Linguistics 251 | Spring 2013
Variation in Phonology | Hayes/Zuraw

Class 12, 5/9/13: More on Gaps and Frequency

1. Assignments etc.

- Hand in your Seuss exercise.
  - Note: this will be the basis as well of our last exercise, on model comparison.
- Read: nothing; the last exercise is substantial so we're giving you time off…
- Talk: with us re. projects.

THE -ABLE QUESTIONNAIRES

2. Gradient gappiness (per Kie)

- This happens so often in a wug test: participants do express a preference, but express reluctance to endorse any outcome.
  - The experimenter feels guilty; because no difference has been recorded between this situation and one in which there *is* an outcome that makes the participants fully happy.

3. Could there be a test that allowed for gradient gappiness?

- Surely — rating every item on a Likert scale would do it.
  - Getting a crummy rating for all options surely diagnoses semi-gaphood.
- I tried to get data that could be easily modeled — let "no answer" serve as a choice equal among others.

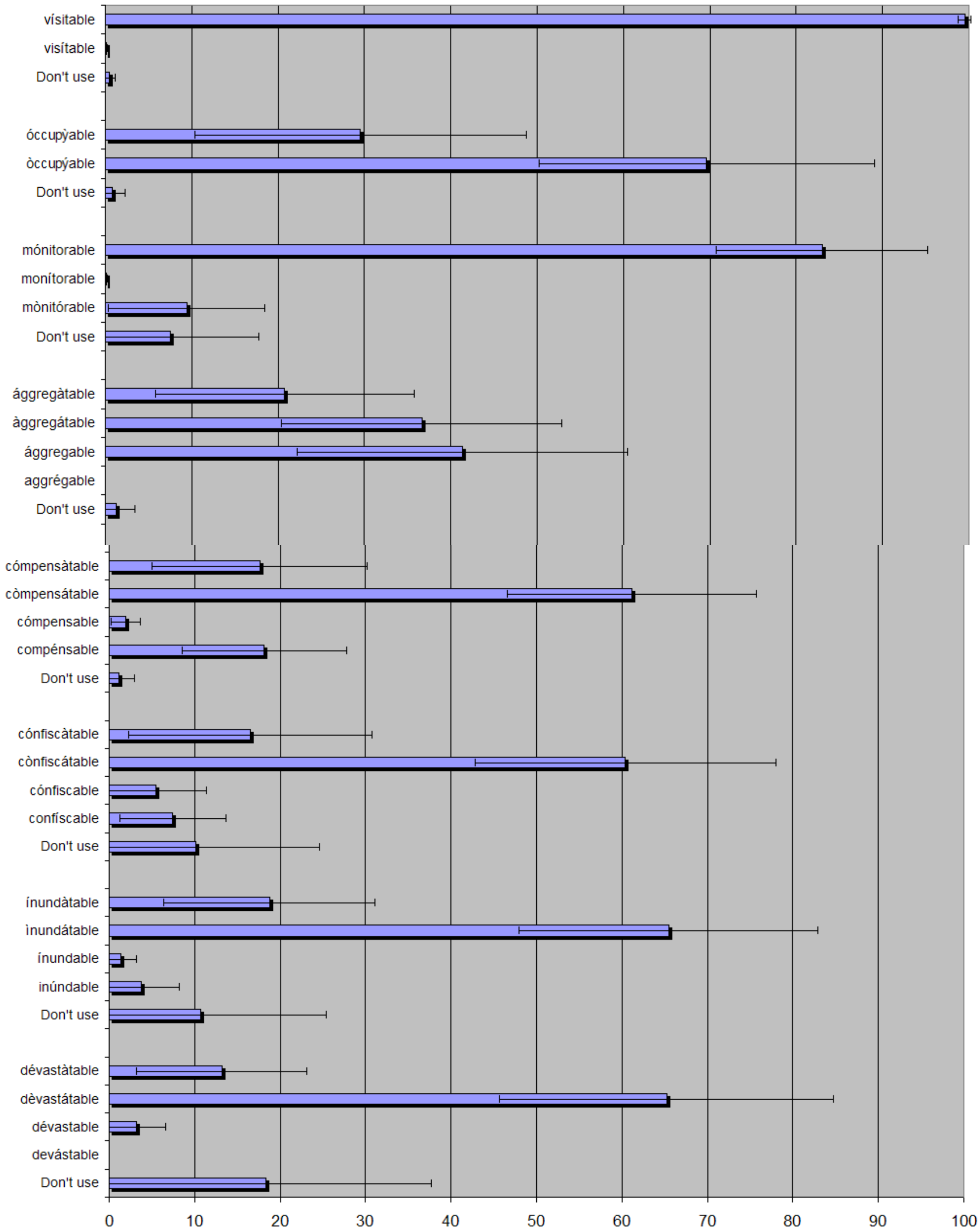
4. In practical terms, my test was a flop

