Class 11: The zero variant

1   Paradigm gaps

I've discovered that I'm lactose-intolerant. Sadly, from now on, I'll have to forgo dairy. This afternoon I _____ the ice cream I normally would have relished on such a hot day.

- The funny thing is that speakers know exactly what the options are and can produce them—*forgoed, forwent*—but they just feel icky about both options.
- The most famous and surprising paradigm gaps involve productive inflectional morphology:
    - 'He abolishes' in Spanish: *abuele, *abole (see Albright, Hayes & Andrade 2001)
    - 'I win' in Russian: *pobežu (see Daland, Sims & Pierrehumbert 2007)
    - 'We fry' in French: *nous frions, *nous fritons, *nous frisons (Baronian 2009)

2   Overview of today

- Variable, phonologically driven paradigm gaps in derivational morphology.
- One way of modeling these: having the grammar assign absolute goodness scores
- A case study or two
- Where else might we want goodness scores?

3   A well-known English allomorphy: *-ar* vs. *-al*

- *-ar* is supposed to be the allomorph that occurs when the stem's last liquid is /l/
    - accident-a**l**   vs.   consu**l**-a**r**   (Raffelsiefen 1999)
- Counts from CELEX (XXref) English lemma list, Noun+*al/ar* → Adj only

noun stem has	all nouns[1]	-al suffixed	-ar suffixed
no liquid	*4499*	*113 (2.5%)*	*0*
just [r]	381	115	0
[r...r]	474	25	0
[l...r]	496	10	0
[r..l...r]	23	3	0
[l..l...r]	3	0	0
[l..r...r]	11	0	0
[r..r...r]	3	0	0
[r...r...r...l]	1	0	0
subtotal	*1392*	*153 (11.0%)*	*0*
just [l]	2599	24	12
[l...l]	107	0	2
[r...l]	678	3	3
[l...r...l]	15	0	0
[r...l...l]	2	0	0
[r...r...l]	11	0	0
[r...l...r...l]	1	0	0
subtotal	*3413*	*27 (0.8%)*	*17 (0.5%)*

Counts too big for Fisher's Exact Test.
Chi-square test
(2x2: last liquid=l or no, takes suffix or no)
gives *p*< 0.00001

⇒ if no liquid, or last liquid is [r], use –*al*. If last liquid is [l], variation.

---

[1] Nouns with no hyphen or space, not tagged as affixed or compound.

- But also: nouns whose last liquid is [l] are less likely to take this suffix at all!
- We could say that –*al/-ar* has a <u>partial paradigm gap</u> for these stems.

- This looks like a gradient version of a starker gap noted (non-quantitatively) by Raffelsiefen 1999.
- The –*al* that forms nouns from verbs has only one allomorph. If there's an *L...L issue, the word is simply <u>unutterable</u>:
  - deny, deni-a**l**   vs. re**l**y, *re**l**i-a**l**, *re**l**i-a**r**

verb stem has	all verbs[2]	-al suffixed
no liquid	*506*	*10 (2.0%)*
just [r]	483	22
[r...r]	73	4
[l...r]	44	0
[r..l...r]	1	0
[r..r...r]	3	0
total	*604*	*26 (4.3%)*
just [l]	318	0
[l...l]	6	0
[r...l]	69	0
total	*393*	***0 (0%)***

Chi-square test (2x2): *p*=0.0004

4   One way to model variable gaps: absolute rather than relative well-formedness

(Bruce will discuss another way Thursday—competition with the null parse)

- <u>Schematically</u>: typical MaxEnt says these tableaux are the same—50-50 variation:

input1	CONSTRAINT1 *weight*=1	CONSTRAINT2 *weight*=1	harmony	probability
output1a	*		e^{-1}	0.5
output1b		*	e^{-1}	0.5

input2	CONSTRAINT1 *weight*=1	CONSTRAINT2 *weight*=1	harmony	probability
output2a	**		e^{-2}	0.5
output2b		**	e^{-2}	0.5

- But in absolute terms, *output1a* is better than *output2a*.
- So what if we try to model utterability as a function of <u>harmony</u> (not of relative harmony, i.e. probability)?
  - Coetzee & Pater 2005 propose this, but using Harmonic Grammar (so no exponentiation)
- This allows us to capture the difference between two good options and two bad options.

[2] Nouns with no hyphen or space, not tagged as affixed or compound.

