Bruce Hayes
UCLA

Chicago Linguistic Society
March 10, 2014

Comparative Phonotactics[1]

[1] Thanks to audiences at UCLA and the University of Pisa for comments on earlier versions of this talk.

TWO KINDS OF PHONOTACTICS: ABSOLUTE AND COMPARATIVE

1. Absolute phonotactics

- What is the phonological well-formedness of a particular word?
- How is it learned?
- Hayes and Wilson (2008) suggested both a grammar framework and a learning system.
  - Framework: the maxent variant (Goldwater and Johnson 2003) of harmonic grammar (Legendre et al. 1990), with the overall constraint-based architecture borrowed from Optimality Theory (Prince and Smolensky 1993)

- The system assigns a predicted probability to every word; difference in probability correspond to differences in well-formedness.
- An algorithm selects phonological constraints and weights them so as to maximize the predicted probability of the set of existing words.
- To some extent, the output probabilities assigned by this system succeed in matching linguists' phonotactic descriptions and human phonotactic intuitions.

2. Comparative phonotactics

- Assume two populations of strings, A and B.
- Assume the same maxent framework (constraints, weights, etc.)
- Seek a grammar whose output probabilities accurately predict whether any given string will belong to population A or population B.
- To do this, the constraints must be comparative — make distinctions between the A and B populations.
- Is comparative phonotactics a useful idea for phonology or phonological learnability?

3. Plan of talk

- An example: a comparative phonotactics for the English Latinate/Native distinction
- An example: environment discovery by stem-sorting: comparative phonotactics in Hungarian vowel harmony
- Address general questions about the approach.

APPLICATION I:
The Latinate/Native distinction of English

4. The Lexical strata hypothesis

- Chomsky and Halle (1968, 373)2 proposed that languages with heavy admixtures of loanwords develop synchronically arbitrary lexical strata — groupings of vocabulary that:
  - ➢ have a purely diachronic origin (native vs. adapted foreign words)
  - ➢ are nevertheless apparent to native speakers as a synchronic phenomenon
- In English the strata are thought to be Native and Learned/Latinate, perhaps with a Greek subdivision of the latter.

---

[2] A tiny sampling of other work: McCawley (1968), Ito and Mester (1995), Moreton and Amano (1999)

5. I think strata are real

- As a native speaker I feel I have a strong sense of the "Latinity" of English words, even though I know no Latin.
- This sense is gradient:
  - very Latinate: *protectionism*, *veterinarian*, *sexuality*, *vaporization*, *industrialization*
  - Not Latinate at all: *warmth*, *fresh*, *swath*, *shove*, *pooch*, *yank*, *beige*, *snot*
  - Fairly Latinate: *palate*, *oblique*, *motor*, *postal*, *suitor*
  - See analysis below, which predicts these distinctions.

6. What could constitute the language learner's evidence for strata?

- **Morpheme cooccurrence**: if you have -*ation*, then you likely have *con*-. (49/613, in my data)
- **Alternations**: Latinate words undergo different phonological alternation types, such as Trisyllabic Shortening (*SPE*)
- **Phonotactics**: Latinate and native words are phonotactically different.
  - ➢ This is just what Ito and Mester (1995) proposed re. the strata of Japanese.

7. Where does the native speaker's sense of strata come from? A proposal

- They internalize a comparative phonotactics
  - Population A = Native
  - Population B = Latinate
- The contrasting strata are bootstrapped in some way, building up from initially simple information, making use of morphology.

8. Getting started: an operational definition of Latinity

- Let's suppose that any word of at least seven letters ending in one of these suffixes is Latinate:

  -able, -acy, -al, -ance, -ancy, -ant, -ary, -ate, -ated, -ation, -ator, -atory, -ence, -ency, -ent, -graphy, -ia, -iac, -ian, -ible, -ic, -ical, -ician, -ific, -ify, -ine, -ism, -ist, -ity, -ium, -ive, -ize, -ular, -logy, -or, -ory, -ous, -sis, -tion, -ure, -us

9. Is this acceptable as a heuristic criterion?

- I checked an electronic lexicon (my own edited version of the Carnegie-Mellon Pronouncing dictionary; www/linguistics.ucla.edu/people/hayes/EnglishPhonologySearch)
- At least intuitively, this criterion seems not too bad to me as an ad hoc way of identifying words that seem Latinate.
- We'll be in a better position to check, shortly.

10. Finding constraints for the grammar

- In a fully-principled approach these would be located by algorithm, as in Hayes and Wilson (2008).
- In this exploration, I checked constraints that seemed likely to me to work (i.e., I carried out traditional problem-set analysis).

11. Candidates in the tableaux

- They are statuses: [Latinate], [Native]

12. Sample constraint

- PREFER [Native] IF [$_{word}$ [+strident][+nasal]
- "Assess a penalty to [Latinate] status for any word that begins with a sibilant followed by a nasal."

