakers (Ethnologue 2005, data from Zuraw 2009's corpus ; see also Schachter & Otanes 1972)

d → ɾ / V__V :

dunoŋ	'knowledge'	ma-**ɾ**unoŋ	'intelligent'
dinig	'heard'	ma-**ɾ**inig	'to hear'
dupok		ma-**ɾ**upok	'fragile'

But, there are also words like this

daʔig	'beaten'	ma-**d**aʔig	'beaten'
dulas	'slipperiness'?	ma-**d**ulas	'slippery'
daʔan	'road'	ma-**d**aʔan-an	'passable'

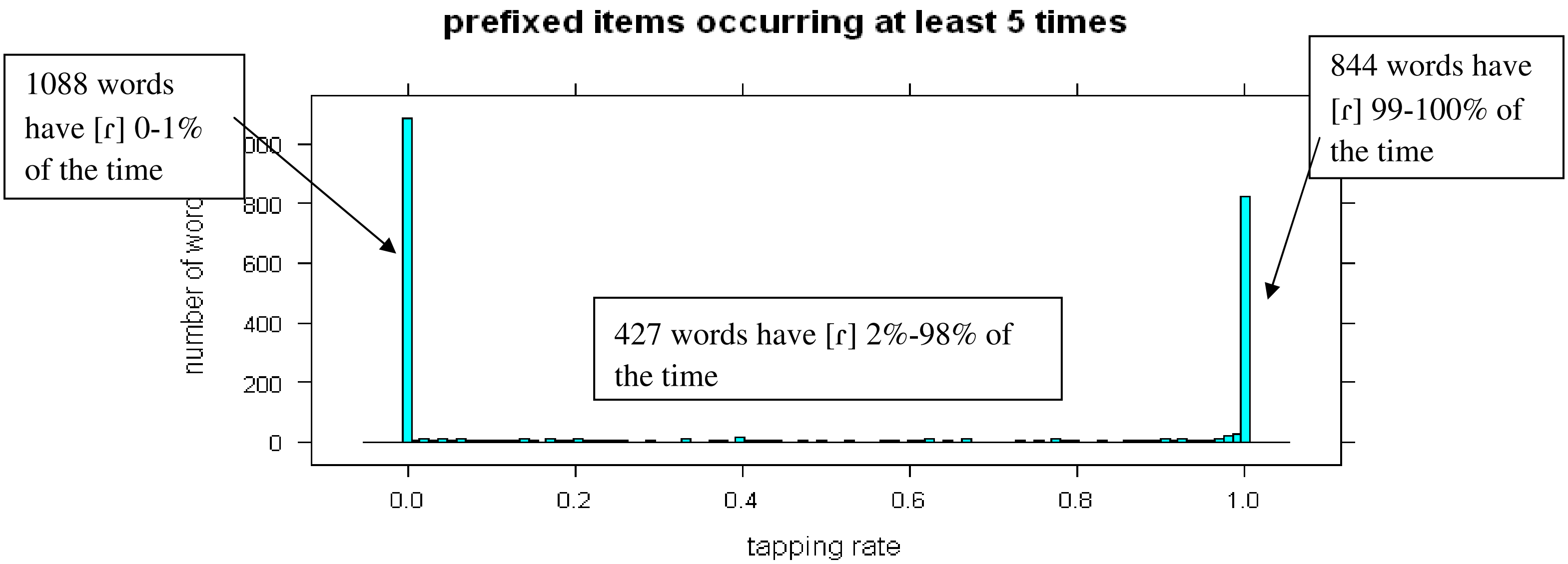
and like this

duŋis	'dirt on face'	ma-**ɾ**uŋis ~ ma-**d**uŋis	'dirty (face)'
dumi	'dirt'	ma-**ɾ**umi ~ ma-**d**umi	'dirty'

How often does each have each variant:

word	*# with d*	*# with ɾ*	*% ɾ*
ma-_unoŋ	33	9130	99.6%
ma-_inig	97	3517	97.3%
ma-_upok	0	235	100.0%
ma-_ulas	348	23	6.2%
ma-_aʔan-an	132	6	4.3%
ma-_aʔig	102	0	0.0%
ma-_umi	319	708	64.4%
ma-_uŋis	59	52	46.8%

Then, we can count up how many words are 0-<5%, how many 5-<10%, 10%-<15%, etc., and make a **histogram**.



