7 March 2011

Week 10, prosem. on lexical access and the phonology of morphologically complex words
Attempts at synthesis

(1) Overview

- We've seen a lot of (sometimes conflicting) results about when and whether related words affect each other in processing.
- Rather than summarize those results again—since we can always turn back to our handout collection—I want to speculate forwards a bit.
- First we'll look at a couple of attempts to make lexical access and grammar communicate.
- Then we'll revisit some phonological proposals and think about how amenable they'd be to such attempts.

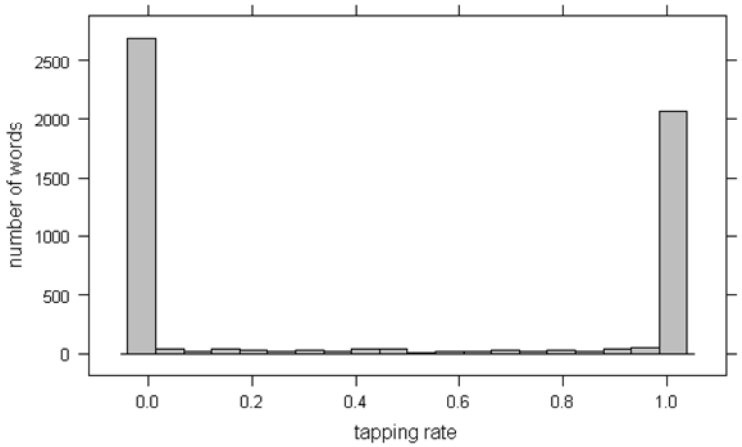
1 How can models of lexical access communicate with grammar? Zuraw 2009

This is work that Kevin Ryan and I got started on—he did all the phonetic work—but it stalled somewhere short of a full model. I'd hoped that this proseminar could fill in some gaps for me so that we could return to it...

(2) 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'

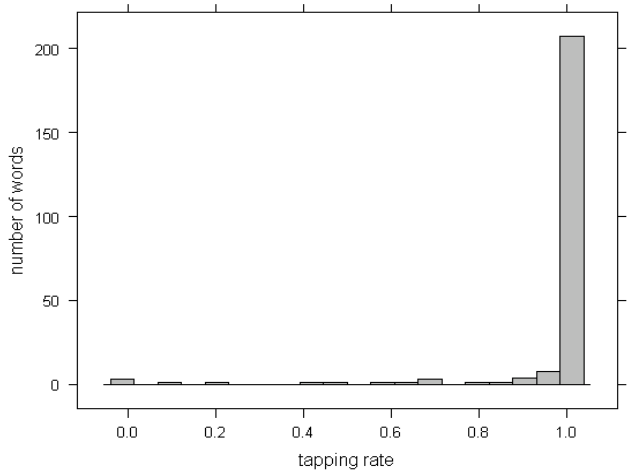
Each word seems to have a consistent behavior (using spelling data in corpus):



The more frequent the word, the more likely tapping, suggesting that processing route matters (with the result maybe getting lexicalized?)

(3) Tapping in suffixed words: obligatory

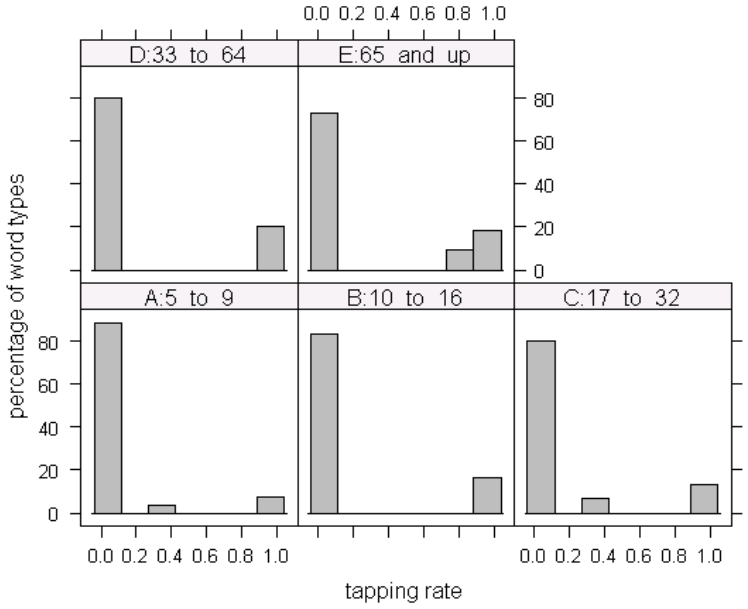
lakad 'walk' *laka**r**-an* 'to be walked on'