CONSTRAINTS I:

Those that penalize Latinity

13. Latin had stricter phonotactics than English

- These are lacking:
  - Initial sibilant-nasal clusters (due to a sound change: proto **snurus* became *nurus* 'daughter-in-law').
  - No [f] before obstruents ([ft])
  - Palato-alveolars /ʃ, ʒ, tʒ, dʒ/. These arose later in English by alveolar palatalization, but only in "ambisyllabic" positions (*nation*, *vision*, *natural*, *gradual*).

14. English sounds without Latin sources

- Various English sounds just happen not to be the way that Latin sounds normally get rendered; e.g. [ʊ], [aʊ].

15. The Latin sounds were transmitted to English in particular ways

- [w] is rendered as such only in the clusters [kw] and [gw]; else it appears as [v]; so [w] is missing in other positions.
- [k, g] undergo Velar Softening to [s, dʒ] before (what used to be) nonlow front vowels ([aɪ, ɪ, i, ɛ]).
- **u* is [ʌ] before nonfinal coda consonants (*ungulent*), else [u] after coronals (*duplicate*), else [ju] (*circuitous*).

CONSTRAINTS II:

Those that penalize nativeness

16. Just plain length

- Latinate words are longer; in our culture we say “long words” when we difficult, rare, learned words.

17. Some sound sequences are abundant in Latinate words and not in native words

- [VpʃV], [VkʃV]
- stressless [iə]
- [mn]

18. Certain phonemes are overrepresented in Latinate words

- [n], [t], [v]

19. The full grammar I set up: constraints

Prefer Native	**Prefer Latinate**
INITIAL [sn]	STRESSLESSVOWEL
MONOSYLLABIC	[mn]
NONAMBISYLLABIC	[iə]
PALATOALVEOLAR	
INITIAL [ʃ]	AT LEAST 5 SYLS
ALVEOLARSTOP [l]	At LEAST 4 SYLS
[ft]	[ərə]
DISYLLABIC	{[p], [k]}+[ʃ]
w NOT AFTER [k], [g]	Individual segments: [n], [v], [t], [ʃ]
*V:CC	(four constraints)
FINAL MAIN STRESS	
INITIAL [j] NOT BEFORE [u]	
[k,g] + VELAR SOFTENING TRIGGER	

[ʌ] IN OPEN SYLLABLE
TAKER OF [ju] BEFORE [u]
GENERAL BIAS AGAINST
LATINITY (intercept)
TRISYLLABIC SHORTENING
Individual segments: [θ], [ŋ],
[ʊ], [aʊ]
(four constraints)

20. Finding the weights that best fit the data:

- I used R to find them; R script on request

21. Best-fit weights for my constraints

Prefer Native		**Prefer Latinate**	
INITIAL [sn]	11.84	STRESSLESSVOWEL	0.12
MONOSYLLABIC	6.47	[n]	0.34
NONAMBISYLABIC			
PALATOALVEOLAR	4.08	[v]	0.64
INITIAL [ʃ]	3.91	[t]	0.84
ALVEOLARSTOP [l]	2.60	[mn]	1.15
[ft]	1.77	[iə]	1.17
DISYLLABIC	1.53	AT LEAST 5 SYLS	1.27
w NOT AFTER [k], [g]	1.39	At LEAST 4 SYLS	1.33
*V:CC	1.35	[ʃ]	1.61
FINAL MAIN STRESS	1.24	[ərə]	1.61
INITIAL [j] NOT BEFORE [u]	1.23	{[p], [k]}+[ʃ]	1.91
[ʊ]	1.22		
[k,g] + VELAR	1.10		

SOFTENING TRIGGER	
[aʊ]	1.02
[ʌ] IN OPEN SYLLABLE	1.00
TAKER OF [ju] BEFORE [u]	0.71
GENERAL BIAS AGAINST LATINITY (intercept)	0.58
[ŋ]	0.49
[θ]	0.37
TRISYLLABIC SHORTENING	0.08

22. Computing probability of Latinness for one form: *frustration* [ˌfrʌsˈtɹeɪʃən]

- Frustration violates four simple constraints penalizing non-Latinity:

	Weight
PREFER LATINATE IF [n]	0.341
PREFER LATINATE IF [t]	0.843
PREFER LATINATE IF [ʃ][3]	1.610
PREFER LATINATE IF STRESSLESS VOWEL	0.119
Total weight ("harmony")	2.904

[3] [ʃ] per se is actually favored in Latinate words; the preference is overridden by stronger anti-Latinate constraints on [ʃ] that are applicable when it is not in its preferred ambisyllabic position.