==> Most words have a fixed behavior, though some do vary

- Let's sketch out a grammar with variable constraint ranking. What problems do we run into in modeling these data?
- Let's discuss the pros and cons of simply listing all the prefixed words in the lexicon, with /d/ or /ɾ/ in their lexical entries.

10. Modeling lexical variation: indexed constraints

Probably the best-developed theory of lexical variation is **constraint indexing** (Pater 2009, Becker 2009, Mahanta 2009[1])

The basic idea, as applied to our Tagalog case:

*VdV$_{\text{type A words}}$ >> IDENT(cont) >> *VdV$_{\text{type B words}}$

- Draw tableaux for /ma/+/dunoŋ/$_{A}$ and /ma/+/daʔig/$_{B}$

Richer example, from Becker 2009: <u>Turkish</u> (Altaic language from Turkey with 50 million speakers, Ethnologue 2005)

Three kinds of word-final obstruents in Turkish (p. 19)

always voiceless

atʃ	'hunger'	atʃ-ɨ	'hunger-*possessive*'
anatʃ	'female cub'	anatʃ-ɨ	'female cub-*possessive*'

always voiced (rarer—examples from Kaisse 1990*)*

ofsajd	'offside'	ofsajd-ɨ	'offside-*possessive*'
serhad	'Serhad (name)'	serhad-ɨ	'Serhad's'

alternating

tatʃ	'crown'	tadʒ-ɨ	'crown-*possessive*'
amatʃ	'target'	amadʒ-ɨ	'target-*possessive*'

[1] There are earlier references from the same authors, but I chose works that seemed to represent the most current versions of the authors' approaches.

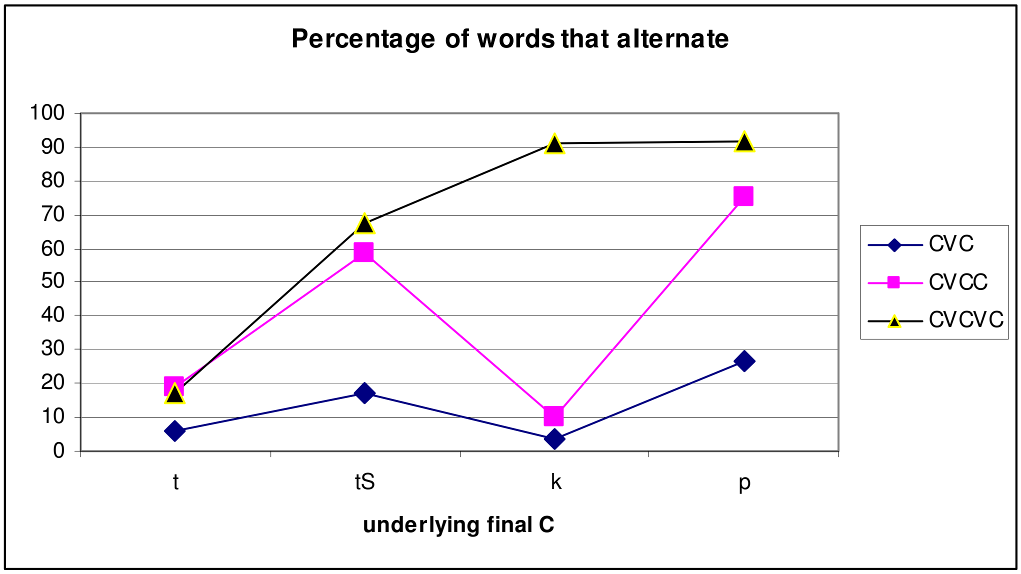
Becker takes the unaffixed form as underlying. The grammar then needs let some words undergo intervocalic voicing.

o Develop an indexed-constraint analysis of the Turkish data so far.

