Class 19: Phonology-lexicon and phonology-processing interfaces

To do
- Mini-conference on Tuesday (10-1, but we should be done around 12:40) in conference room. Prepare handout for 15-minute presentation, 5 minutes of questions.
- Papers due Friday (PDF by e-mail is fine).

Overview: We'll look at a bunch of phonological phenomena that show frequency effects—including a case from my own research—and consider where in our model of language those effects could reside.

1 Classic frequency effect: English irregular past tense

There are only about 200 of them, but they are disproportionately likely to be frequent (e.g., Bybee & Slobin 1982).
Top 25 most frequent verbs (Oxford English Corpus)—irregulars are in bold:

1. be	**8. know**	**15. give**	**22. feel**
2. have	**9. take**	16. use	23. try
3. do	**10. see**	**17. find**	**24. leave**
4. say	**11. come**	**18. tell**	25. call
5. get	**12. think**	19. ask	
6. make	13. look	20. work	
7. go	14. want	21. seem	

Locus of explanation?

Diachrony
- In order to learn an irregular past tense form, you have to be exposed to it enough times
  → low-frequency verbs will tend to regularize from one generation to the next (*bode* > *bided*).
- Kirby 2001: simulation study

Processing
Dual-route model (see Pinker 2000 for overview and application to this case)
- When you want to say a past tense, there's a race between retrieving a stored form (which might be irregular) and creating the form via the *–ed* rule.
- The more frequent the stored form, the higher its resting activation → more likely to win the race.
  → low-frequency verbs may get pronounced as regular, even if speaker knows irregular form.

Grammar?
I don't think anyone has proposed it for *this* case, but it's a logical possibility:
- Some constraints are sensitive to frequency.

/bowd/, cf. [bajd]	I-O FAITH(hi freq)	O-O FAITH	I-O FAITH(lo freq)
bowd		*!	
☞ bajdɨd			*

or split O-OFAITH by frequency.

- Or there's just one I-O FAITH constraint, but its ranking is a function of frequency

With these three possibilities in mind, let's look at some more phonological cases and how they've been analyzed.

2 Ng 2010: Singapore English prosodic boundaries

- Singapore English has strong glottalization at prefix-stem, stem-stem, but not stem-suffix boundary
  - *mis-understand* [misʔɑndəstæn] , *stop-over* [stɔpʔovə], *magic-al* [mædʒikØəɯ](p. 8)
  - Ng analyzes this in terms of p-word structure: let's sketch it out

- Stress is realized as tone: ($L_0M^*M_0$)H or H (p. 11)
  - last syllable is H: *see* ˈH
  - first (non-final) stressed syllable gets M tone: *apple* ˈMH
  - sylls between first stress and final get M except the last: *elephant* ˈMMH, *Indonesia* ˈMMˌMH
  - syllables preceding first stress get L: *hibiscus* LˈMH, *machine* LˈH, *America* LˈMMH

- Domain of tone assignment ≈ p-word
  - tone pattern generally re-starts in compounds: *century egg* (ˈMH)(ˌH) (p. 13)
  - tone pattern may or may not restart at prefix-stem boundary: *un-install* (ˌH)-(LˈH) ~ (L-(LˈH)) (p. 12)
  - tone doesn't restart at stem-suffix boundary: remove-able (LˈMMH) (p. 12)

- Much interesting analysis follows, but let's focus on initialisms (e.g. *NUS* 'National University of Singapore')
  - Initialisms show varying degrees of prosodic merger:

Society for the Prevention of Cruelty to Animals, Anglo-Chinese Junior College, National Registration Identity Card, National Trade Unions Congress (supermarket)

		Least merger			*Most merger*
a.	*SPCA*	((((ˈH)ˌH)ˌH)ˌH)			
b.	*ACJC*	((((ˈH)ˌH)ˌH)ˌH)	(((ˈH)ˌH)ˌMH)		
c.	*NRIC*	((((ˈH)ˌH)ˌH)ˌH)	(((ˈH)ˌH)ˌMH)	((ˈH)ˌMMH)	
d.	*NTUC*	((((ˈH)ˌH)ˌH)ˌH)	(((ˈH)ˌH)ˌMH)	((ˈH)ˌMMH)	(ˈMMMH)

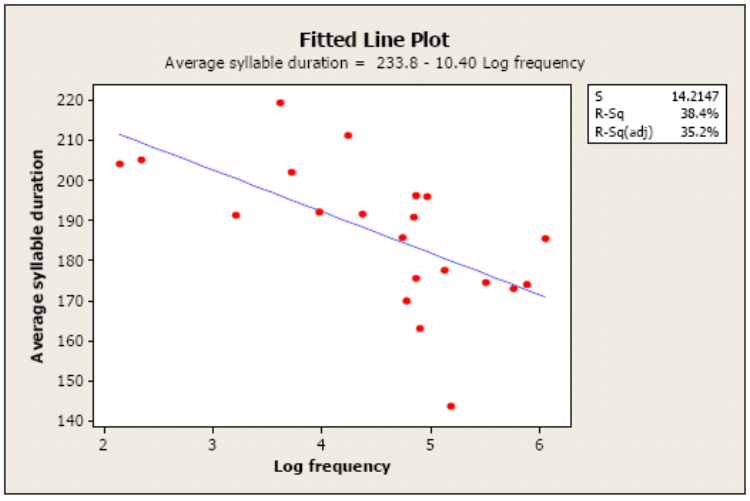
(p. 23)

- Ng finds a correlation between which group an initialism belongs to and its number of Google hits.

