Class 3, 4/7/2020: Statistical testing, phonotactic analysis, more on frameworks

1. Assignments

- Read:
  - Michael Becker, Andrew Nevins, and Jonathan Levine (2012) Asymmetries in generalizing alternations to and from initial syllables. *Language* 88:2, pp. 231–268.
  - On course web site
- Continue homework on medial clusters.
  - Due in a week, Thursday April 16.

2. Outline for this class

- Statistical testing and other virtuous MaxEnt spreadsheet activities
- Basics of the GEN-theory of phonotactics
- Demo for the homework: Warlpiri — techniques of cluster analysis
- Intuitivity of MaxEnt
- Noisy Harmonic Grammar as an alternative to MaxEnt

3. Forthcoming

- Framework comparison
- UG bias and how to implement it in grammar learning (next reading)
- If time: the Observed/Expected fallacy and how to avoid it in MaxEnt

IMPROVING THE /F/ VOICING ANALYSIS, AND STATISTICAL TESTING

4. Fixing an error

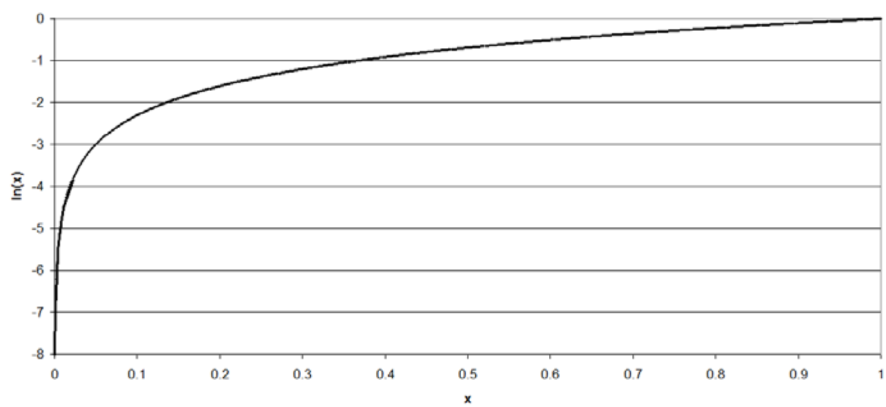
- Thanks to Jinyoung for pointing this problem out after the last class.
- We only included *special factors* in the /f/ voicing analysis.
- But surely we need a *baseline*, giving general dispreference for /f/ voicing.
- This is indeed part of standard phonological theory (McCarthy and Prince 1995): IDENT(voice)
- Let’s try to put it in, and see what happens.
- We will also take this as an opportunity to teach a couple things:
  - log likelihood
  - the Likelihood Ratio Test in statistics

5. Log likelihood

- We have already covered the "maximum likelihood" method for setting model parameters (here, constraint weights)
    - Assign as big as possible a share of the "probability mass" (total 1) to existing forms.
    - This assigns minimal probability to non-existing forms, which is what a grammar should do.

6. Shifting to log likelihood

- For any realistic problem, likelihood values are extremely low, so to keep Excel from crashing we instead take the *natural logarithm* of the likelihood ("log likelihood").
    - Natural logarithm is the log to base *e* (about 2.72)
- This preserves the relative goodness of different solutions (monotonic) but keeps the numbers manageably small.



7. Likelihood Ratio Test: a source and citable reference

- Wasserman, Larry (2004) *All of statistics: a concise course in statistical inference.* New York: Springer, p. 164.

8. Recipe

- We are already calculating log likelihood
- Consider two **nested** analyses, called Big and Little:
    - Big has all the constraints of Little, plus more
- Let **degrees of freedom** be the number of extra constraints in Big.
- Take the **difference** in log likelihood of the two models.
    - This is a measure of how much improvement you get by adding the new constraints to Little.
- Double this difference.
- Look up a p-value from this chi-distribution
    - You can use CHIDIST( ) function in Excel

- =CHIDIST(A1,1) means:
  - Cell A1 has the doubled log likelihood difference
  - 1 is the number of degrees of freedom
- Let's do this now for adding IDENT(voice) to our existing model.
- Criterial value for p is a social construct: very strict in sciences that can afford it; less strict in sciences with messy data.
- p-values themselves are enormously controversial, as you can Google; stats is reforming and this recipe may soon be obsolete.
- Still, it's nice to do at least *something* to argue you are not deluding yourself.