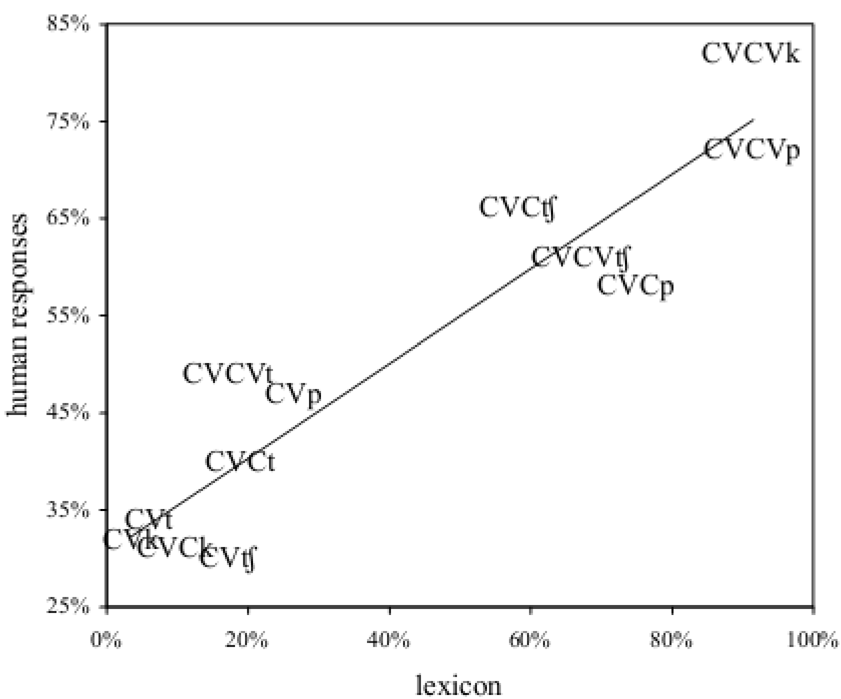
11. Patterned exceptions

The distribution of always-voiceless vs. alternating isn't random, though.

Based on Becker's p. 25:



And however this came about, Turkish speakers don't ignore it. In a wug test (Berko 1958), speakers followed the pattern closely, though overall closer to 50-50:



(Becker p. 37)

==> We want to get this information into the grammar

12. Grammar for Turkish

Becker proposes that Turkish learners have access to these constraints:

- IDENT(voice)
- IDENT(voice)$_{\sigma 1}$: the [voice] value of an output segment in a word's first syllable must be the same as the [voice] value of its input correspondent
- *VtV, *VtʃV, *VpV, *VkV
- *RtV, *RtʃV, *RpV, *RkV : don't have [p] (etc.) preceded by a sonorant C and followed by a vowel

The learner encounters an inconsistency...

/anatʃ-ɨ/ → [anatʃ-ɨ] means IDENT(voice) >> *VtʃV
/amatʃ-ɨ/ → [amadʒ-ɨ] means *VtʃV >> IDENT(voice)

The learner **clones** the IDENT constraint and re-does the ranking (see Becker's ch. 4 for how to choose which constraint to clone).

- IDENT(voice)$_{<*\text{VtʃV,anatʃ}>}$: Don't change voicing values in the lexical item /anatʃ/; conflicting constraint is *VtʃV
- IDENT(voice)$_{<*\text{VtʃV,amatʃ}>}$

○ Make tableaux for /anatʃ-ɨ/, /amatʃ-ɨ/, with the two IDENT clones and *VtʃV

Eventually, the learner ends up with constraints like...

IDENT(voice)$_{\{<*\text{VtʃV,anatʃ}>,\ <*\text{VtV,sepet}>,...\}}$ (49 <*VtʃV, *X*> items)
IDENT(voice)$_{\{<*\text{VtʃV,amatʃ}>,\ <*\text{VtV,kanatt}>,\ ...\}}$ (101 <*VtʃV, *X*> items)

When it's time to take the wug test, experimental participant must choose which IDENT constraint to assign the new word /hevetʃ/, with the conflicting constraint being *VtʃV.

The more <*VtʃV,X> items belong to the constraint, the more likely the new word is to be assigned to it.