Application to adjectival –al/-ar

- I don't know what the function from harmony to observed probability should look like, but suppose we want it to be linear. Then, these weights work pretty well:

gloam+{al,ar}	*L...L w=2.635	*R...R w=2	*LoneLiquid w=1.465	Prefer-al w=3.1	harmony	observed probability
gloamal	*				0.072	27/3413 = 8%
gloamar				*	0.045	17/3413 = 5%

stobe+{al,ar}	*L...L w=2.635	*R...R w=2	*LoneLiquid w=1.465	Prefer-al w=3.1	harmony	observed probability
stobal			*		0.231	113/4499 = 3%
stobar			*	*	0.010	0/4499 = 0%

criff+{al,ar}	*L...L w=2.635	*R...R w=2	*LoneLiquid w=1.465	Prefer-al w=3.1	harmony	observed probability
criffal					1.000	153/1392 = 11%
criffar		*		*	0.006	0/1392 = 0%
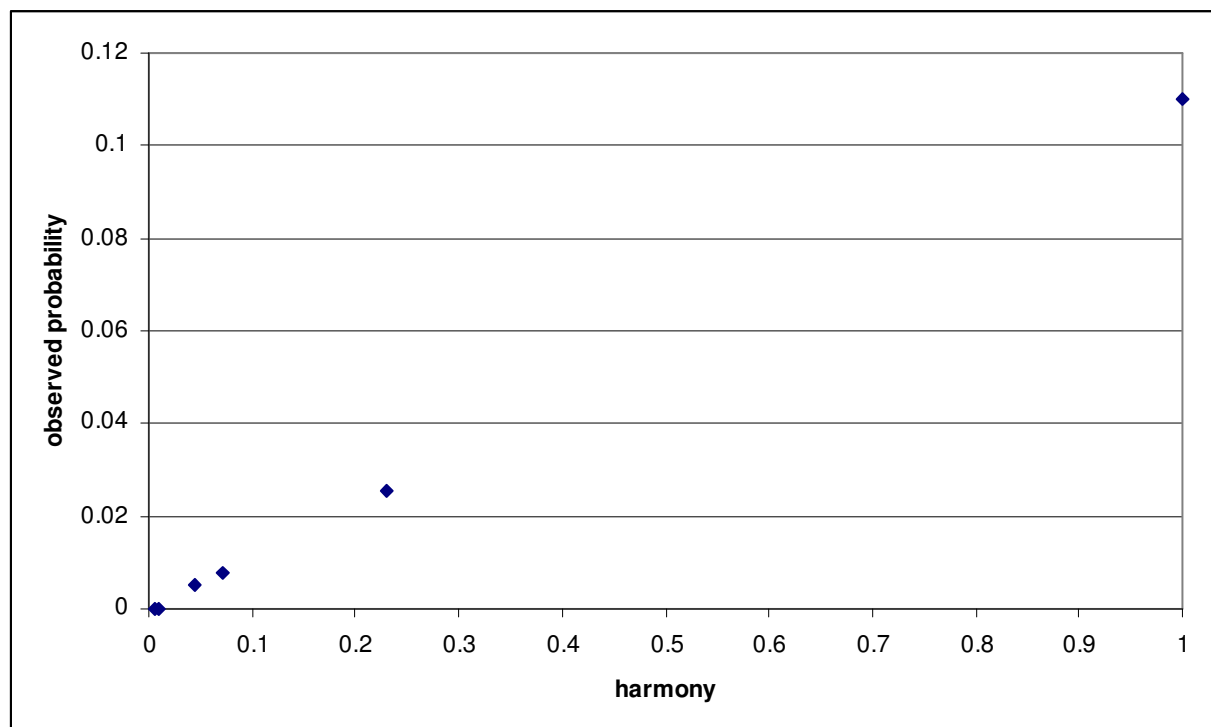
- The idea of absolute well-formedness is well-established in phonotactics.
    - It's not a question of *what* the best output for /smin/ is, but rather *how good* that output is.
    - See Frisch, Pierrehumbert & Broe 2004, Coetzee & Pater 2005; Daland et al. 2011 for overview and empirical comparison.

5   How big should the tableau be? How specific is the input?

- Smallest tableau

hobby-horse + {-al, -ar}	*L...L	*R...R	*LONELIQUID	PREFER-AL
hobby-horsal				
hobby-horsar		*		*

- A little broader

hobby-horse + {-al, -ar, -ical}	*L...L	*R...R	*LONELIQUID	PREFER-AL	ENDDACTYL?
hobby-horsal					*
hobby-horsar		*		*	*
hobby-horsical[3]				*	

- Broad

hobby-horse → *adjective*	*L...L	*R...R	...
hobby-horsal			
hobby-horsar		*	
hobby-horsical			
hobby-horsic			
hobby-horsy			
hobby-horsish			
hobby-horsian			
hobby-horsoid			
hobby-horsesque			
...			