- Perhaps some people didn't understand the instructions? Most never assigned a positive value to "no answer".
- Rebellious participants said "no way" and used a Likert scale instead of following the instructions.
- Occasionally, people gave 100% to "no answer" (accounting for most such responses).
- So, in what follows, beware "garbage in, garbage out".
  - I still thought the results were interesting, though, because they actually fitted the model I had had in mind.

5. How I compiled the data

- If people entered no value in "No answer" I took this as a zero.
- All responses not summing to 100 (including the rebellious Likert scale values) were summed and the results divided by 100. This overcame any addition errors.

6. Raw results with error bars



THEORY I: THE NULL PARSE THEORY OF ILL-FORMEDNESS

7. Origins

- This dates from early, nonstochastic OT; the ursource is Prince and Smolensky (1993/2004).
- At crudest level: non-output is a candidate competing with the others at all times.
- Variants of the theory: just what constraints are violated by the null parse candidate?
- Prince and Smolensky (p. 51) say: it really depends on the constraints in question:

> The key, we suggest, is that among the analyses to be evaluated is one which assigns no structure at all to the input: the Null Parse, identical to the input. The Null Parse will certainly be superior to some other possibilities, because it vacuously satisfies any constraint pertaining to structures that it lacks. For example, FTBIN says if there is a foot in the representation, then it must be binary; violations are incurred by the presence by nonbinary feet. The Null Parse therefore satisfies FTBIN, since it contains no feet of any kind. Similar remarks hold for syllable structure constraints such as ONS, because the Null Parse contains no syllables; for structural constraints such as FILL, which demands that empty nodes be absent (they are). Of course, the Null Parse grossly fails such constraints as PARSE, which demands that segments be prosodically licensed, to use Itô's term, because the input will always contain segments. The Null Parse will fail LX.PR [every lexical word must be big enough to be a prosodic word; cf. Seuss's *[nʌ]] when the input string is a lexical category, because the constraint applies to all items in the category; it says that all such items must be parsed as prosodic words.

- In the end, they adopt something akin to DON'T NOT HAVE AN OUTPUT, though expressed a bit more opaquely than this:

> Failure to achieve morphological parsing is fatal. An unparsed item has no morphological category, and cannot be interpreted, either semantically or in terms of higher morphological structure. This parallels the phonetic uninterpretability of unparsed segmental material. The requirements of higher order prosody will parallel those of higher order morphology and syntax: a phonological Null Parse, which assigns no Prosodic Word node, renders a word unusable as an element in a Phonological Phrase (Selkirk 1980, Nespor & Vogel 1986, Inkelas 1989), which is built on prosodic words. This is the structural correlate of phonetic invisibility. Members of the PARSE family of constraints demand that the links in the prosodic hierarchy be established; let us use **MPARSE** for the constraint which requires the structural realization of morphological properties.

8. The Null Parse is not the same thing as saying nothing.

- You can tell this when you put true null into a sentence.

- Let us suppose that **fishish* is utterly ungrammatical in English.
- Then two outcomes for “It has a fishish smell”:
  - It has a smell.
  - [“I can’t say this sentence because one of the words in it doesn’t exist.”]
  - Cf. the word *have*, which arguably sometimes has an utterly null realization in English; it reduces to [əv], which loses [ə] due to a preceding hiatus and [v] due to a following consonant: fast speech “Y’know, on some occasions [aɪ gɑtən sɪk] from eating that stuff.”

9. A modern update on the null parse: McCarthy and Wolf (2010)

- Ref.: McCarthy, John J. & Matthew Wolf (2010). Less than zero: Correspondence and the null output. In Curt Rice (ed.) *Modeling Ungrammaticality in Optimality Theory*, London: Equinox. Earlier version at ROA 722.
- They have a pleasingly simplified view of the constraint violations of the null parse:
  - It violates *no* constraints other than MPARSE.
- Result: it’s like a **harmony threshold**: no form can surface if all of its candidates violate some constraint ranked higher than MPARSE.