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

(4) Tapping in p-word reduplication: nearly forbidden

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



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

(5) Analysis: grammar refers to outcome of lexical access

ALIGN(AccU,L,PWd,L): the left edge of any accessed lexical unit must coincide with the left edge of some prosodic word.

→ outcome for prefixed word depends on access mode:

(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)$_\omega$		*(!)	*(!)	*	*	

(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)$_\omega$		*(!)	*(!)		*	

(6) Analysis: grammar refers to outcome of lexical access

- Accessed units partly determine p-word structure.
- Outcome for suffixed or p-word-reduplicated words is fixed, because the constraint that depends on access mode is low-ranked:

(3)

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

(same outcome if just *lakaDan* accessed)

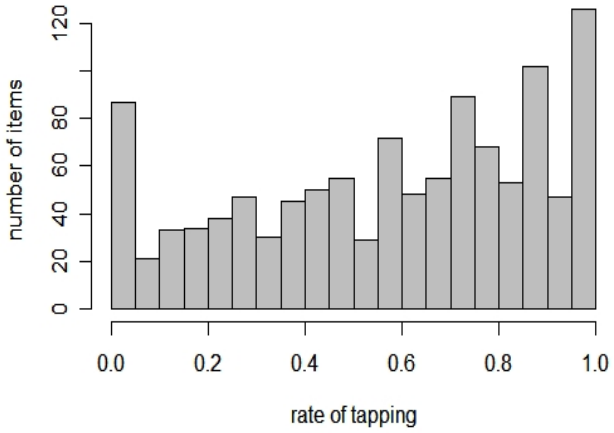
(4)

	accessed: DalaDala	*(...VdV...)	MINIMALITY	STEMISHEAD	ALIGN (ACCU, L, PWd, L)	NORECURSION	*ɾ
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)$_\omega$]$_\varphi$			*!		*	

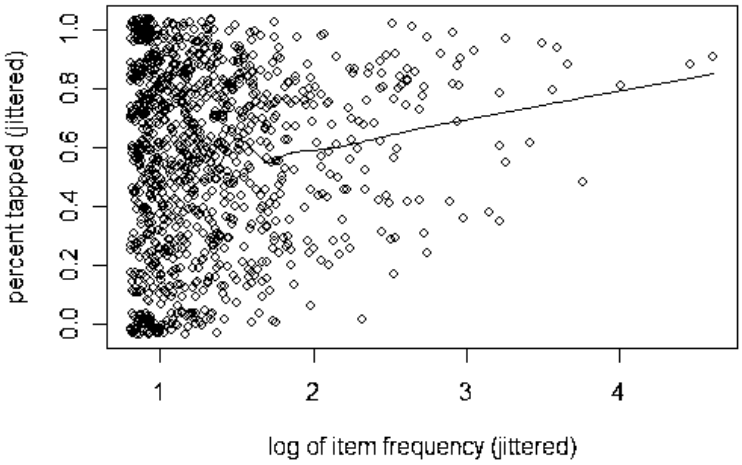
(same outcome if *Dala* accessed)

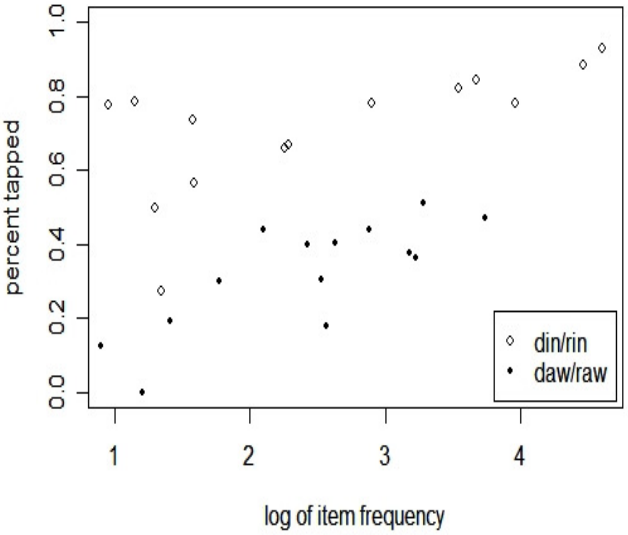
(7) Is any of this really online? Or is it all lexicalized?