Why?

- Ng notes that frequency determines speed of production, perhaps because of faster access:

Figure 4: Frequency and duration



(p. 31)

- Constraints are then sensitive to speed, e.g. "Grammatical word accessed at speed *n* allows only *n* levels of stress"
  - Result is a prosodification of higher-frequency words that results in fewer stresses.
  - This is an interesting way of removing the need for the grammar to refer to frequency
    - Predicts that if we can manipulate speaking rate independent of word frequency, we'll get similar effects.

(58) *Do not restore faithful stress to destressed initialisms*

[[M][O][E]]$_{S_2}$	WRAP	*SCLASH$_S$	STRESS
☞ a. (('M),OE)			*
b. (('M),O,E)		W*	L
c. (('M)(,O)(,E))		W*	L
d. ('M)('O)('E)	W*		L

accessed at "speed 2" (S_2), so allows only two levels of stress (b and c have tertiary stresses)

(p. 33)

3 Hammond 1999: English rhythm rule

thìrteen mén or *thirtèen mén*?

- In survey, shift is more likely if adjective is more frequent: *nàive fríend* vs. *obèse chíld*
- Hammond proposes morpheme-specific faithfulness constraints, whose ranking depends on the word's frequency.

4 Löfstedt 2010: frequency-specific constraints

- We saw these earlier: Famous paradigm gaps in Swedish result when vowel shortening produces too much of a quality change.

STEM	NEUTER	GLOSS
bl[oː$_T$]	bl[ɔ$_L$] + t:	'blue'
v[iː$_T$]t	v[ɪ$_L$] + t:	'white'
v[iː$_T$]d	v[ɪ$_L$] + t:	'wide'

quality change (from Tense to Lax) is not too big

(p. 152)

STEM	NEUTER	GLOSS	ALLÉN (PL)	GOOGLE (-A)
gr[ɑː]d	INEFFABLE[51]	'straight'	0	7,140
l[ɑː]t	INEFFABLE	'lazy'	0	581,000

quality change (would be from [ɑː] to [a]) is too big

(p. 154)

- But! Sufficiently frequent words don't have a gap

gl[ɑː]d	gl[a] + t:	'happy'	29	2,110,000

(p. 154)

- For each of the vowels that can show a gap, there seems to be a frequency cut-off above which there's no gap. (Löfstedt shows this for some phenomena in other languages too) E.g.,

STEM	NEUTER	GLOSS	ALLÉN (PL)	GOOGLE (-A)
gr[ɑ:]d	INEFFABLE[51]	'straight'	0	7,140
l[ɑ:]t	INEFFABLE	'lazy'	0	581,000
gl[ɑ:]d	gl[a]+t:	'happy'	29	2,110,000

frequency counts from different corpora

(p. 154)

- Löfstedt's solution: faithfulness constraints penalizing vowel changes are indexed to frequency:

	/glad + t:/ 'happy' neut. Cf. [glɑ:d] (*Freq* /glad/ = 2,110,000)	$\sigma_{\mu\mu}$ ↔ [+stress]	[+LONG] ↔ [+TENSE]	IDENT [Long C] /V_	*MAP (ɑ,a) (7140)	*MAP (ɑ,a) (581,000)	M-PARSE	*MAP (ɑ,a) (2,110,000)
a.	glɑ:t:	*!						
b.	glat:		*!					
c.	glɑ:t			*!				
d. >	glat:							*
e.	⊙						*!	

(p. 167)

5 Boersma 1999: lexical-access *constraints*

- The problem: in Dutch, you want to be able to recognize [rɑt] as either /rɑt/ or /rɑd/.
- If you try to use a standard grammar to map perceived form to underlying form, you'll always pick the faithful one:

(7) *Failure to recognize the wheel*

[rɑt]	*VOICEDCODA	MAXVOI
☞ |rɑt| 'rat'		
|rɑd| 'wheel'		*

This is a comprehension tableau:
input = perceived phonetic form
output = lexical entry

(p. 4)

- So, Boersma proposes a family of constraints *LEX(*x*) "don't recognize any utterance as lexical item *x*" (one for each lexical item).
- Ranking depends on word's frequency:

(10) *A strong tendency to recognize the rat*

[rɑt]	*LEX (|rɑd| 'wheel')	*VOICED CODA	MAXVOI	*LEX (|rɑt| 'rat')
☞ |rɑt| 'rat'				*
|rɑd| 'wheel'	*!		*	

(p. 5)

- Actually, it's a bit more complex: *LEX(*x*/context=*y*) to allow for semantic context to matter

6 Zuraw 2009: Tagalog tapping

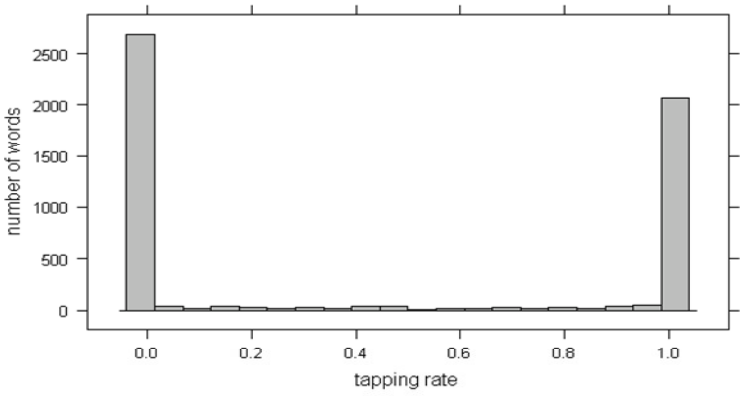
This is work that Kevin Ryan and I got started on—he did all the phonetic work.

Tapping in prefixed Tagalog words: variable

d →ɾ (spelled *r*) / V__V *dumi* 'dirt' *ma-**r**umi* 'dirty'

but *dahon* 'leaf' *ma-**d**ahon* 'leafy'

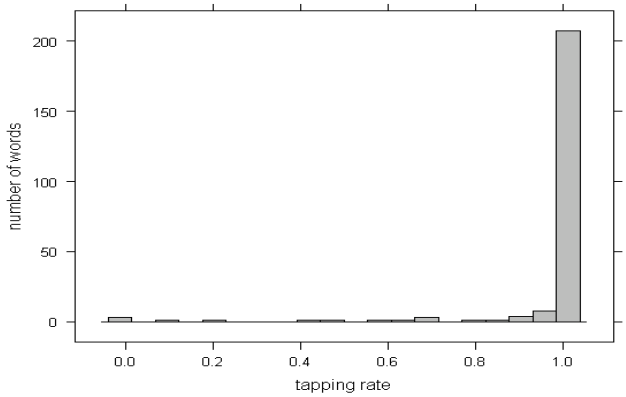
Each word seems to have a consistent behavior (using spelling data in corpus):



Not shown in this graph: The more frequent the word, the more likely tapping is.

Tapping in suffixed words: obligatory

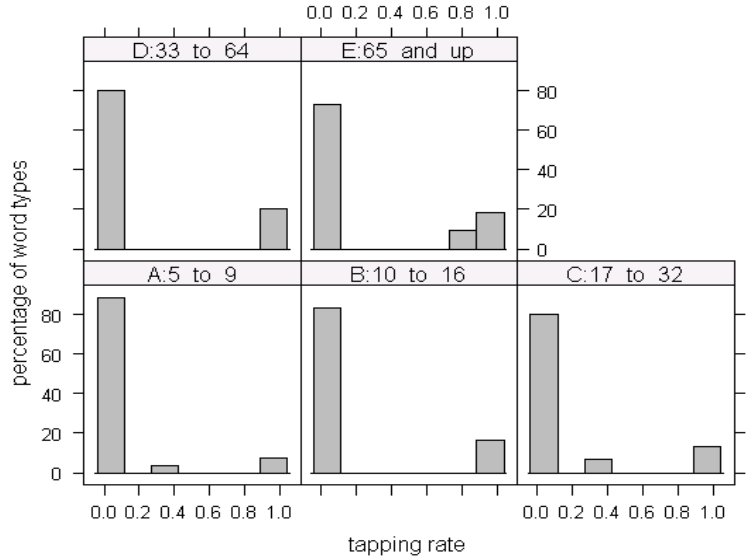
lakad 'walk' *laka**r**-an* 'to be walked on'



The grammar probably has to enforce the change here, since even low-frequency words undergo.

Tapping in p-word reduplication: nearly forbidden

dala 'carry' *dala-**d**ala* 'load carried'



Even high-frequency words (D and E) rarely show tapping. (only 84 word types, though)

→ Maybe grammar should prevent the change from applying in this context.

2009 analysis: grammar refers to outcome of lexical access

ALIGN(AccU,L; PWd,L): L edge of any accessed lexical unit must coincide with L edge of some p-word.