10. The null parse and stochastic grammars

- In stochastic grammar, we have the possibility of assigning some probability between zero and one to the null parse.
- Zero: the actual winner, if just one of them, will be utterly perfect [complications for multiple winners]
- One: this form is utterly ungrammatical
- It is tempting to assign interpretations to intermediate values. Perhaps they predict that the experimental participant experiences some degree of discontent, as noted above.

THEORETICAL BACKGROUND II:
THE -ABLE PROBLEM AS A CASE OF LEXICAL CONSERVATISM

11. Source

- “Lexical conservatism and the notion base of affixation”, on line at Donca Steriade’s web site.

12. Basic idea

- We store all the allomorphs of a stem.[1]
- E.g. if we know:

compensate [ˈkɑmpəns-eɪt]

[1] Perhaps not those from allophonic rules…

compensatory [kəmˈpɛns-əˌtɔri]

then we store [ˈkɑmpəns] + [eɪt] and [kəmˈpɛns] + [əˌtɔri]

- Why store the whole word (parsed)? See below — we'll need to know the word that the allomorph comes in in order to enforce a special constraint type.

- **Lexically conservative word formation**: you can take affixes and add new suffixes to form new words.
  - Typically you pick the one that produces the best fit to the well-formedness principles of the language.
  - Not all word formation processes are lexically conservative,[2] but English formation of deverbal adjectives in *-able* appears to be.

13. An example of a stem allomorph choice that is a good fit to the principles of the language

- Forming an -able adjective from *compensate,* it's best to pick [kəmˈpɛns], not [ˈkɑmpəns]. Why?
  - English dislikes stress more than three from the end: words like ˈ*hesitancy* are rare.
  - English doesn't like the sequence: *stressed syllable + closed stressless syllable + stressless syllable*: words like *galaxy* [ˈgæ.lək.si] are rare.

14. Plausibly, there is also a penalty for using semantically mismatched bases

- *Compens*(*at*)*able* means "able to be compensated", not "able to be treated in compensatory fashion"[3]
- Hence, it makes sense to prefer the allomorph [ˈkɑmpəns], not [kəmˈpɛns]; the former is the "true base".

15. What happens if you violate Lexical Conservatism?

- This happens in cases like *inundable* *[ɪˈnʌndəbəl] 'able to be inundated'.
- *Inundate* has no allomorph [ɪˈnʌnd] in its derivational paradigm.

16. Steriade on a roll?

- Lexical conservatism plays a pretty marginal role in English.

[2] Or so I think. If I remember rightly, German *-chen* 'diminutive' can produce newly umlauted stems.
[3] Cf. "I felt that the defendant was probably compensatable"; ??"I thought that the deleted nasal was probably compensatable and therefore anticipated lengthening on the preceding vowel".

- It plays a large role in French adjectival liaison, Romanian word formation (effect of palatalization), Ukrainian and Russian word formation (effect of stress; cf. Igor's talk of coauthored work last week).

MODELING THE -ABLE RESULTS

17. Constraint set

Constraint	*Abbreviation*	*Comment*
*NULLPARSE		See above
*NOT RIGHTMOST-NONFINAL MAIN	**ALLIGATOR*	In most English words, main stress falls on the rightmost nonfinal stressed syllable (S. Schane, 1970's). Think of *ˈopeˌrate, ˌopeˈration.*
OO-MAINSTRESS		Keep the main stress on the same syllable where it was located in the base. An output-to-output Faithfulness constraint (Benua 1997).
*UNSOURCEDFULL		Do not have a full vowel quality in a derived form that does not have a matching vowel in some allomorph of the derivational paradigm (Steriadean lexical conservatism). Violated by *iˈhundable* [ɪˈnʌndəbəl].
*FOOT-MEDIAL HEAVY	**GALAXY*	Avoid the sequence *stressed syllable* + *heavy stressless syllable* + *stressless syllable*. Violated by ˈ*compensable* [ˈkɑmpənsəbəl].
*DISTAL BASE		Do not use a listed form of the stem that fails to occur in the base form from which the candidate word is derived; cf. *compensable* [kəmˈpɛns-əbəl], using [kəmˈpɛns] from *compensatory*.