- Really broad

express the idea of 'related to a pet topic'	...
hobby-horsal	
hobby-horsical	
hobby-horsian	
characteristic of a hobby-horse	
pet-topic-related	
having to do with his favorite subject	
or even just behave suitably:	
you know how he always likes to talk about fortifications and stuff	
How about this weather we're having?	
[get up to refill guest's glass]	
...	

- If we assign each candidate a goodness score, it doesn't matter if they're in the same tableau or different tableaux.

---

[3] Lawrence Sterne's choice in *Tristram Shandy*.

6   Some more trends in Raffelsiefen 1996, Raffelsiefen 1999

Raffelsiefen is careful to point out: English borrowed many words already containing these suffixes, which can violate generalizations below—generalizations apply only to new coinages.

- *-ize* doesn't like to attach to after a stressed syllable (1996 p. 194)
  - síster-ìze   *obscéne-ize
  - sálmon-ìze   *ápt-ize

- *-eer* also doesn't like to attach after a stressed syllable (1996 p. 207, 1998 p. 231)
  - mùffin-éer   *baguètt-éer (though this could be an issue with /t/--uncirculated talk by Albright)
  - jàrgon-éer   *strìke-éer
  - càmel-éer   *giràff-éer
  - pìgeon-éer   *dòve-éer

  but *–ee* is fine with it:
  - assìgn-ée
  - appòint-ée

- *-ize* and *–ee* don't like to attach after [i] (1996 p. 202)
  - *sílly-ìze
  - *mónkey-ìze
  - rèscue-ée   *càrry-ée

  except where deletion is allowed:
  - mémor-ìze (< mémory)
  - apóstroph-ìze (< apóstrophe)

- *-ee* doesn't like to attach after any vowel (1998 p. 246)
  - mùsket-éer   *bazòoka-éer

- *-ese* doesn't like to attach after any V, but can delete a V or insert a C (1998 p. 246)
  - Chìn-ése (< Chína)
  - Jàvan-ése (< Jáva)
  - Bàlin-ése (< Báli)

- *-ize* also doesn't like to attach after Obstruent$_i$RimeObstruent$_i$ (1996 p. 200)
  - *crísis-ìze
  - *thésis-ìze
  - cótton-ìze   *línen-ìze
  - vígor-ìze   *hórror-ìze
  - Stálin-ìze   *Lénin-ìze
  - líquid-ìze   *cándid-ìze

except where it can delete to solve the problem:
> émphas-ìze (< émphasis)
> óptim-ìze (<óptimum)
> fémin-ìze (<féminine—cf. másculin-ìze, not *máscul-ìze)
> áppet-ìze (<áppetite)

similarly (1998, p. 243):
> trepíd-ity   *candíd-ity
> lámb-ify   *béef-ify

- *-ish* doesn't like to attach right after a postalveolar (1998 p. 241)
  > shéep-ish   *físh-ish
  > cánnibal-ish   *rúbbish-ish

- similarly... (1998 pp. 241-243)
  > flávor-ous   *bías-ous
  > pìstol-éer   *revòlver-éer
  > kìtchen-étte   *clòset-étte
  > shórt-age   *lárge-age
  > vást-ity   *fáint-ity

- The *–en* that forms V from A requires a monosyllable ending in an obstruent
  > wéak-en   *vívid-en   *wárm-en

  doesn't apply to other –ens:
  > wool-en (N → A)
  > fall-en (past participle)

7   A case to model: the choice of –y, -ly, -ish, -ful

- I extracted the ~11,000 noun lemmas from CELEX
- About 700 of them take at least one of these suffixes
- I threw in some plausible constraints:

*MONOSYLLABLE+Y	*MONO+LY	*MONO+LIKE	MONO+ISH	*MONO+FUL
*STRESSED+Y	*STRESSED+LY	*STRESSED+LIKE	*STRESSED+ISH	*STRESSED+FUL
*UNSTRESSED+Y	*UNSTR+LY	*UNSTR+LIKE	*UNSTR+ISH	*UNSTR+FUL
*V+Y	*C+LY	*C+LIKE	*V+ISH	*C+FUL
*[i]+Y	*[l]+LY	*[l]+LIKE	*[i]+ISH	*[f]+FUL
	*[l]...+LY	*[l]...+LIKE	*[ʃ]+ISH	*[f]...+FUL
			*[ʃ]...+ISH	
			*POSTALV+ISH	
			*POSTALV...+ISH	

- Fitting a MaxEnt model (with just one input, "dummy input") and all noun-suffix combinations as outputs (some with freq. 0, others with freq. 1), was taking too long.

- So, I went with logistic regression. Dependent variable is attestedness.
  - This means some of the constraints can actually bestow bonus points rather than penalties.