- *Frustration* violates one constraint penalizing Latinity, the default constraint:

GENERAL PREFERENCE AGAINST LATINITY	0.578
Total weight	0.578

- The standard maxent formula (e.g. Goldwater and Johnson 2003, (1)) tells us:
- P(*frustration* is Latinate) = $\frac{e^{-0.578}}{e^{-0.578} + e^{-2.904}} = 0.911$
- So *frustration* is claimed to be fairly Latinate, but not utterly Latinate.

PERFORMANCE OF THE LATINITY-DETECTING GRAMMAR

23. Words predicted to be "very Latinate"

- Highest scoring words that I had pre-classified as Latinate (see (8)), all with probabilities at least .996:

  protectionism, veterinarian, sexuality, vaporization, geriatrician, industrialization, perfectionism, reactionary, generalization

24. Words predicted to be "very Native"

- Lowest-scoring words pre-classified as non-Latinate
  - Sampling at random from the bottom 500, all with scores less than .001:

    warmth, fresh, gulch, swath, preach, shove, pooch, yank, beige, snot

25. The intermediate forms mentioned earlier

- *palate*, *oblique*, *motor*, *postal*, *suitor* all about .23

26. Looking at errors I: Lowest-scoring words that I had pre-classified as Latinate

- These appear almost entirely to be misclassifications like *sardine*.
- A few are interestingly deviant words with true Latinate suffixes (over)

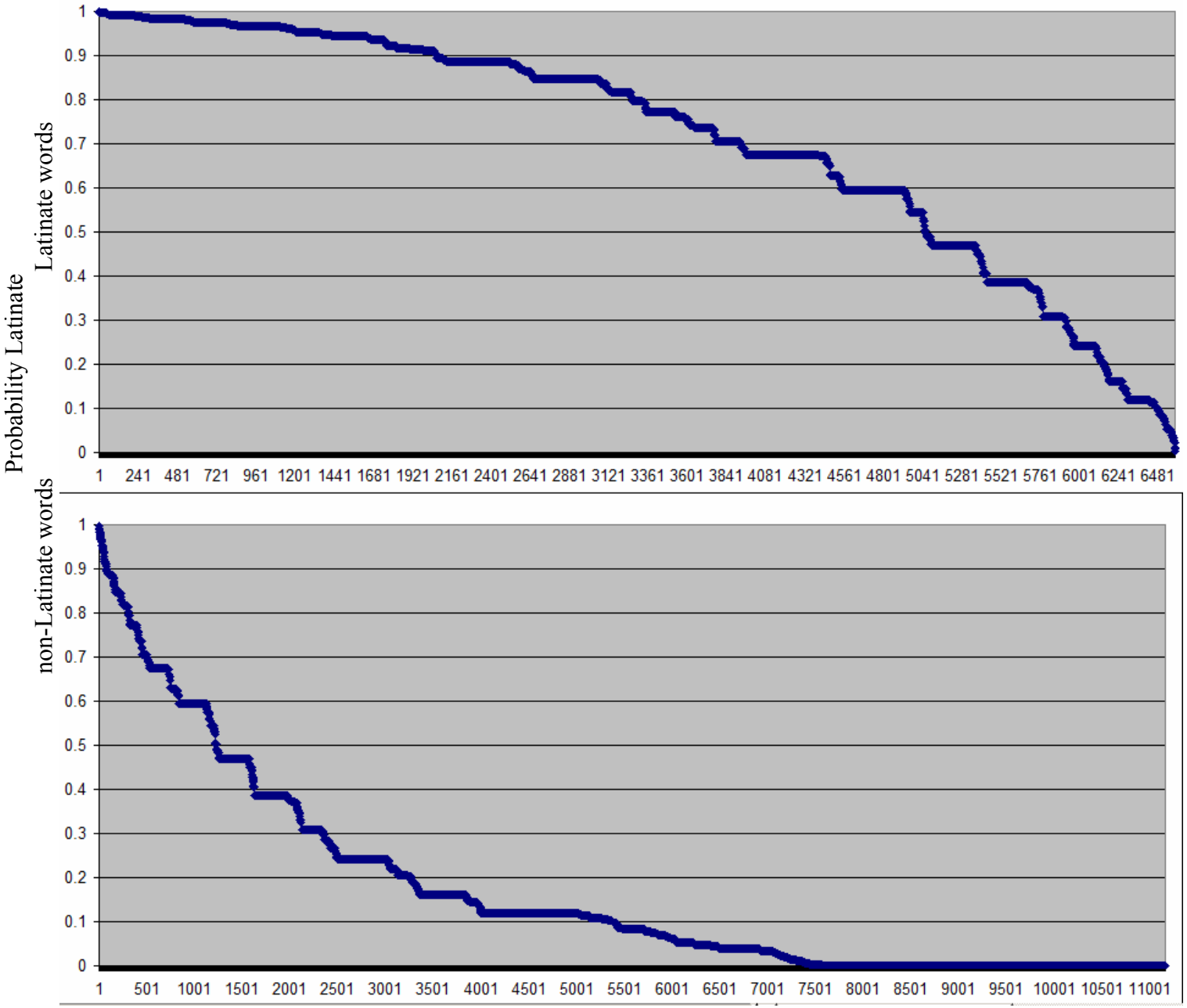
public	0.048	[ʌ] in open syllable
wondrous	0.045	unusual attachment of Latinate suffix to native stem
warrior	0.033	
vegetable	0.045	palato-alveolar in coda, due to syncope [ˈvɛ**dʒ**.tə.bəl]
psychic	0.044	Velar Softening not applied, because Greek (*SPE* suggests a separate sub-stratum for Greek)
seismic	0.034	long V in closed syllable, because Greek