→ outcome for prefixed word depends on access mode:[1]

(1)

accessed: ma, Dahon, (*and maybe* maDahon)		*(...VdV...)	MINIMALITY	STEMISHEAD	ALIGN (AccU,L;PWd,L)	NORECURSION	*r
☞ *a*	(ma(dahon)$_\omega$)$_\omega$					*	
b	(ma(rahon)$_\omega$)$_\omega$					*	*!
c	(madahon)$_\omega$	*!			*		*
d	(marahon)$_\omega$				*!		
e	(ma)$_\omega$(dahon)$_\omega$		*!				
f	((ma)$_\omega$ dahon)$_\omega$		*(!)	*(!)	*	*	

less-frequent word: synthesis route (prefix, stem) should tend to win.

(2)

accessed: maDami		*(...VdV...)	MINIMALITY	STEMISHEAD	ALIGN (AccU,L;PWd,L)	NORECURSION	*r
g	(ma(dami)$_\omega$)$_\omega$					*!	
h	(ma(rami)$_\omega$)$_\omega$					*!	*
i	(madami)$_\omega$	*!					
☞ *j*	(marami)$_\omega$						*
k	(ma)$_\omega$(dami)$_\omega$		*!				
l	((ma)$_\omega$ dami)$_\omega$		*(!)	*(!)		*	

more-frequent word: whole-word retrieval route should tend to win.

- Outcome for suffixed words is fixed, because constraint that refers to access mode is low-ranked:

(3)

accessed: lakaD, an, (*and maybe* lakaDan)		*(...VdV...)	MINIMALITY	STEMISHEAD	ALIGN (AccU,L,PWd,L)	NORECURSION	*r
a	(lakad(an)$_\omega$)$_\omega$		*(!)	*(!)		*	
b	(lakadan)$_\omega$	* !			*		
☞ *c*	(lakaran)$_\omega$				*		*
d	(lakad)$_\omega$(an)$_\omega$		*!				
e	((lakad)$_\omega$ an)$_\omega$				*	* !	

[1] Access route should depend on more than just word frequency. See Hay 2003.

- Similarly, outcome for 2-syll reduplicated words is fixed:

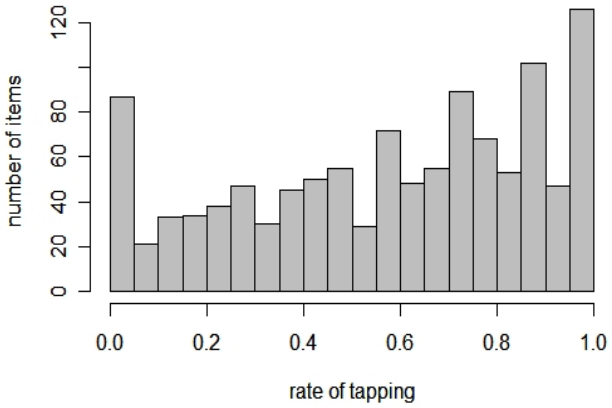
(4)

accessed: DalaDala		*(...VdV...)	MINIMALITY	STEMISHEAD	ALIGN (ACCU, L, PWd, L)	NORECURSION	*r
h	[(dala(dala)$_\omega$)$_\omega$]$_\varphi$			*!		*	
l	[(dalarala)$_\omega$]$_\varphi$			*!			*
☞ *j*	[(dala)$_\omega$(dala)$_\omega$]$_\varphi$						
k	[(dala)$_\omega$(rala)$_\omega$]$_\varphi$						*!
l	[(dala)$_\omega$]$_\varphi$[(dala)$_\omega$]$_\varphi$						
m	[(dala)$_\omega$]$_\varphi$[(rala)$_\omega$]$_\varphi$						*
n	[((dala)$_\omega$ dala)$_\omega$]$_\varphi$			*!		*	

(same outcome if *Dala* accessed)

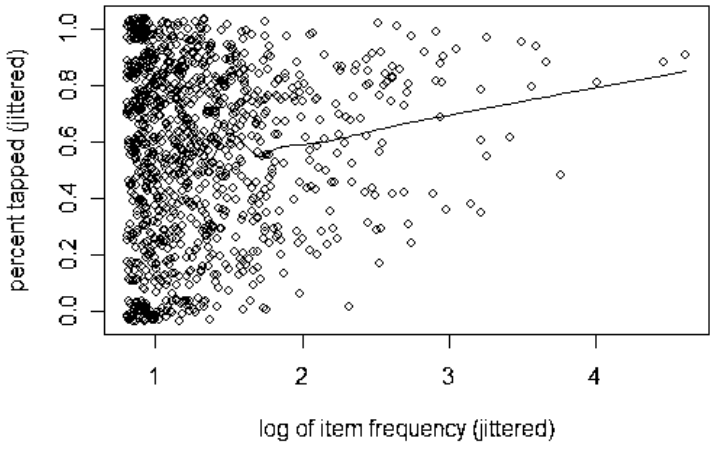
Is any of this really online? Or is it all lexicalized (reflecting diachronic effects)?