18. Violations (OTSoft file)

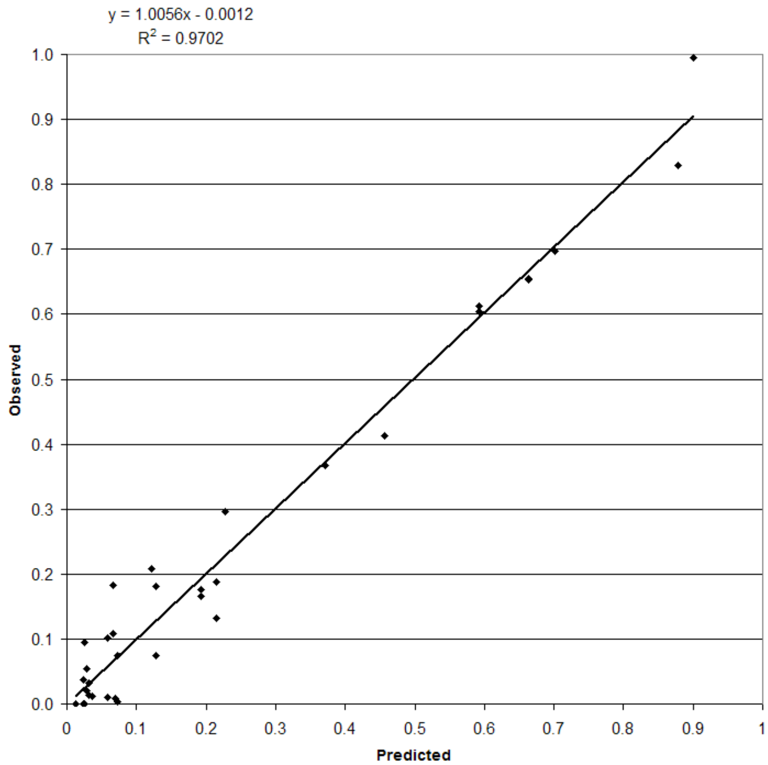
		Frequency	*NullParse	**alligator*	OO-Main	*UnsourcedFull	**galaxy*	*DistalBase
visit	vísitable	99.5						
	visítable	0.08			*	*		
	Null parse	0.42	*					
occupy	óccupỳable	29.5		*				
	òccupýable	69.6			*			
	Null parse	0.83	*					
monitor	mónitorable	83						
	monítorable	0.08			*	*		
	mònitórable	9.44			*	*		

	Null parse	7.5	*					
aggregate	ággregàtable	20.8		*				
	àggregátable	36.7			*			
	ággregable	41.3						
	aggrégable	0			*	*		
	Null parse	1.25	*					
compensat e	cómpensàtable	17.6		*				
	còmpensátable	61.2			*			
	cómpensable	1.97					*	
	compénsable	18.2			*			*
	Null parse	1.08	*					
confiscate	cónfiscàtable	16.6		*				
	cònfiscátable	60.4			*			
	cónfiscable	5.45					*	
	confíscable	7.44			*			*
	Null parse	10.1	*					
inundate	ínundàtable	18.7		*				
	ìnundátable	65.5			*			
	ínundable	1.34					*	
	inúndable	3.71			*	*		
	Null parse	10.8	*					
devastate	dévastàtable	13.2		*				
	dèvastátable	65.2			*			
	dévastable	3.21					*	
	devástable	0			*	*		
	Null parse	18.4	*					

19. Fitted weights (from Maxent GrammarTool)

Constraint	*Weight*
*NullParse	2.51
**Alligator*	1.33
OO-MainStress	0.21
*UnsourcedFull	3.37
**galaxy*	3.25
*DistalBase	1.53

20. Scattergram: model fit



- Bear in mind that this is six weights predicting 35 (averaged) observations.