27. Looking at errors II: Highest-scoring words that I had pre-classified as non-Latinate

- There are few, virtually all are words with Latinate suffixes that didn't make it into my suffix list.

28. Aggregate performance

- For these charts, I separated Latinate and non-Latinate (by my preclassification), then sorted by descending predicted probability.


Probability Latinate
Latinate words
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 241 481 721 961 1201 1441 1681 1921 2161 2401 2641 2881 3121 3361 3601 3841 4081 4321 4561 4801 5041 5281 5521 5761 6001 6241 6481
non-Latinate words
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 501 1001 1501 2001 2501 3001 3501 4001 4501 5001 5501 6001 6501 7001 7501 8001 8501 9001 9501 10001 10501 11001

29. What we need for more serious validation

- Some kind of wug test: "Please rate on a scale of 1-7 whether this word would be likely to be used in scientific or scholarly writing."
  - Wug 1: *tennecation* (model prediction: .978)
  - Wug 2: *wepechation* (model prediction: .016)

APPLICATION II:

Finding the environments for phonological processes by sorting the stem inventory

30. Learning environments by stem-sorting

- I follow Becker and Gouskova (2012), who suggest this for Russian data.
- We have some affix that exists in two allomorphic forms **a** and **b**.
- We suppose that the stems that take these allomorphs form populations **A** ("a-takers") and **B** ("b-takers")
- Proposal: language learners sometimes perform comparative phonotactics on the two populations and use the result to distribute the affix allomorphs.

31. Comparison: how this is analyzed as "pure phonology" in OT

- We assume some appropriate underlying forms, one per morpheme.
- The GEN component creates the possible suffix allomorphs, like Hungarian dative [-nak]/[-nɛk]
- Assume some appropriate set of constraints (EVAL), perhaps from Universal Grammar (Prince and Smolensky 1993).
- GEN and EVAL don't even look at stems or suffixes; you need only find the right constraint ranking and the system will give you the right answer.

32. Can we argue for stem-sorting? A Hungarian example

- Hayes, Zuraw, Siptár, and Londe (2009) studied Hungarian vowel harmony.
- They used stem-sorting as a research heuristic.
- Two populations of stems:
  - A: those that take front-vowel suffixes
  - B: those that take back-voweled suffixes
- What we didn't realize: stem-sorting is actually essential to part of their analysis.

33. The most effective way to separate the populations: vowel harmony constraints

- E.g. stems whose rightmost vowel is front rounded are always in Population A.
- Stems whose rightmost vowel is back are always in Population B

34. Simple cases where this works perfectly

- Constraints like these are never violated in winning candidates and get huge weights in a maxent grammar:

  AGREE(back) trigger: [−back,+round]
  AGREE(back) trigger: [+back]

Input	**Candidate**	**Harmony**	**Predicted freq.**	**AGREE (back)-front rounded trigger**	**AGREE (back)-back trigger**
				50	50
yʃt-nAk	yst-nak	50	0	*	
	☞ yst-nɛk	0	1		
ablak-nAk	☞ ablak-nak	0	1		
	ablak-nɛk	50	0		*