Result:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept: ful)	-4.345519	1.008568	-4.309	1.64e-05	***
suffix=ish	0.385621	1.361636	0.283	0.77702	
suffix=like	-0.448268	1.436270	-0.312	0.75496	
suffix=ly	-0.631347	1.377151	-0.458	0.64663	
suffix=y	2.361594	1.603819	1.472	0.14089	
y____mono	-4.978219	1.120391	-4.443	8.86e-06	***
y____final_stress	2.511399	1.649652	1.522	0.12791	
y____final_stressless	-0.003126	1.460663	-0.002	0.99829	
V_y	-0.830426	0.308917	-2.688	0.00718	**
i_y	-3.530299	1.533468	-2.302	0.02133	*
ly____mono	-0.638485	0.495280	-1.289	0.19735	
ly____final_stress	-0.346992	1.247249	-0.278	0.78085	
ly____final_stressless	-0.234919	1.223998	-0.192	0.84780	
C_ly	0.849465	0.412582	2.059	0.03950	*
l_ly	-1.064240	0.855053	-1.245	0.21326	
l_dot_dot_dot_ly	-0.502848	0.337780	-1.489	0.13657	
like____mono	-2.464277	1.240523	-1.986	0.04698	*
like____final_stress	0.035957	1.473864	0.024	0.98054	
like____final_stressless	-0.468508	1.327181	-0.353	0.72408	
C_like	0.087791	0.590673	0.149	0.88185	
l_like	-0.730362	0.891083	-0.820	0.41243	
l_dot_dot_dot_like	0.266909	0.460607	0.579	0.56227	
ish____mono	-2.381054	0.766904	-3.105	0.00190	**
ish____final_stress	0.141440	1.306048	0.108	0.91376	
ish____final_stressless	0.226993	1.236040	0.184	0.85429	
sh_ish	-0.396604	1.961895	-0.202	0.83980	
postalv_ish	-2.371327	1.521538	-1.559	0.11911	
sh_dot_dot_dot_ish	1.207102	1.068041	1.130	0.25839	
postalv_dot_dot_dot_ish	-1.369953	1.006718	-1.361	0.17357	
V_ish	-1.174535	0.742487	-1.582	0.11367	
i_ish	0.326065	0.931847	0.350	0.72640	
ful____mono	0.080480	0.273154	0.295	0.76828	
ful____final_stress	-1.805414	1.045057	-1.728	0.08406	.
ful____final_stressless	0.665338	1.017251	0.654	0.51308	
C_ful	-0.209413	0.298990	-0.700	0.48368	
f_ful	-2.139546	1.454674	-1.471	0.14134	
f_dot_dot_dot_ful	0.195713	0.351844	0.556	0.57804	

Annotations alongside the table:

- Each suffix gets its own baseline productivity
- *-y* hates attaching to monosyllables, Vs, and especially [i]
- *-ly* is pretty easy-going; prefers to attach to C rather than V
- *-like* is also pretty easy-going; doesn't like attaching to monosyllables,
- *-ish* also doesn't like attaching to monosyllables. Surprisingly, OCP effects aren't significant.
- *-ful* is easygoing

## 8   What would this look like in language use?

- Imagine the regression model as a MaxEnt grammar instead (i.e., with suitable adjustments to avoid negative weights)
- You want to make *cash* into an adjective.
- A suffix comes to mind, such as *–ish*
- You run it through the grammar, and see how good the winner sounds.

	probability under log. reg. model above
cash-y	0.1206
cash-ly	0.0067
cash-like	0.0056
cash-ish	0.0013
cash-ful	0.0200

- If not good enough, derivation crashes and you try again.
  - See Orgun & Sprouse 1999 for the idea of CONTROL, a component that checks the output of the normal Gen-Eval process; if it violates certain constraints, derivation crashes.
- If nothing acceptable comes to mind in time, you paraphrase, hesitate, or change the subject.
  - E.g., Martin's Turkish case that we discussed earlier: if a compound isn't good enough to be a "lexical compound", you can make it an izafet compound instead.