21. Significance testing

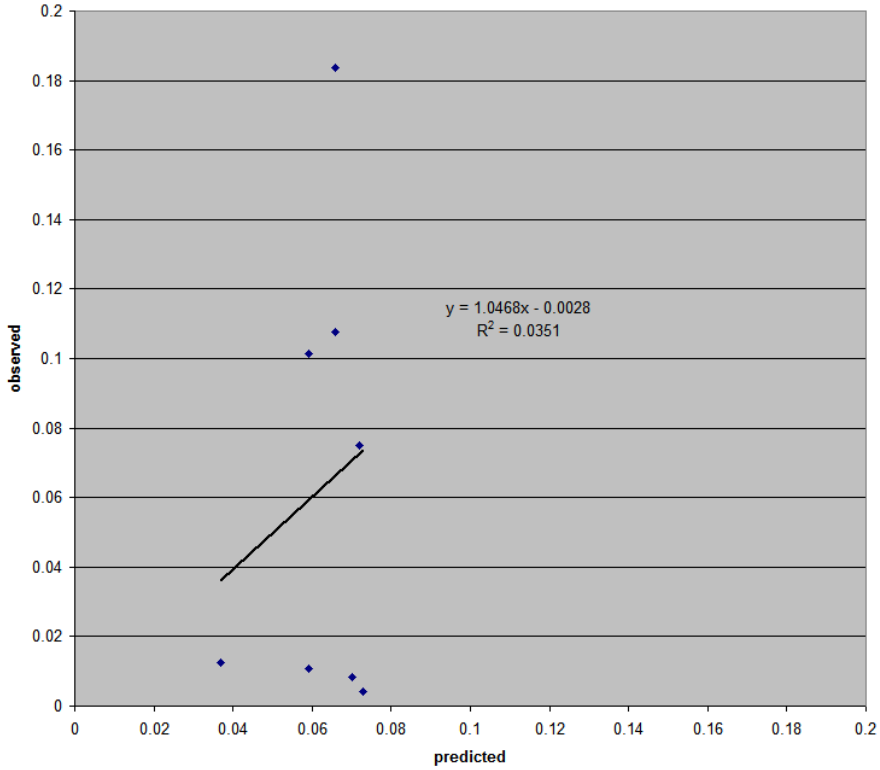
- I removed each constraint from the grammar one at a time.
- I did a likelihood-ratio test in each case. [We will cover this soon.]
- All constraints tested highly significant except OO-MainStress.
  - ➢ Why not significant? It's in close opposition to **Alligator*, and maxent typically likes to let just one of two opposed constraints bear weight.

22. The essential Steriadean comparisons

- *iˈnundable* and *deˈvastable* should be worse than *conˈfiscable* and *comˈpensable*, due to *UNSOURCED FULL.
  - ➢ Values from experiment: 3.71%, 0% vs. 7.44%, 18.2%
  - ➢ Values from model: 0.23%, 0.23%, 12.8%, 12.8%
- *ˈaggregable* should be better than *ˈconfiscable*, *ˈinundable*, *ˈdevastable*, and *ˈcompensable* because of **galaxy*
  - ➢ values from experiment: 41.3% vs. 5.4%, 1.3%, 3.2%, 2.0%
  - ➢ values from model: 45.7% vs. 1.3%, 3.2%, 3.2%, 2.8%
- i.e. we get both Lexical Sourcing and simple Markedness effects.

23. Fit of the model re. the Null Parse candidate

- Not good at all:



24. Where could the model fit be improved?

- There ought to be a penalty for preantepenultimate stress, seen in the inferior status of ˈ*monitorable* relative to ˈ*visitable*, participants 83%, 99.5%, model 90.1%, 87.9%.

25. What does it all mean?

- I'm amazed that such an ill-designed experiment (12 subjects (one rebellious, 3 drunk), few stimuli, colossal response variances) got results that were so linguistically interpretable and so closely matched to my theoretical expectations.
  - ➢ Perhaps a more careful replication would yield otherwise.
  - ➢ I wouldn't trust many replications because I think you need a subject population comfortable with words like *confiscatory*.
- If the results are correct, they vividly illustrate Cowart's Law of Psycholinguistics:[4]
  - ➢ Result that look like garbage on the individual level can aggregate to meaningful and theory-matching findings in the aggregate.
- I can't see any real evidence for or against the Null Parse theory of ungrammaticality here, given that the task didn't resonate well with the subjects and model fit was poor.

[4] See Wayne Cowart's remarkable book, *Experimental Syntax*.

TWO MODELS (OR ARE THEY?) OF GAPPINESS AND FREQUENCY

26. Kie's modeling last time

- This was a heroically-scaled examination of frequency in English word formation.
- It raised an issue I'm not really solid with, sort of as a side issue, which I'd like to informally explore.

27. Kie's two methods

- I: one "dummy input", then every logically-possible output (combination of existing stem, existing suffix) with frequencies as training data
  - These frequencies are type frequencies, hence either zero or one.
  - call this **global choice**
- II: every logically-possible combination of existing stem, existing suffix is an input. Output is 1 for "exists", 0 for "does not exist".
  - call this **local choice**
- In practice, the Maxent Grammar Tool choked on I, so Kie went to R and did II.
- My naïve query: Is there any difference between these two methods?