Clitics show real variation: *daw* 'reportedly' *ako raw ~ ako daw* 'me, reportedly'
din 'also' *ako din ~ ako rin* 'me too'.



Weak (and prob. non-linear) frequency effects:



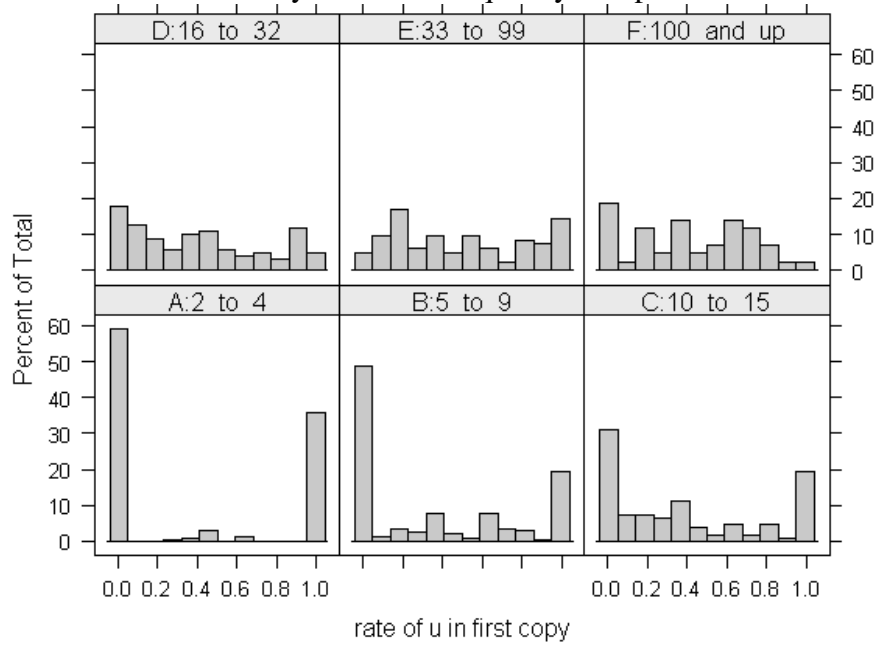


Just the clitic+clitic combinations (*ako pa rin* 'still me also'), so fewer data points

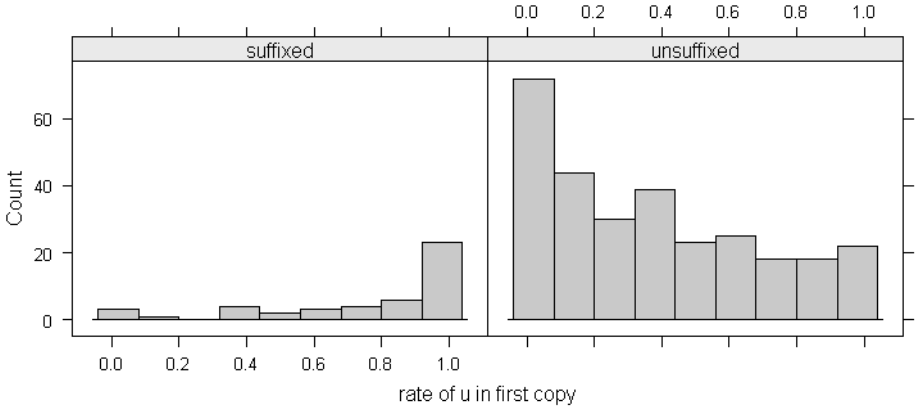
(8) Similar results vowel-height alternation

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')