13. A different model (see Zuraw 2010 for details and learnability)

Suppose that Turkish speakers just have lexical entries all the affixed words they know:

/anatʃ-ɨ/, /amadʒ-ɨ/

Known words surface faithfully:

	/anatʃ-ɨ/, cf. /anatʃ/	IDENT-IO(voice) R.V.: 110	*VpV 98	IDENT-OO(voice) 97.5	*VtʃV 97
☞ *a*	[anatʃ-ɨ]				*
b	[anadʒ-ɨ]	*		*	

But new words are subject to lower-ranking constraints:

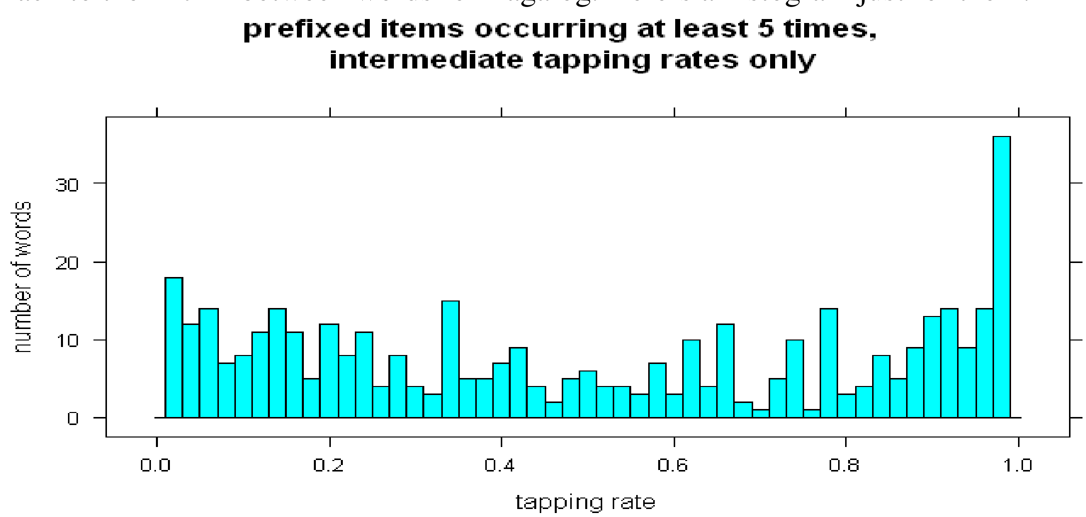
	possessive of /hevetʃ/	IDENT-IO(voice) R.V.: 110	*VpV 98	IDENT-OO(voice) 97.5	*VtʃV 97
62% *c*	[hevetʃ-ɨ]				*
38% *d*	[hevedʒ-ɨ]			*	

	possessive of /hevep/	IDENT-IO(voice) R.V.: 110	*VpV 98	IDENT-OO(voice) 97.5	*VtʃV 97
27% *e*	[hevep-ɨ]		*		
73% *f*	[heveb-ɨ]			*	

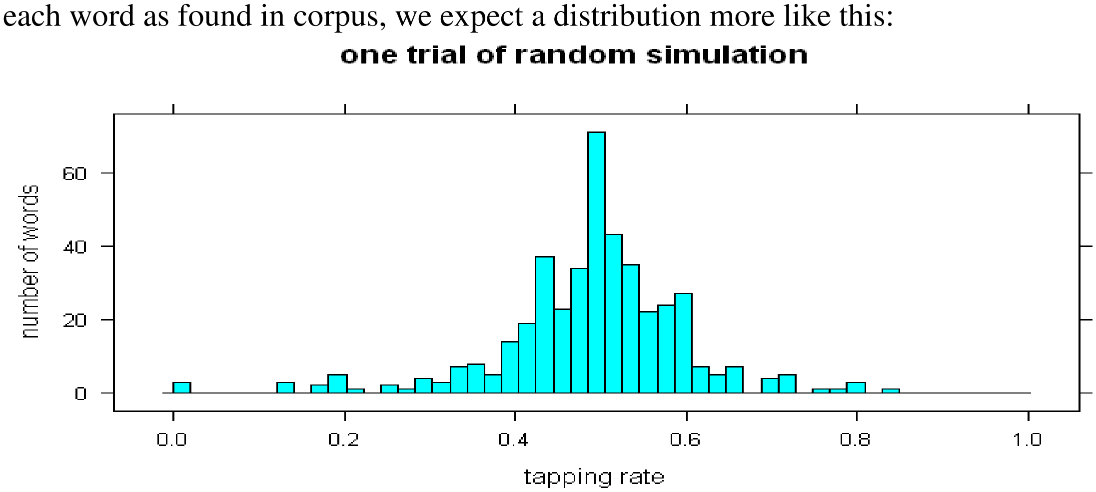
- In this model, how can we rule out pairs like hypothetical /sat/ ‘frisbee’ /fim-ɨ/ ‘frisbee-*poss*’?
- In this model, how can we ensure that the various suffixed forms of the same stem all have the same voicing behavior?