28. A background comment

- Both global choice and local choice use "**synthesized negative evidence**".
- I.e., you use a basic knowledge of the domain (these stems exist, these suffixes exist) and assume free combination to create a large universe of candidates.
- The grammar is then set up to allocate maximal probability, among this universe, to the existing candidates.
- Difference:
  - global choice: maximize probability to "existors"
  - local choice: maximize probability to "yes" for existors, to "no" for nonexistors.

29. Global vs. choice in the research literature

- Hayes and Wilson (*LI* 2008) use global choice for English phonotactics.
  - The synthesized negative evidence took the form of a gigantic finite state machine, representing all possible strings below a certain length. Program runs slow …
- Albright (2012[5]) and McClelland and Vander Wyck (ms.)[6] represent local-choice models.

---

[5] Albright, Adam (2012) Additive markedness interactions in phonology. Handout for colloquium at UCLA. On course web site.

[6] James L. McClelland and Brent C. Vander Wyk. 2006. Graded Constraints on English Word Forms. Ms. Posted at http://psych.stanford.edu/~jlm/papers/GCEWFs_2_18_06.pdf.

- If I understand Albright correctly, he thinks a local-choice model works better, capturing "hyperadditive" effects.

30. A toy run contrasting local and global choice

- Bottom line in advance:
  - they yield results in essentially a perfect ratio.
  - the local-choice model allows (and perhaps sometimes needs?) an additional parameter?

31. A scaled down version of Kie's model

- In Pseudo-Language, the stems can be vowel-final or consonant-final:

  pa, ta, ca, ka, fa, sa, sha, xa, ma, na, nja, nga
  pang, tang, cang, kang, fang, sang, shang, xang, mang, nang, njang, ngang

  - These are all nouns.

- There are two suffixes that attach to nouns to form adjectives, namely [-oid] and [-like].[7]
- [-like] is a more productive than [-oid].
- [-like] has a tendency to attach to vowel-final stems, [-oid] to consonant final stems.
- They are not entirely mutually exclusive, nor is there always an output.
- The numbers:

Stem type	Takes -like	Takes -oid	Takes both	Takes neither
V-final	10	0	1	1
-C-final	6	5	1	0

32. Constraints employed

- *VV — penalizing [-oid] attached to vowel stems
- *CC — penalizing [-like] attached to consonant stems
- *oid — reflecting slightly more productive status of [-like].

33. A local choice model

- One further constraint: EXIST, penalizing the "don't exist"
  - This is analogous to classical OT MPARSE.

[7] Pronounce them as IPA; the resemblance to English is accidental but mnemonic.

- Input file (sample):

		Winner	Exist	*VV	*CC	Avoid - oid
palike	Yes	1				
	No		1			
paoid	Yes			1		1
	No	1	1			
panglike	Yes	1			1	
	No		1			
pangoid	Yes					1
	No	1	1			

34. Local choice model: constraint weights

Constraint	Weight
Exist	2.397744
*VV	2.397842
*CC	2.061257
Avoid oid	2.397744

- The identical weights reflect some 50/50 probability in the data.

35. Local choice model: predicted probability of existence for the four data categories

0.083	-oid after V
0.5	-oid after C
0.917	-like after V
0.583	-like after C

36. A global choice model: constraints employed

- *VV — penalizing [-oid] attached to vowel stems
- *CC — penalizing [-like] attached to consonant stems
- *oid — reflecting slightly more productive status of [-like].

- EXIST is not employed.

37. Input file — sample

			*VV	*CC	Avoid oid
Input	palike	1			
	paoid		1		1
	panglike	1		1	
	pangoid				1
	shanglike	1		1	
	shangoid	1			1
	ngalike				
	ngaoid		1		1

38. Global choice model: constraint weights

Constraint	Weight
*VV	1.79
*CC	0.45
Avoid oid	0.61

- Weights are smaller.