Clitics show real variation: *daw* 'reportedly' *ako **r**aw ~ ako **d**aw* 'me, reportedly'
din 'also' *ako **d**in ~ ako **r**in* 'me too'.

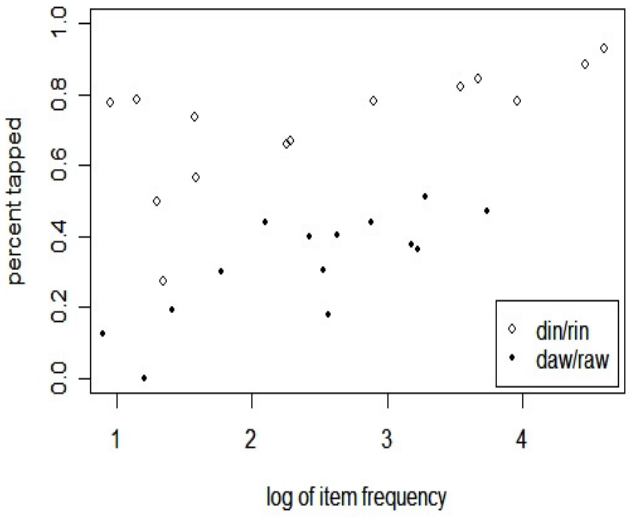


Weak, non-linear frequency effects:

all word+clitic combinations



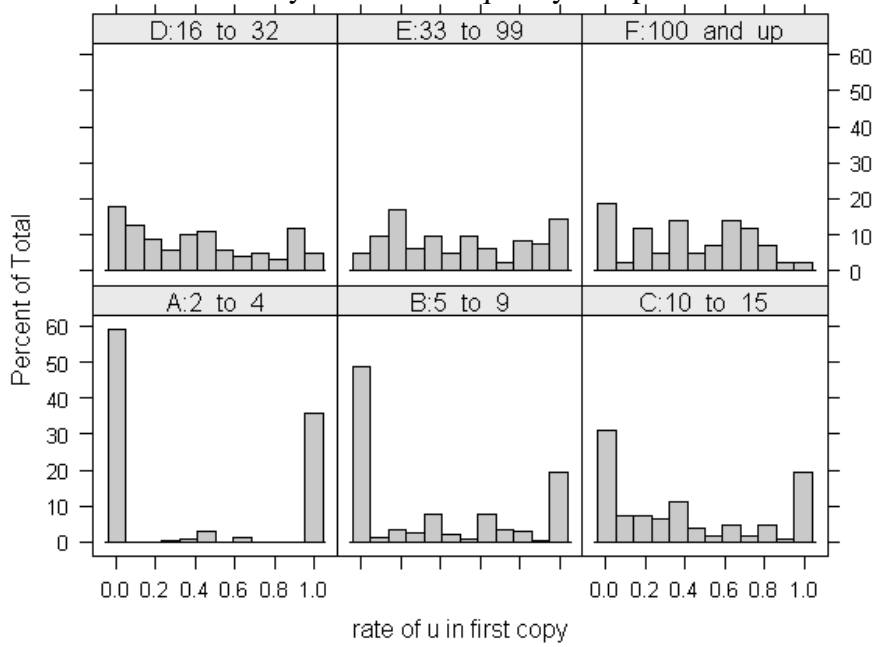
just the clitic+clitic combinations



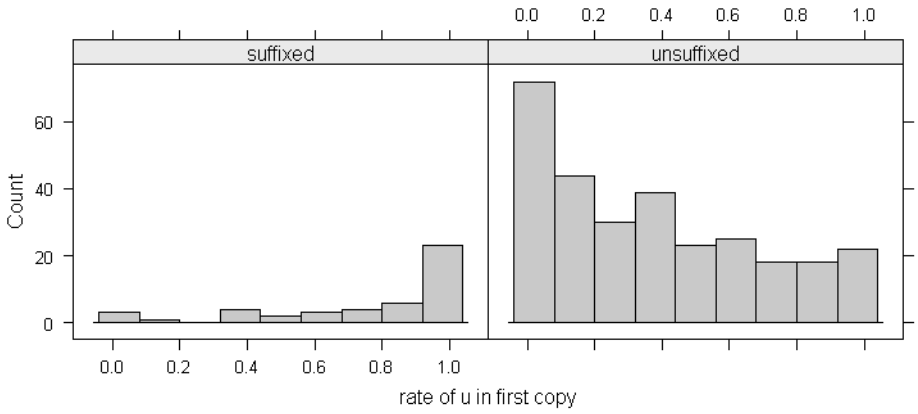
Similar results vowel-height alternation

halo ‘mixture’ hal**o**-halo ~ hal**u**-halo ‘(a dessert)’
(final-syllable [o] alternates with non-final syllable [u])