35. The harder cases of Hungarian vowel harmony

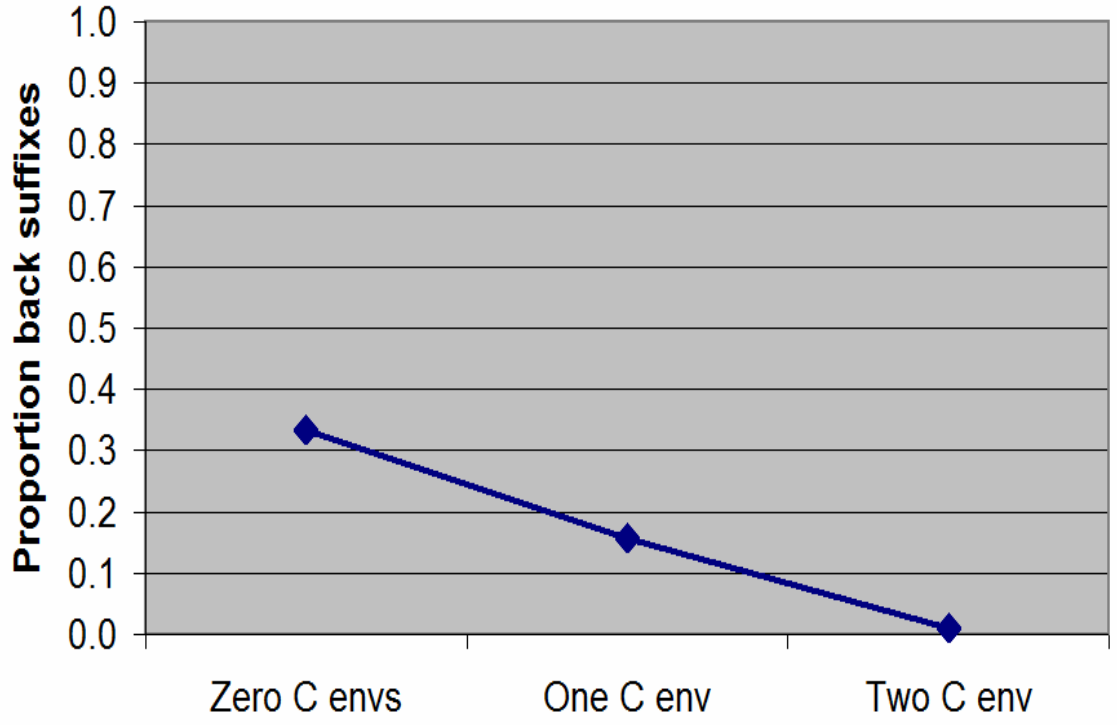
- In the "zones of lexical variation" (about 900 stems), harmony is unpredictable.
  - These are stems whose rightmost two vowels are Back + Neutral, or rightmost three vowels are Back + Neutral + Neutral
  - "Neutral" = [i, iː, eː, ɛ]
  - Behavior of each stem must be memorized, though there are strong statistical patterns present.

36. Consonant effects in Hungarian vowel harmony

- In the "zones", there is a surprise: stem-final consonants statistically affect harmony (Hayes, Zuraw, Siptár, and Londe 2009).
- Front suffixes occur more often when the stem ends in:
  - a bilabial consonant
  - a sibilant
  - a coronal sonorant
  - a consonant cluster

- The effect is surprisingly large: about 1/3 back suffixes when none of these environments is met; close to zero when two are present.

Proportion of back suffixes according to number of consonant environments present



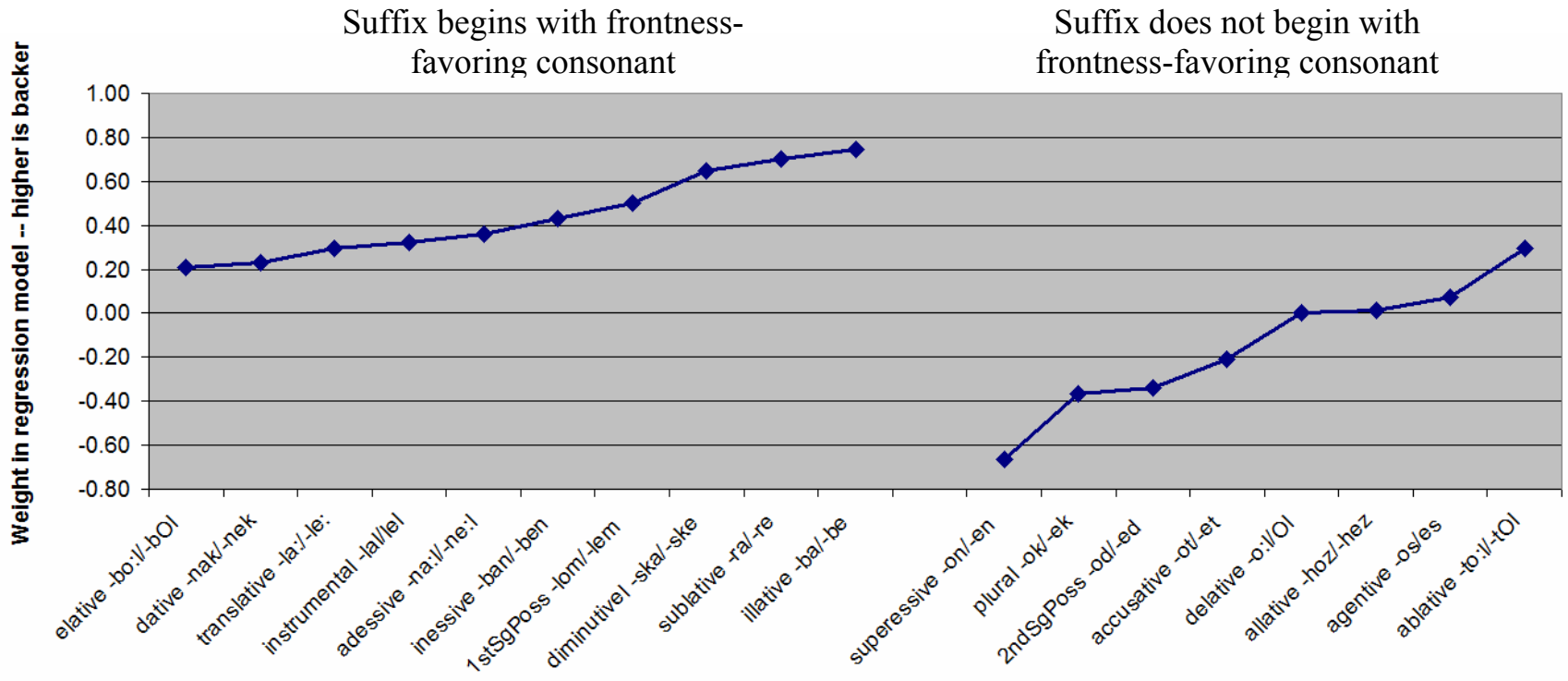
- To our surprise, the effect is productive, as shown by highly significant results in our wug test.