39. Analogous weights in the two models have no obvious relation

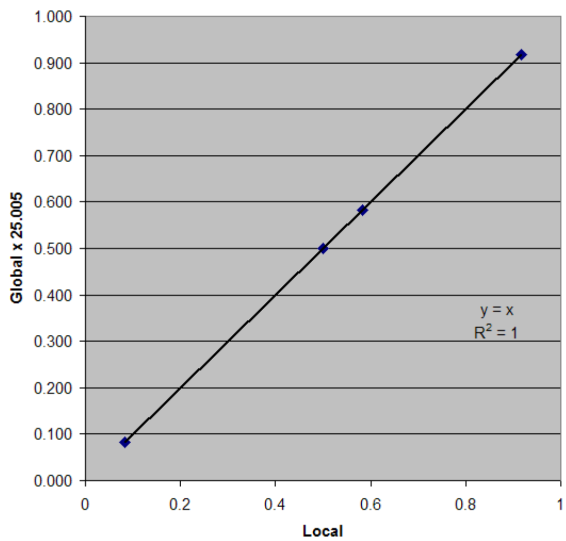
	Global	Local
*VV	1.79	2.40
*CC	0.45	2.06
Avoid oid	0.61	2.40

40. Global choice model: predicted probability for the various types

0.023	-like after C
0.037	-like after V
0.003	-oid after C
0.02	-oid after V

- These are small numbers, since they reflect the probability assigned to individual types in a large universe.

41. Probabilities of the two models are perfectly correlated



- You get the local probabilities by multiplying the global ones by 25, which is the total number of data.

- (25.005 fits better, perhaps because the weights were calculated with a small Gaussian prior.)
- It makes sense that these probabilities will match, since both models were trying to maximize the overall probability of the data.

42. Did the local choice model use an extra parameter?

- If you leave out the constraint EXIST, the local-choice model does horribly.
- Weights:

Constraint	Weight
*VV	2.40
*CC	0
Avoid oid	1.60E-05

- This makes sense: if existence is only penalized, then winners can never get a probability greater than .5 (i.e. zero violations).
- Kie's original model included an "AVOID" constraint for every suffix; AVOID *-like* would serve as an equivalent to our EXIST.

43. The local choice model is richer, but perhaps in a trivial way

- Imagine that EXIST has an extremely high weight, with other constraints having their old sensible weights.
- Then everything gets a fairly high probability for Yes, though the nuances remain.
- Ditto: relatively low weight for EXIST;[8] everything sounds pretty bad, with nuances remaining.
- The global-choice model conflates these distinctions.

A NOTE ON THE LOFSTEDT READING

44. Basic story

- Swedish vowel length is contextually determined; essentially short before consonant clusters, long elsewhere.
- The nine Swedish long-short vowel pairs have two outliers, notable for their perceptual distance:
  - ɑː vs. a
  - ʉː vs. ɵ
- These behave differently in various places in the phonology.

[8] Not *too* low! Else EXIST will squeeze the conflicting markedness constraints downward toward zero, wiping out the nuances.

45. Example of special behavior

- Optional (but normal) coalescence of r + voiced alveolar to a single retroflex, with accompanying vowel made long
  - Careful speech: børd
  - More normal speech: bøːɖ
- But if it's one of the distant-pairs, you keep the /r/:
  - Careful speech: gard
  - More normal speech: also gard, not *gɑːɖ[9]
- Analytic strategy: the constraint favoring retroflex coalescence is outranked by *Map(a, ɑː) but not by *Map(ø, øː).

46. Paradigm gap

- To form a neuter adjective in certain contexts, suffix a [-t].
- If the stem ends in a [t] or [d], this will create a geminate.
- … which will in turn shorten the stem vowel.
  - heːt 'hot' + -t → heːtt → hett → hɛtt 'hot-neuter'
- But if you try to do this with a distant-vowel pair, you get a paradigm gap.
  - lɑːt 'lazy' has no neuter; if it existed it would be *[latt]
- Lofstedt gets this in the classical way, ranking *Map(ɑː, a) over MPARSE.

47. Might Wilsonian/Whitian methods be applied to Lofstedt's work?

- In Chapter 5 (not assigned), Lofstedt observed that highly frequent forms actually *can* alternate distant vowels: [glɑːd] ~ [glatt] 'happy'.
- He atomizes *Map() by frequency, with the *Map() penalizing more frequent forms ranked lower.
- An alternative to consider is diachronic simulation, explaining how Swedish got where it is today.
- Assume a P-map-based set of weights for the vowel pairs.
- This could serve as the basis for biased learning: you need more input data to override your P-map-based reluctance to alternate vowel quality.

[9] Random comment: might this be an exception to the "no look-ahead" principle of Harmonic Serialism (McCarthy 2007 et seq.)?