9   What would this look like over generations? Martin 2007 in a nutshell

- Uncontroversial assumption: language production involves *competition*
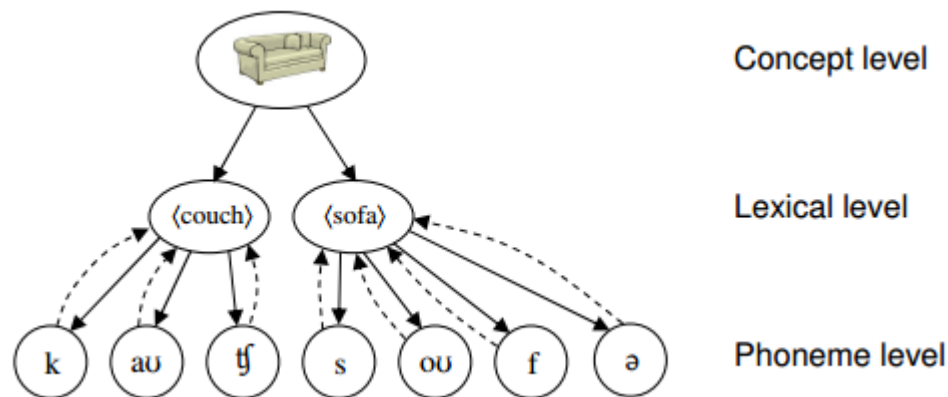
(p. 29)

- Competitors get an edge from...
  - phonological goodness (being connected to phonological structures with high resting activation)
  - phonological typicality (ditto)
  - frequency (high resting activation) ← self-reinforcing over the generations

- If speakers ("agents") communicate, given certain assumptions they'll tend to converge on just one term per meaning (*fish-y*) but not always (*man-ly, man-ful*):

  (12) **Set of cooperative agents ($\beta > \alpha$): agents come to agree on a single word**
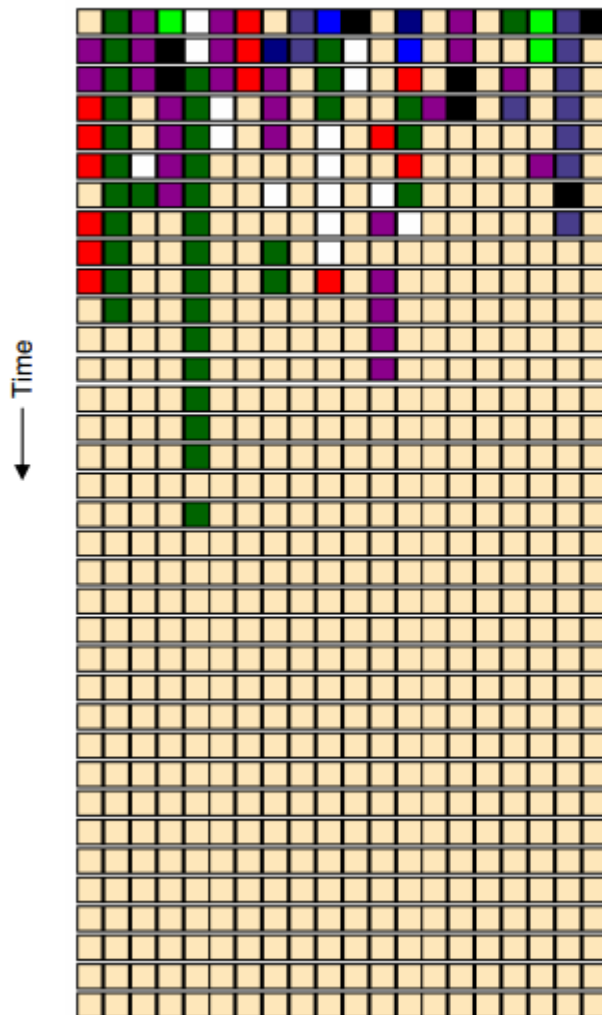
β and α are parameters in Martin's model that determine relative importance of one's own utterances to others'.

(p. 25)