(9) Similar results for nasal substitution

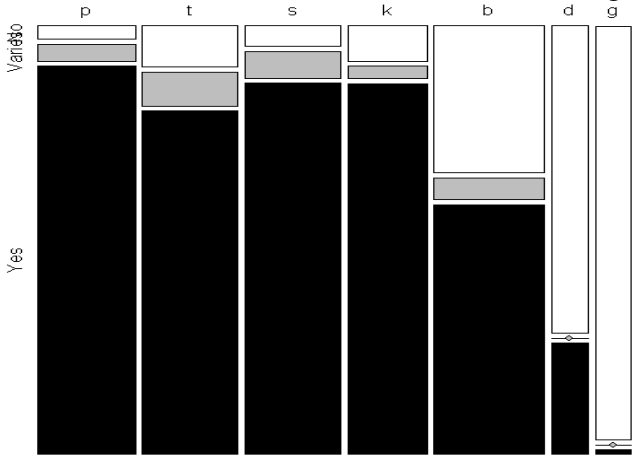
(see also Zuraw 2010 for more on this rule)

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

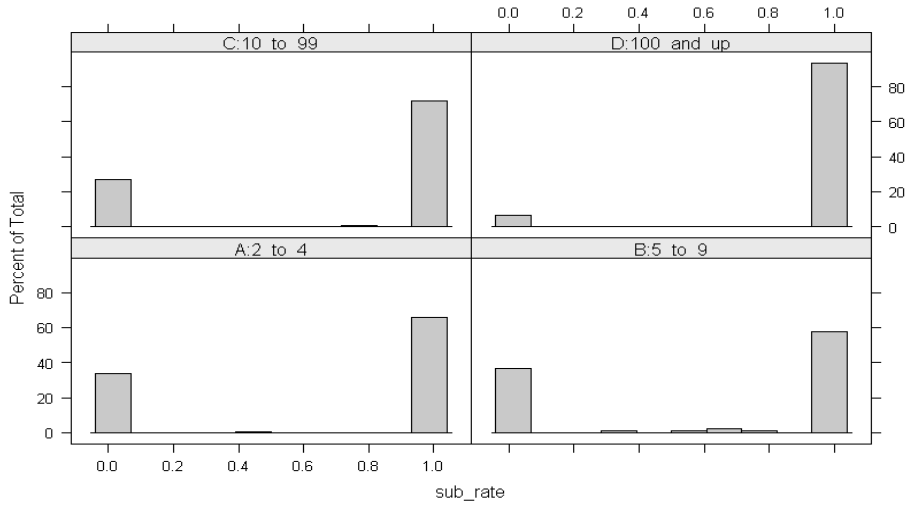
/paŋ+pasko/ 'for Christmas'

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 (here comes the graph I've made a million times)



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



(10) How to revisit this in light of the course material

- Your suggestions and thoughts, please!
- Just about any model of lexical access in production can plug into this (how to test which one works best?), as long as it allows both decomposed and non-decomposed access of the crucial items.
- How to test which parts of this are online and which parts are lexicalized?

2 Lexical access communicating with grammar II: Bermúdez-Otero forthcoming

(11) Previous views of storage vs. computation

- Chomsky & Halle 1968: an alternation exhibiting any psychological reality (operationalization to be determined) belongs in the grammar
  - because the alternative is for it to be in the lexicon, which is « cognitively inert »
  - e.g. *string-strung* ablaut must be in the grammar, because of innovations like *sneak-snuck*
  - → we need diacritics ([+ablaut]) to determine which words a rule applies to
- Evaluation metric enforces this
  - if a rule is *n* symbols long and saves you *m* symbols in the lexicon, if *m*>*n*, learn the rule
  - (cost of computation is ignored)

example

- Pure listing costs about 78 symbols :
  - *sing*, *sang, ring, rang, shrink, shrank, drink, drank, sink, sank, begin, began, swim, swam, stink, stank, spring, sprang...* (78)
- Having a rule costs less in total, even counting the rules' symbols ungenerously :
  - $sing_{[abl]}$ $ring_{[abl]}$ $shrink_{[abl]}$, $drink_{[abl]}$, $sink_{[abl]}$, $begin_{[abl]}$,$swim_{[abl]}$, $stink_{[abl]}$ (48)
  - $V \rightarrow$ [+ablaut] / $[_V [_V \_\_C_0]_{[abl]\ PAST}]$ (12 ?)
  - [i,+ablaut] → [æ] (3 ?)