37. Stem-sorting, or ordinary whole-word phonology?

- Vowel constraints like those given above work fine as normal phonology — the suffix allomorph that better AGREE's with the stem vowel will surface as the winner.
- But for consonants, things are different — the truth seems to be that speakers use stem-sorting.

38. Evidence for stem-sorting I: suffix behavior

- About half of the Hungarian suffixes begin with a consonant in one of the four classes given above, like dative [-nɛk]/[-nɔk], with a coronal sonorant.
- But these suffixes do not take front allomorphs more often than the others; if anything, it is the reverse.


Suffix begins with frontness-favoring consonant
Suffix does not begin with frontness-favoring consonant
Weight in regression model -- higher is backer
1.00
0.80
0.60
0.40
0.20
0.00
-0.20
-0.40
-0.60
-0.80
elative -bo:l/-bOl
dative -nak/-nek
translative -la:/-le:
instrumental -lal/lel
adessive -na:l/-ne:l
inessive -ban/-ben
1stSgPoss -lom/-lem
diminutivel -ska/-ske
sublative -ra/-re
illative -ba/-be
superessive -on/-en
plural -ok/-ek
2ndSgPoss -od/-ed
accusative -ot/-et
delative -o:l/Ol
allative -hoz/-hez
agentive -os/es
ablative -to:l/-tOl

39. Evidence for stem-sorting II: stem behavior

- The consonant effects on vowel backness fail to show up when you inspect the stem inventory — they are simply not part of Hungarian gradient phonotactics.[4]

	After consonants that favor frontness in suffixes	After consonants that do not favor frontness in suffixes
% of front vowels in stems	42.4	44.7

[4] Thanks to Kie Zuraw, who kindly prepared a spreadsheet proving this point when the question arose.

40. So what's in Hungarian grammar?

- The relevant constraints, however stated, must embody generalizations like:
  - Use front vowels after stems that end in bilabial consonants.
  - NOT: Use front vowels after bilabial consonants.
- And: constraint of the correct type are those learnable by stem-sorting.
- For the main effects (like back vowels, front rounded vowels), perhaps stem-sorting is overkill: straightforward OT Markedness-Faithfulness ranking could easily do the job.
- Thus, stem-sorting is perhaps a "desperation measure", used where simple predictability is unattainable.

THREE GENERAL QUESTIONS ABOUT COMPARATIVE PHONOTACTICS

QUESTION I:

Does comparative phonotactics solve some problems better than "absolute" phonotactics would?

41. The argument from "wrecked forms"

- Suppose Latinity = "gets good score on a grammar of absolute phonotactics for Latinate words"
- Consider then our *tennecation* [ˌtɛnəˈkeɪʃən] vs. *wepechation* [ˌwɛpəˈkeɪʃən].
- Modify them to *[ˌtɛnəˈkeɪʃənp] vs. *[ˌwɛpəˈtʃeɪʃənp].
- These would get horrible scores from any decent phonotactic grammar.
- My intuition is that [ˌtɛnəˈkeɪʃənp] is nonetheless more Latinate than [ˌwɛpəˈkeɪʃənp].
- I.e. direct comparison would work better than comparison to an absolute standard.
- The badness of these forms should be attributed to the absolute phonotactics of English as a whole.