- Phonological-goodness example: Latin, countertypologically, had more *d* than *b*.
  - The Romance languages fixed this, in part by preferentially retaining *b*-words (largely not the same ones, except in Spanish-vs-Portuguese!)
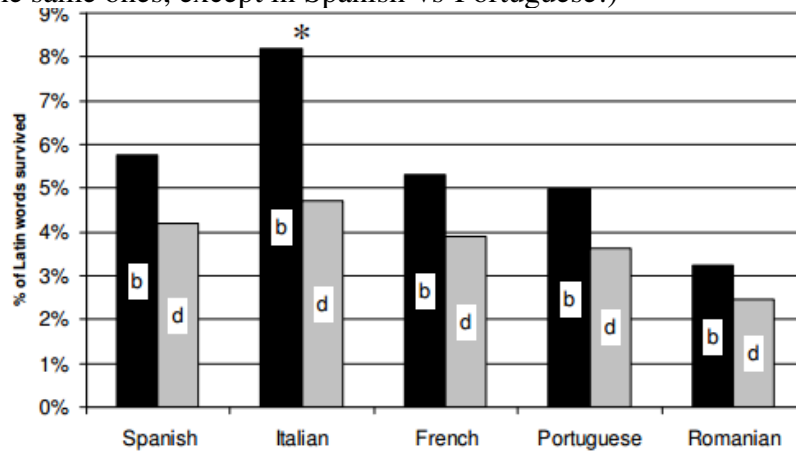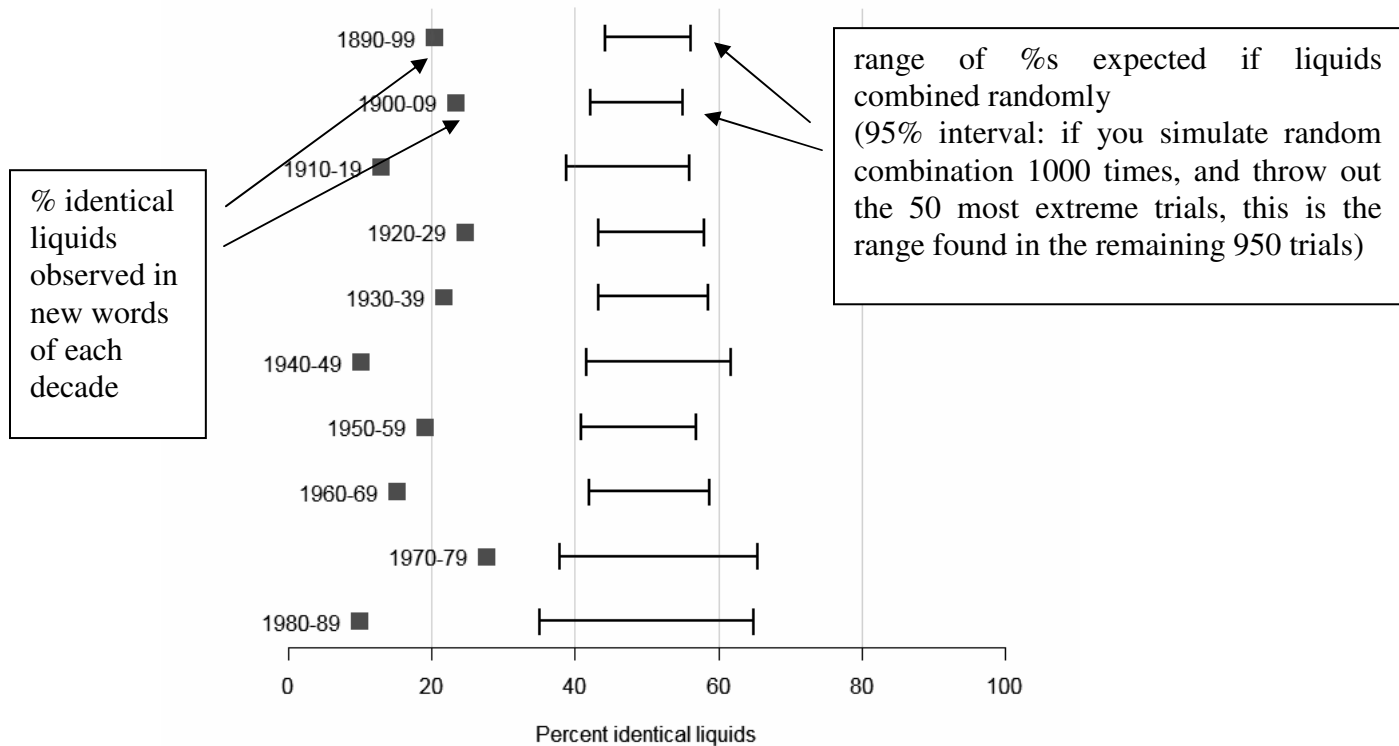

(p. 59)

  - But more than that, in French case study, *b*-words spawned far more derivatives:

|			/b/-initial words	/d/-initial words
(a)	inherited from Latin	**B ≈ D**	14	15
(b)	re-borrowed from Latin	**D > B**	31	77
(c)	borrowed from another language	**B > D**	64	31
(d)	derived from existing French word	**B > D**	212	77

(p. 62)

- Another good one: liquids in English.
  - Martin shows that...
    - In the English lexicon overall, if a word has two liquids, they're more likely to be *l...r* or *r...l* than *l...l* or *r...r*.
    - In Old English, about 35% of words with two liquids have identical liquids, compared to ~55% expected by chance.
    - In Middle English, it's about 25% (expect ~50%)
    - Today, it's about 25% (expect ~50%)
  - Even though current English retains only ~10-15% of the Old English vocabulary!
  - → English gained and lost many, many words, but always tended to respect the constraint against *l...l* or *r...r*.