(12) Bermúdez-Otero's view: two types of listing

If the lexicon is *not* inert, then some psychologically real phenomena can be left there

- 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 (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
  - 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

So, we have some regularities residing in listed outputs of the stem level.
Others emerge through associative memory (*string-strung*, etc.)

(13) Sample lexical entries

(19) *Nonanalytic listing of a stem-level form:* idyllic

a. ✓ IDYLLIC ↔ [ω [Σ′ ɪ[Σ° 'dɪ$^{\mu}$.lɪ$^{\mu}$k]]] = (18a)

b. × IDYLLIC ↔ [SL ɪdɪl - ɪk]

p. 23

(21) *Nonanalytic lexical entry for the adjective stem* Arabic

a. ✓ ARABIC ↔ [ω [Σ′ [Σ° 'æ$^{\mu}$.ɹæ$^{\mu}$]bɪk]] (C-extrametricality and penult stress blocked)

b. × ARABIC ↔ [SL æɹæb - ɪk] (→SL *[ω [Σ′ æ[Σ° 'ɹæ$^{\mu}$.bɪ$^{\mu}$k]]])

p. 24

(20) *Analytic listing of a word-level form:* loaded

a. ✓ <LOAD, past> ↔ [WL [ω [Σ lə$^{\mu}$ʊ$^{\mu}$d]] - d]

b. × <LOAD, past> ↔ [ω [Σ 'lə$^{\mu}$ʊ$^{\mu}$]dɪd]

p. 23

(14) 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* (say you've 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*)

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

(p. 28)

(15) Chung's generalization (from Chung 1983**)**

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

- High-ranking faithfulness needed to ensure that */orígina**l**+ity/* → *orìginálity*
- 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° σ̀σ̆] [Σ° σ́]σ]		1!	2+1=3
	☞ [ω [Σ° σ̀σ̆] 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)

(16) Blocking can break down, though, because it happens in processing

- Nonanalytic entry [ω [Σ′ [Σ° ˈæ$^{\mu}$.ɹæ$^{\mu}$]bɪk]] (Árabic) races against synthetic [$_{SL}$ æɹæb - ɪk]
- If the whole word isn't frequent enough, it can lose out to synthesis, and therefore regularization, in production
- and if the exceptional form isn't produced often enough, next generation won't learn it

(17) Frequency effects

- Classic cyclicity :

a.	cómp[ə]nsàte	còmp[ə]nsát-ion
	cónt[ə]mplàte	còmt[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

- Let's step through how this would work

3 Some case studies—let's discuss them in light of lexical access

(18) Ng 2010**: Singapore English prosodic boundaries**

- Strong glottalization at prefix-stem, stem-stem, but not stem-suffix boundary
  - *mis-understand* [mis**ʔ**ɑndəstæn], *stopover* [stɔp **ʔ**ovə], *magic-al* [mædʒikØəɯ] (p. 8)
  - analyzed in terms of p-word structure

- Stress is realized as tone: $L_0M^*M_0H$ (first stressed syllable gets M tone, preceding syllables get L, following syllables get M except the last, which gets H; if just one syllable, H)
  - tone pattern generally re-starts in compounds: *century egg* (ˈMH)˛H) (p. 13)
  - tone pattern may or may not restart at prefix-stem boundary: *ˌun-inˈstall* (˛H)-(LˈH)˛, *un-inˈstall* (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'), where there are frequency effects
  - higher-frequency initialisms seem to show more prosodic merger:

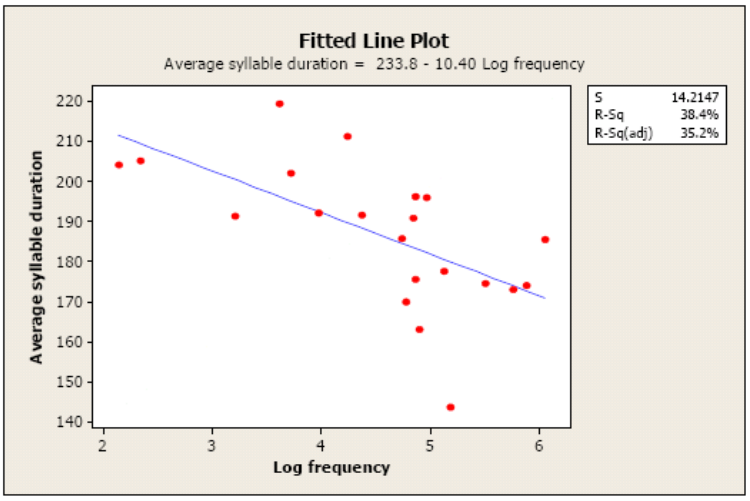
Society for the Prevention of Cruelty to Animals, Anglo-Chinese Junior College, National Registration Identity Card, National Trade Unions Congress (supermarket)