- ‘o’ forms are mostly in lowest-frequency reduplicated words:



- Grammar matters too: strong reduplicative identity effect
  - if second copy is forced to be [u] by suffixation, first copy is usually [u] too (*ka-tap**u**s-tapus-an* ‘very last’)



Similar results for nasal substitution

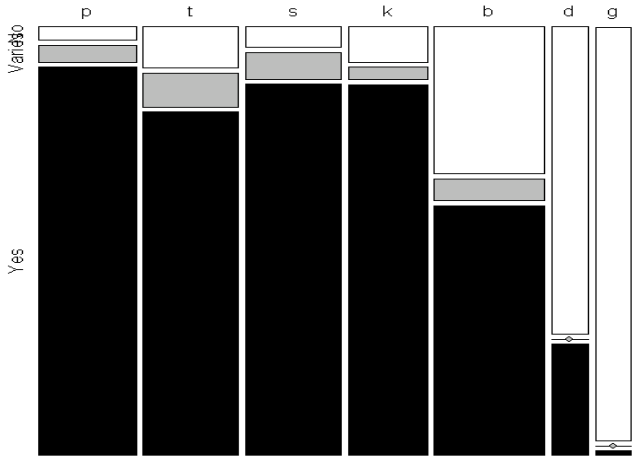
(see Zuraw 2010 for more on this rule)

Prefix-final nasal can fuse (or not) with following obstruent:

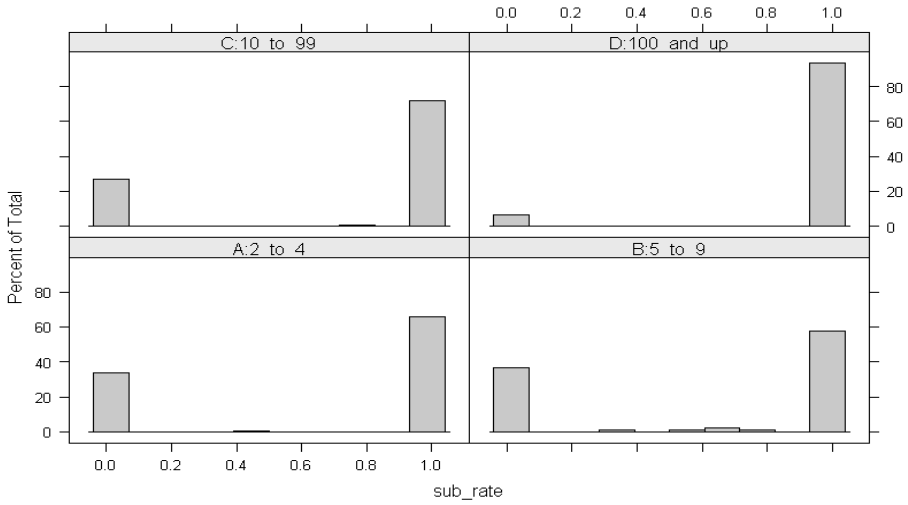
/paŋ+pasko/ 'for Christmas'

a.	non-assimilation	paŋ-pasko	<pang-pasko>
b.	assimilation	pam-pasko	<pam-pasko>
c.	nasal substitution	pamasko	<pamasko>

- Which obstruent it is matters a lot:



- But within the /b/s, where there are plenty of both types, frequency matters:



So where is this effect, really?

- Giving grammar a role seems to work well.
- But what if the grammatical effects be achieved by a diachronic model? Maybe this is all just information stored in lexical entries, perhaps reflecting lexical-access events from long ago.
- Sabbatical plans: if lexical access really is involved, it should be possible to affect a word's pronunciation through priming (temporarily perturbs the item's activation).

7 More proposals in which grammar refers (at least somewhat) directly to frequency

Can we think of ways to determine whether grammar makes direct reference to frequency, or sees only to the outcome of lexical access?

- Coetzee 2008: a lexical item's frequency determines how likely it is to be assigned to a given lexical class on any production occasion
- Myers 2005: how can lenition be both postlexical and sensitive to lexical frequency?
  - proposes a diachronic solution, where high frequency results in a more lenited lexical entry over time, but plays no synchronic role
  - diachronic and synchronic explanations should make different predictions about effects of priming on production...
- Alcántara 1998 (English): high-frequency exceptions can be protected by high-ranking idiosyncratic constraints
- Carlson & Gerfen 2011 (not a proposal about grammar, but a cool case): when a Spanish diphthong loses stress (say, because of suffixation), it should monophthongize. But it' variable:

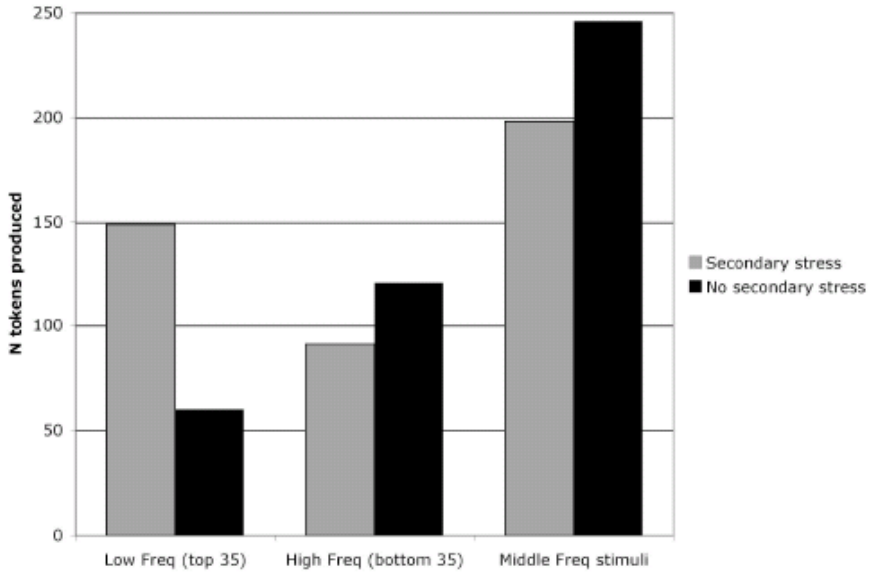
STEM		DERIVED FORM	
n[jé]ve	'snow'	n[e]vóso	'snowy'
verg[wé]nza	'shame'	verg[o]nzóso	'embarrassing'
v[jé]jo	'old'	v[je]jecíto	'little old man'
p[wé]blo	'town, people'	p[we]blíto	'little village'
cal[jé]nte	'hot'	cal[je]ntíto	'warm/cozy'
		cal[e]ntíto	

(p. 512)

The more productive the suffix (by corpus measures), the more likely to keep the diphthong.

- Gouskova & Roon 2008: in Russian compounds, the constraint requiring each stem to bear a prominence is ranked low, but there's a higher-ranked version of the constraint for low-frequency stems, forcing a secondary stress:

Figure 1: Effect of frequency on secondary stress realization



(p. 56)

If we have time, here's one more proposal I'd like to discuss...

8 Bermúdez-Otero forthcoming: two types of listing

- **Non-analytic listing**: output of stem level goes into the lexicon (fully prosodified)
  - such a listing blocks application of stem-level phonology, e.g. stress assignment, if faithfulness ranked high enough
  - → allows exceptional stress to survive (*Árabic*)
  - existence of such a listing **blocks** morphosyntactic synthesis in processing (you can't just compose *Arab+ic* or *drive+d*)
- **Analytic listing**: output of word level may go into the lexicon, but if it does it's listed as a concatenation of inputs to the word level
  - example: <LOAD, PAST> = $[_{\text{WORD LEVEL}}\ [_{\text{p-word}}\ (\text{lə}^{\mu}\upsilon^{\mu}\text{d})\ ] - \text{d}]$ (p. 23)
  - unable to block application of word-level phonology, e.g. [-d]~[-t]~[-ɨd] allomorphy
  - → no exceptions to word-level phonology allowed
- vs. plain old computation

Illustrated with a classic example

- *-al* is a stem-level suffix
- so */orígina**l**/* is listed non-analytically
- if you then want to derive *originality* (if you'd never heard it), you have to start with */orígina**l**/*; can't start with */origin+al/*
- faithfulness is ranked high: */orígina**l**+ity/* → *orìginálity*, not **òriginálity* (cf. *àbracadábra*, *dèlicatéssen*, *Mèditerránean*)

listed stem-level output: has full prosodification (Σ = foot)

	MAX-Head(Σ)	ALIGN(ω,L;Σ°,L)
[ω o[Σ° rí.gi]nal] – ity		
[ω [Σ° ò.ri]gi[Σ° ná.li]ty]	*!	
[ω o[Σ° rì.gi][Σ° ná.li]ty] ☜		*

(p. 28)

Chung's generalization (from Chung 1983)

A stem-level process can "cyclically misapply" iff it can have lexical exceptions in monomorphemes

Bermúdez-Otero's OT interpretation:

- High-ranking faithfulness are needed to ensure */orígina**l**+ity/* → *orìginálity* (= cyclic misapp.)
- This means you could have monomorphemic exceptions to the 'abracadabra rule' too : *Epàminóndas*, apparently (ancient Greek statesman)