42. Becker and Gouskova's idea (2012): comparative phonotactics is compared absolute phonotactics

- Learn the absolute phonotactics of Population A
- Learn the absolute phonotactics of Population B
- Then, probability that a form x belongs to A is

$$\frac{x\text{'s phonotactic probability construed as A}}{x\text{'s phonotactic probability as A} + x\text{'s phonotactic probability as B}}$$

- This strikes me as intriguing but oblique — why not solve the problem as directly as possible? The non-contrastive information will probably just be noise.
- My own efforts at applying the BG method to Latinity yields less accurate results, as measured by summed log probability. (BG: −7594.9, comparative −7208.8)

QUESTION II:

Why would it be sensible for language learners to engage in comparative phonotactics?

43. Basic answer

- Because it makes them **better speakers**, in many different ways.
- Here are examples.

44. Part of speech

- Comparative phonotactics permits guessing of part of speech (cf. *SPE*, Smith 2011), perhaps useful to the acquisition of syntax (work by Morten Christiansen and colleagues)

45. Grammatical gender

- The existing research literature indicates that grammatical gender is surprisingly predictable on phonological (also morphological) grounds.
  - For models that do this pretty well for French, see Karmiloff-Smith 1979, Lyster 2006, Glewwe 2014
- A comparative phonotactic grammar for gender permits speakers to make better guesses about grammatical gender for new words, or better understand other people's mistaken or dialectally-varying productions.

46. Style

- A comparative phonotactics for strata helps a speaker command different styles
- Learned style:
  - Even a child or an ill-educated person is aware when a speaker is larding their discourse with Latinate terms

47. Speech perception

- In a widely-adopted view, human speech perception is guided by a Bayesian "forward model" that assigns prior probabilities to the possible interpretations of the signal.
  - ➢ See "Shortlist B" (Norris and McQueen 2008), an explicit model that implements this idea.
- An remarkable phonological experiment, Moreton and Amano (1999), shows that Japanese listeners use knowledge of lexical strata when they perceive vowel length.

48. The Moreton/Amano experiment

- Basic phenomena:
  - Initial [rj] and [hj] do not occur in the Native stratum.
  - Long [aː] does not occur in the Sino-Japanese stratum.
- Paradigm:
  - play nonce words like [rjotaː], [potaː] ([p] not confined to Sino-Japanese)
  - smoothly vary the length of the [aː]
  - calculate the perceptual boundary between [a] and [aː]
- Result: when initial consonants are [rj], you need more phonetic length to perceive phonological [aː]!

QUESTION III:

What sort of grammatical architecture could accommodate comparative phonotactics?

49. Monolithic vs. atomistic approaches to phonological grammar

- Example of monolithic: Smolensky (1995), defending the use of a single constraint hierarchy for both production and perception.
- Atomistic approaches are varied:
  - Edward Flemming, Paul Boersma each propose systems with separate grammars for production and perception.
  - Sharon Inkelas and colleagues propose "co-phonologies" for separate grammatical constructions.
- Comparative phonotactics is clearly more compatible with an atomistic approach.

50. The virtuosic child

- Experimental phonology seems to be flourishing — what is it telling us?
- I think it is revealing human children as highly virtuosic extractors of phonological patterns from the data they receive.

- Two examples:
  - Albright and Hayes's (2003) "islands of reliability" for English past tenses: children detect e.g. that all verbs ending in voiceless fricatives are regular.
  - Ernestus and Baayen's demonstration of "predicting the unpredictable" in Dutch: speakers know the probability of voiceless and voiced consonants when you ask them to "undo Final Devoicing" on novel stems.
- There are isolated and puzzling exceptions (notably Becker, Nevins, and Ketrez 2011), but I think the pattern that is emerging is:
  - When in doubt, *bet on the language learner to notice things*.

51. The general scheme

- Language acquirers are virtuosic and assiduous.
- They behave atomistically, noticing individual problems involving predictability, and solve them with appropriate individualized grammatical subsystems.
- Such detailed knowledge is useful to them in production, perception, detection of style, synthesis of novel forms, and gender-guessing.
- The particular kind of phonotactics (absolute or comparative) that they learn may be whatever is best adapted to the particular problem they are trying to solve

References

Anttila, Arto. 1997. Deriving variation from grammar. In Frans Hinskens, Roeland van Hout and Leo Wetzels (eds.), *Variation, change and phonological theory*, Amsterdam, John Benjamins. pp. 35-68

Baković, Eric (2011) Opacity and ordering. In John Goldsmith, Jason Riggle, and Alan C. L. Yu (eds.) *The handbook of phonological theory*, 2nd ed. Oxford: Wiley-Blackwell.