		Least merger			*Most merger*
a.	*SPCA*	((((ˈH)˛H)˛H)˛H)			
b.	*ACJC*	((((ˈH)˛H)˛H)˛H)	(((ˈH)˛H)˛MH)		
c.	*NRIC*	((((ˈH)˛H)˛H)˛H)	(((ˈH)˛H)˛MH)	((ˈH)˛MMH)	
d.	*NTUC*	((((ˈH)˛H)˛H)˛H)	(((ˈH)˛H)˛MH)	((ˈH)˛MMH)	(ˈMMMH)

(p. 23)

- 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 at all...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	*SCLASHS	STRESS
☞ a. ((ˈM)˛OE)			*
b. ((ˈM)˛O˛E)		W*	L
c. ((ˈM)(˛O)(˛E))		W*	L
d. (ˈM)(ˈO)(ˈE)	W*		L

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

(p. 33)

(19) 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.

(20) Löfstedt 2010**: frequency-specific constraints**

- 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'

(p. 152)

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

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

(p. 154)

quality change is too big

- 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[ɑː]d	INEFFABLE[51]	'straight'	0	7,140
l[ɑː]t	INEFFABLE	'lazy'	0	581,000
gl[ɑː]d	gl[a] + t:	'happy'	29	2,110,000

(p. 154)

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

	/glad + t:/ 'happy' neut. Cf. [glɑːd] (*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.	glɑːt:	*!						
b.	glat:		*!					
c.	glɑːt			*!				
d. >	glat:							*
e.	⊙						*!	

(p. 167)

(21) 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'		*

(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 words' 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

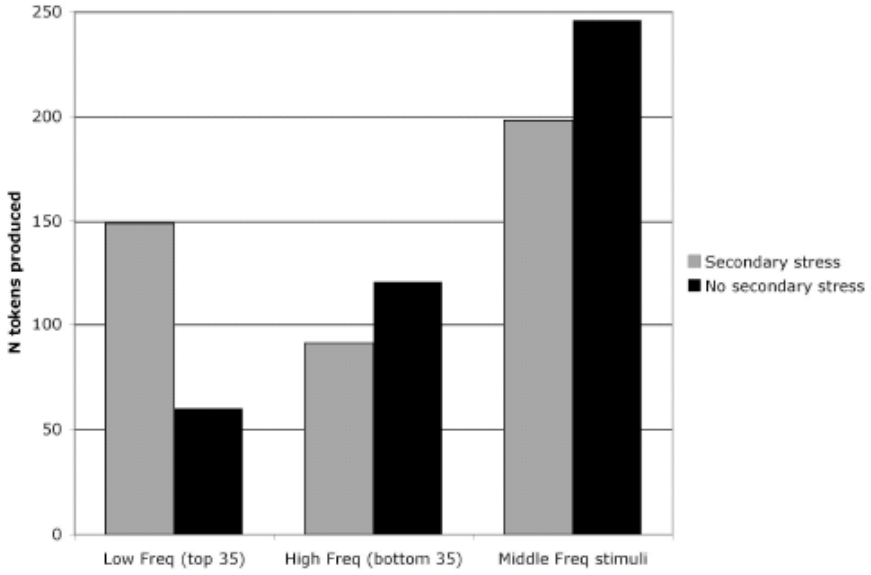
(22) More proposals in which grammar refers (at least somewhat) directly to frequency

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

- 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)

Next week we'll hear from some of you about things you've been working on or thinking about.

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.

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. 2008. Phonological variation and lexical frequency. In , *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.

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.

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.

Zuraw, Kie. 2009. Frequency influences on rule application within and across words. In , *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.