▪ Martin also uses the Oxford English Dictionary, which gives dates of earliest attested use for each word, to look at words newly entering the language.
  ▪ In every decade, new words avoid identical liquids:

% identical liquids observed in new words of each decade

range of %s expected if liquids combined randomly
(95% interval: if you simulate random combination 1000 times, and throw out the 50 most extreme trials, this is the range found in the remaining 950 trials)

Percent identical liquids

(p. 78)

● Phonological-typicality example (maybe): increasingly, Americans agree on which sound their baby's name should begin with:

Entropy of English word-initial consonant distribution

Entropy

Decade

(p. 41)
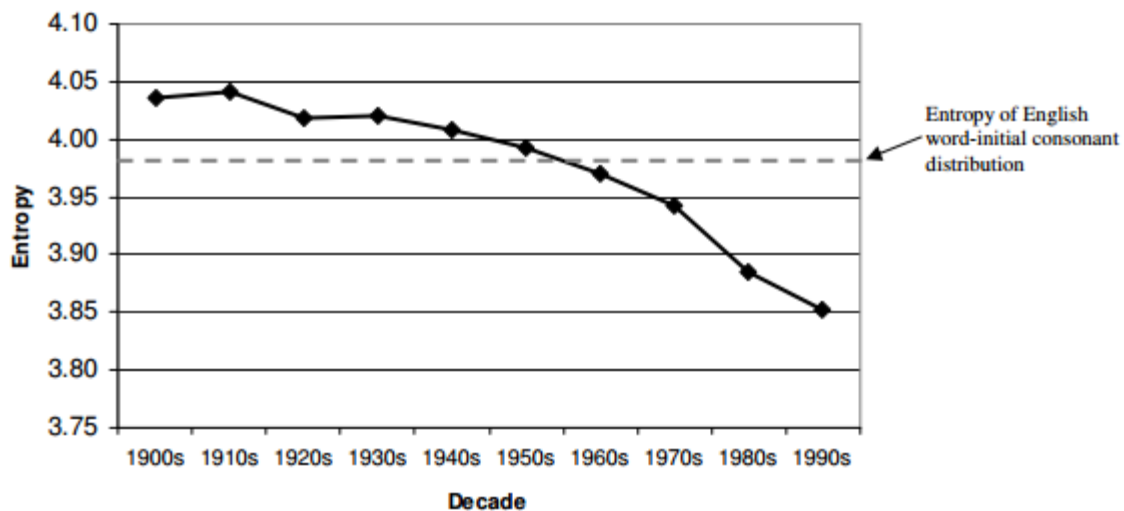
11

Other places where phonological goodness can play a role in a competition

10   Imperfect puns

- Zwicky & Zwicky 1986 (*with fronds like these, who needs anemones*), Fleischhacker 2006, Kawahara 2010
- E.g., in Japanese, puns don't have to be identical, but the more similar they are, the more likely they are to be attested

        Arumikan-no ue-ni aru mikan   'An orange on an aluminum can.'
        Aizu-san-no aisu   'Ice cream from Aizu' (Kawahara)
- Various interesting implications for what counts as similar, how similarity is calculated
- The competition is at the levels of deciding whether to make the pun, deciding whether to repeat a pun you've heard

11   Imperfect rhyme (e.g., Steriade 2003, Kawahara 2007)

- Although rhymes don't have to be perfect, closer is better

        Kettobase kettobase; kettobashita kashi de gettomanee
        'Kick it, kick it; with funky lyrics, get money'   (Kawahara)

12   Metrical properties of verse

- Hayes 2009 takes on, among other issues, the problem of the "missing remedy"—attempted line derivations that simply crash:
    - As Halle & Keyser pointed out, the following isn't a line of iambic pentameter, and it's not clear how to fix it to make it one:
        Ode to the West Wind by Percy Bysshe Shelley
- Hayes proposes that a line will simply fail to be usable if the outputs of the phonological grammar matches the output of the metrical grammar (which will attempt crazy repairs if forced to).

- In humorous verse you sometimes see what looks like the output of the metrical or rhyming grammar, despite its unacceptable faithfulness violations:

    Gast[õ] y a l'téléf**[õ]** qui s**[õ]**, et y a jamais pers**[õ]** qui y rép[õ] (Nino Ferrer)
        The bold ones should be [ɔn], but are unfaithful in order to rhyme

13   Naming

- Which names people choose for babies (as we saw), fantasy role-playing characters, and pharmaceuticals (Martin 2007)
- Which first-name/last-name combinations people choose (Shih 2012)

14   Blends

- There's isn't a machine you can just plug any two words into.
    - *breakfast + lunch → lunch* is fine
    - but *breakfast + reception → ??breception ??breakception ??recepfast* crashes, I think

- Ahn 2011 example from Korean
    - *kunde* 'army' + *lotteria* 'Lotteria [fast-food chain]' → *kund**eria*** 'army mess hall'
    - As Ahn points out, Lotteria seems to have been chosen (rather than MacDonald's or whatever) because it produces a good blend.