Becker, Michael, Nihan Ketrez, and Andrew Nevins (2011) The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language* 87:1, pp. 84–125.

Becker, Michael and Maria Gouskova. 2012. Source-oriented generalizations as grammar inference in Russian vowel deletion. Ms., SUNY Stony Brook and NYU.

Bybee, Joan and Carol Lynn Moder. 1983. Morphological classes as natural categories. *Language* 59:251-270.

Cedergren, Henrietta and David Sankoff. 1974. Variable rules: Performance as a statistical reflection of competence. Language, 50: 333-355.

Chomsky, Noam and Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.

Christiansen, Morten, & Monaghan, Padraic. 2006. Discovering verbs through multiple cue integration. In K.Hirsh-Pasek & R. M.Golinkoff (Eds.), Action meets word: How children learn verbs (pp. 88–110). New York: Oxford University Press.

Gelman, Andrew, Alek Jakulin, Maria Grazia Pittau, and Yu-Sung Su. 2008. A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics* 2.1360–83.

Gelman, Andrew, Yu-Sung Su, Masanao Yajima, Jennifer Hill, Maria Grazia Pittau, Jouni Kerman, and Tian Zheng. 2009. arm: data analysis using regression and multilevel/hierarchical models. R package. http://cran.r-project.org/web/packages/arm

Glewwe, Eleanor. 2014. Developing a Phonological/Morphological Model to Predict the Gender of French Nouns. Ms., Dept. of Linguistics, UCLA.

Goldwater, Sharon, and Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, ed. by Jennifer Spenader; Anders Eriksson, and Osten Dahl, 111–120. Stockholm: Stockholm University Department of Linguistics.

Hayes, Bruce and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39: 379-440.

Hayes, Bruce, Kie Zuraw, Péter Siptár, and Zsuzsa Londe. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85: 822-863.

Hulstaert, Gustaaf. 1961. *Grammaire du lo̱mo̱ngo̱*. Tervuren: Musée royal de l'Afrique centrale.

Itô, Junko and Armin Mester. 1995. Japanese phonology. *The Handbook of Phonological Theory*, ed. J. Goldsmith, 817-838. Oxford: Blackwell.

Kapatsinski, V. 2013a. Conspiring to mean: Experimental and computational evidence for a usage-based harmonic approach to morphophonology. Language 89: 110-148.

Karmiloff-Smith, A. 1979. Production experiments: gender-indicating function of determiners. In A Functional Approach to Child Language, 148-169. Cambridge: Cambridge University Press.

Kawahara, Shigeto, Kohei Nishimura, and Hajime Ono. 2005. Unveiling the unmarkedness of Sino-Japanese. In William McClure, ed., *Japanese/Korean Linguistics* 12. Stanford: CSLI.

Kenstowicz, Michael and Charles Kisseberth. 1979. *Generative phonology: Description and theory.* San Diego: Academic Press.

Kiparsky, Paul. 1973. Abstractness, opacity and global rules. In O. Fujimura (ed.) *Three dimensions of linguistic theory*. Tokyo: Taikusha. l–136.

Kiparsky, Paul. 1982. Lexical phonology and morphology. In I. S. Yang (ed.), *Linguistics in the Morning Calm.* Seoul: Hanshin. 3-91.

Legendre, Géraldine, Yoshiro Miyata and Paul Smolensky. 1990. Harmonic Grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. Report CU-CS-465-90. Computer Science Department, University of Colorado at Boulder.

Lyster, R. (2006). Predictability in French gender attribution: A corpus analysis. *Journal of French Language Studies*, 16(1), 69-92.

Lombardi, Linda. 1999. Positional faithfulness and voicing assimilation in Optimality Theory. *Natural Language and Linguistic Theory* 17: 267-302.

McCawley, James. 1968. *The phonological component of a grammar of Japanese*. The Hague: Mouton.

Moreton, Elliott, and Shigeaki Amano. 1999. Phonotactics in the perception of Japanese vowel length: Evidence for long-distance dependencies. Proceedings of the 6th European Conference on Speech Communication and Technology, Budapest.

Norris, Dennis and James M. McQueen. 2008. Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review* 115:357-395.

Prince, Alan and Paul Smolensky. 1993/2004. *Optimality theory: Constraint interaction in generative grammar*. Technical report, Rutgers University Center for Cognitive Science. [Published 2004: Blackwell]

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Smith, Jennifer. 2011. Category-specific effects. In Marc van Oostendorp, Colin Ewen, Beth Hume, and Keren Rice (eds.), *The Blackwell Companion to Phonology*, 2439-2463. Malden, MA: Wiley-Blackwell.

Tesar, Bruce and Paul Smolensky. 2000. *Learnability in Optimality Theory*. MIT Press.