		MAX-Head(Σ)	ALIGN(ω,L;Σ°,L)	ALIGN(Σ°,R;ω,R)
(a) *default pattern:* σ̆σ̆σ̆[Σ° σ́]σ	[ω σ̆[Σ° σ̆̀σ̆] [Σ° σ́]σ]		1!	2+1=3
	[ω [Σ° σ̆̀σ̆]σ̆[Σ° σ́]σ] ☜			3+1=4
(b) *exception:* [ω a[Σ° pò.the] [Σ° ó]sis]	[ω [Σ° à.po]the[Σ° ó]sis]	1!		3+1=4
	[ω a[Σ° pò.the][Σ° ó]sis] ☜		1	2+1=3

(p. 27)

Blocking can break down, though, because it happens in processing

- Nonanalytic entry [$_{\text{p-word}}$ (([ˈæ$^{\mu}$.ɹæ$^{\mu}$)bɪk)] (*Árabic*) races against synthesis, [$_{\text{STEM LEVEL}}$ æɹæb - ɪk]
- If the whole word isn't frequent enough, the entry isn't accessible enough, so it can lose out to synthesis, resulting in a regularized production.
- And if the exceptional form isn't produced often enough, the next generation won't learn it.

Frequency effects

- Classic cyclicity :

a.	cómp[ə]nsàte	còmp[ə]nsát-ion
	cónt[ə]mplàte	còmp[ə]mplát-ion
b.	cond[ɛ́]mn	còmd[ɛ̀]mn-átion
	imp[ɔ́]rt	ìmp[ɔ̀]rt-átion

(p. 30)

- but :

a.	cons[ɜ́]rve	b.	còns[ə]rv-átion
	tràns[ɔ́]rt		tràns[ə]rt-átion

(p. 30)

- The reason is frequency :

(33)

			(× per 10^6 words in spoken section of COCA)		
			base		derivative
a.	*cyclic stress*				
	cond[ɛ́]mn	còmd[ɛ̀]mn-átion	7.09	>	2.57
	imp[ɔ́]rt	ìmp[ɔ̀]rt-átion	5.15	>	0.62
b.	*variable stress*				
	cond[ɛ́]nse	còmd[ɛ̀-ə]ns-átion	0.28	≈	0.22
c.	*noncyclic stress*				
	cons[ɜ́]rve	còns[ə]rv-átion	1.65	<	9.11
	tràns[ɔ́]rt	tràns[ə]rt-átion	7.23	<	23.54

(p. 32)

See Collie 2008 for a full study

To sum up today

- We looked at several cases of lexical frequency's influencing phonology.
- We considered putting the explanation in diachrony, processing, and/or grammar.

Thursday (last class)

- Getting phonological evidence

References

Alcántara, Jonathan B. 1998. The Architecture of the English Lexicon.. Cornell University.

Bermúdez-Otero, Ricardo. The architecture of grammar and the division of labour in exponence.. In Jochen Trommer (ed.), *The morphology and phonology of exponence: the state of the art*. Oxford University Press.

Bybee, Joan L. & Dan I. Slobin. 1982. Rules and Schemas in the Development and Use of the English past Tense. *Language* 58(2). 265–289. doi:10.2307/414099 (13 March, 2012).

Carlson, Matthew T & Chip Gerfen. 2011. Productivity is the key: morphophonology and the riddle of alternating diphthongs in Spanish. *Language* 87(3). 510–538. (13 March, 2012).

Chomsky, Noam & Morris Halle. 1968. *The Sound Pattern of English*.. Harper & Row.

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

Coetzee, Andries W. 2008. Phonological variation and lexical frequency. *Porceedings of NELS 38*.

Collie, Sarah. 2008. English stress preservation: The case for "fake cyclicity." *English Language and Linguistics* 12(03). 505–532. doi:10.1017/S1360674308002736 (7 March, 2011).

Gouskova, Maria & Kevin Roon. 2008. Interface constraints and frequency in Russian compound stress.. In Jodi Reich, Maria Babyonyshev, & Darya Kavitskaya (eds.), *Proceedings of FASL 17*. Ann Arbor, MI: Michigan Slavic Publications.

Hammond, Michael. 1999. Lexical frequency and rhythm.. In Mike Darnell (ed.), *Functionalism and Formalism in Linguistics, Volume I: General Papers*, 329–358. John Benjamins.

Hay, Jennifer. 2003. *Causes and consequences of word structure*.. Routledge.

Kirby, S. 2001. Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation* 5(2). 102–110. doi:10.1109/4235.918430.

Löfstedt, Ingvar. 2010. Phonetic Effects in Swedish Phonology: Allomorphy and Paradigms.. UCLA ph.d. dissertation.

Myers, James. 2005. Frequency effects and Optimality Theory.. Handout. ROA #810-0306.

Ng, E-Ching. 2010. Reduction, frequency and morphology in Singaporean English prosody.. Manuscript. Yale University, ms. ROA #1102.

Oxford English Corpus. Oxford English Corpus: Facts about the language. http://oxforddictionaries.com/words/the-oec-facts-about-the-language.

Pinker, Steven. 2000. *Words and rules: the ingredients of language*.. HarperCollins.

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.