- A French example

	France	*Quebec*
tweet	le tweet	le gazouillis le tweet
to tweet	teeter	gazouiller tweeter
tweeter, twitterer	tweeteur(e)	gazouilleur/se tweeteur(e)
twitterature	twittérature (twitter + littérature)	***gazouillature** twittérature

15   Next time

- A model that might be easier to work with: competition with the null parse

References

Ahn, Suzy. 2011. Master's thesis. Seoul National University master's thesis.

Albright, Adam, Bruce Hayes & Argelia Andrade. 2001. Segmental Environments of Spanish Diphthongization. *UCLA Working Papers in Linguistics 7 (Papers in Phonology 5)*. 117–151.

Baronian, Luc. 2009. Une analyse de verbes d´efectifs sans sp´ecification lexicale. In Luc Baronian & F Martineau (eds.), *Le franc¸ais d'un continent `a l'autre*, 29–48. Qu´ebec: Presses de l'Universit´e Laval.

Coetzee, Andries & Joe Pater. 2005. *Gradient phonotactics in Muna and Optimality Theory*. University of Michigan and University of Massachusetts.

Coetzee, Andries W & Joe Pater. 2007. Weighted constraints and gradient phonotactics in Muna and Arabic.

Daland, Robert, Bruce Hayes, James White, Marc Garellek, Andrea Davis & Ingrid Norrmann. 2011. Explaining sonority projection effects. *Phonology* 28(02). 197–234. doi:10.1017/S0952675711000145.

Daland, Robert, Andrea D Sims & Janet Pierrehumbert. 2007. Much ado about nothing: a social network model of Russian paradigmatic gaps. *Proceedings of the 45th annual meeting of the Association of Computational Linguistics*, 936–943. Prague: Association for Computational Linguistics.

Fleischhacker, Heidi. 2006. Similarity in phonology: evidence from reduplication and loan adaptation. UCLA Ph.D. dissertation.

Frisch, Stefan A, Janet B Pierrehumbert & Michael B Broe. 2004. Similarity Avoidance and the OCP. *Natural Language & Linguistic Theory* 22(1). 179–228.

Hayes, Bruce. 2009. Faithfulness and componentiality in metrics. In Sharon Inkelas & Kristin Hanson (eds.), *The nature of the word*, 113–148. Cambridge, MA: MIT Press.

Kawahara, Shigeto. 2007. Half rhymes in Japanese rap lyrics and knowledge of similarity. *Journal of East Asian Linguistics* 16(2). 113–144. doi:10.1007/s10831-007-9009-1 (14 February, 2012).

Kawahara, Shigeto. 2010. *Papers on Japanese imperfect puns*.

Martin, Andrew. 2004. The effects of distance on lexical bias: sibilant harmony in Navajo compounds. UCLA master's thesis.

Martin, Andrew. 2007. The evolving lexicon. University of California, Los Angeles Ph.D. Dissertation.

Martin, Andrew. 2011. Grammars leak: modeling how phonotactic generalizations interact within the grammar. *Language* 87(4). 751–770.

Orgun, Cemil Orhan & Ronald L Sprouse. 1999. From "MParse" to "Control": Deriving Ungrammaticality. *Phonology* 16(2). 191–224.

Raffelsiefen, Renate. 1996. Gaps in word formation. In Ursula Kleinhenz (ed.), *Interfaces in phonology*, 194–209. Berlin: Akademie Verlag.

Raffelsiefen, Renate. 1999. Phonological constraints on English word formation. In Geert E Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1998*, 225–287. (Yearbook of Morphology 8). Springer.

Raffelsiefen, Renate. 2004. Absolute ill-formedness and other morphophonological effects. *Phonology* 21(1). 91–142.

Shih, Stephanie. 2012. Linguistic determinants in English personal name choice. Presentation. Paper presented at the LSA annual meeting, Portland, OR.

Steriade, Donca. 2003. Knowledge of perceptual similarity and its uses: evidence from half-rhymes. In M.J. Solé, D Recasens & J Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, 363–366. Barcelona: Futurgraphic.

Zwicky, Arnold M & Elizabeth D Zwicky. 1986. Imperfect puns, markedness, and phonological similarity: with fronds like these, who needs anemones? *Folia Linguistica* 20(3-4). 493–544.