14. Mixed variation

Back to the 427 in-between words for Tagalog. Here’s a histogram just for them:



If all the words underlying had 50% tapping rate, and we sampled the same number of tokens of each word as found in corpus, we expect a distribution more like this:



Instead of free variation, it looks like different words have their own tapping rates (or, there are additional factors contributing to this variation that we haven't identified).

Assuming that's true (and true within individuals)...

o Let's brainstorm ways to combine our theories of lexical and free variation.

15. Wrapping up today

- Idealized free variation can be modeled as variable constraint ranking
  - Ranking preferences can be quantified with Stochastic OT
  - Stochastic OT comes with a learning algorithm, the Gradual Learning Algorithm
  - This algorithm is implemented in the free software OTSoft
- Lexical variation can be modeled by making constraints sensitive to lexical entries (no learning software, unfortunately)
- Lexical variation is often patterned—we want the grammar to represent this

Next time: MaxEnt OT for free variation; lexical selection

Today's bottom line[2]
- There are good tools for dealing with free variation. Unfortunately, things are often more complicated

References

Becker, Michael. 2009. Phonological trends in the lexicon: the role of constraints.. University of Massachusetts Amherst ph.d. dissertation.

Berko, Jean. 1958. The child's learning of English morphology. *Word* 14. 150-177.

Boersma, Paul. 1997. How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* 21. 43–58.

Boersma, Paul & Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32. 45–86.

Ethnologue. 2005. Ethnologue: Languages of the World, Fifteenth edition. (Ed.) Raymond G. Gordon.

Hayes, Bruce, Bruce Tesar & Kie Zuraw. 2003. OTSoft 2.1. http://www.linguistics.ucla.edu/people/hayes/otsoft/.

Kaisse, Ellen M. 1990. Toward a typology of postlexical rules.. In Sharon Inkelas & Draga Zec (eds.), *The Phonology-Syntax Connection*, 127-144. Center for the Study of Language and Information.

Labov, William. 1972. The reflection of social processes in linguistic structures. *Sociolinguistic Patterns*, 110-121. Philadelphia: University of Pennsylvania Press.

Mahanta, Shakuntala. 2009. Morpheme-specific exceptional processes and emergent unmarkedness in vowel harmony.. In Rajendra Singh (ed.), *Annual review of South Asian languages and linguistics: 2009*. Walter de Gruyter.

Pater, Joe. 2009. Morpheme-specific phonology: constraint indexation and inconsistency resolution.. In Steve Parker (ed.), *Phonological argumentation: essays on evidence and motivation*. (Advances in Optimality Theory). Equinox.

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

Zuraw, Kie. 2009. Frequency influences on rule application within and across words. *Proceedings of CLS (Chicago Linguistic Society) 43*.

Zuraw, Kie. 2010. A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language and Linguistic Theory* 28(2). 417-472.

[2] "bottom line": idiom. Originally, the last line of a financial document, which shows the final amount of money that must be paid, or was earned, etc. Metaphorically, refers to profitability or success rate of some venture ("we need to improve our bottom line"). *More* metaphorically, refers to the most important information resulting from some discussion or explanation ("And so, the bottom line is ...").