9. Two other helpful things you can do with a MaxEnt spreadsheet

- 1. Just plain count the violations; it's good to know.
  - And: when there are zero, you know the weight is arbitrary
- 2. For intuitiveness, convert the weights into *probability penalty for violating*
  - Here is the math; probability ratio of Candidate 1 and Candidate 2, differing in only one constraint violation

$$\frac{\text{P(Candidate 1)}}{\text{P(Candidate 2)}} = \frac{\frac{e^{-H}}{Z}}{\frac{e^{-(H+w)}}{Z}} = \frac{e^{-H}}{e^{-(H+w)}} = \frac{e^{-H}}{e^{-H} * e^{w}} = \frac{1}{e^{-w}} = e^{w}$$

where

H = harmony penalty resulting from shared violations
w = penalty for violating constraint C

  - So we can add a row with these intuitive probabilities

PREPARING FOR THE HOMEWORK ON MEDIAL CLUSTER ANALYSIS

10. A paper that actually does this

- (ms.) Bruce Hayes and Jinyoung Jo, Balinese stem phonotactics and the subregularity hypothesis. Posted at https://linguistics.ucla.edu/people/hayes/
- Jinyoung and I have a particular agenda, involving a key claim of the Delaware School …

11. Backdrop to this homework: An encounter I had

- My appointment with a thoughtful grad student at Cornell.
- Says,
  - "Everybody thinks the Syllable Contact Law is relevant to my language's phonology, but I'm not so sure."
  - "It seems to me that other, independently motivated constraints will do the work we attribute to the Law, which is then perhaps not needed."

- I think we need some way of testing such claims, in principle more precisely than "satisfying account".
- Perhaps creating a complete model, assigning a probability to every logically possible medial cluster, might help — does Syllable Contact Law help, at a statistically significant level?

- More generally: in a phonotactics with "imperfect" constraints, how can we assess what the constraints are doing for us, or how good the analysis is.

12. The "Markedness Only" approach to phonotactics

- To my knowledge this was invented by Hayes and Wilson (2008), though the idea is pretty obvious.
  - Bruce Hayes and Colin Wilson. (2008) A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39, 379-440.
- Assign a probability to every form in GEN.
- Or, perhaps, every form in GEN less than 20 phonemes long…
- This can only use Markedness constraints — so things like Positional Faithfulness cannot be used.

13. How phonotactics is done in classical OT (Prince and Smolensky 1993)

- Rich Base: everything can be an input
- Grammar as filter: some inputs get changed to something else.
- The full set of "something elses" and survivors form the set of legal forms.
- I worry about the ability of this system to capture marginal cases
  - The key problem is *unrepaired moderate badness*
  - e.g. ?[pɔɪk] is mildly aberrant to me, but I have no inclination to repair it (e.g. to [paɪk]).

14. The goal at hand (homework)

- Suppose the Markedness Only theory of phonotactics is correct.
- Doing whole languages is a huge job (see Hayes/Wilson, and their software, which uses finite state machines to cover vast sets of strings).
- But medial consonant clusters: V<u>CC</u>V as manageable: GEN is only the square of the number of consonants.
- So: obtain a full, explicit, gradient analysis of some language's phonotactics, using maxent.

15. Note on vowel sequencing as an option

- If you want to do vowels instead, that would be fine with me.
- E.g.,
  - Does a language have tacit vowel harmony?

    - Does a known vowel harmony language obey its own rules within stems?
- I suggest you start with just disyllables, since initial vowels often are very special for vowel harmony systems.

16. Step 1: obtain an electronic lexicon from the Internet

- You want phonemic listings (IPA not essential).
- Hopefully not too huge a consonant inventory
- Perhaps useful not to have too many VCCCV.
- I have found I sometimes have to steal the data one letter at a time.

17. Sample solution

- I did **Warlpiri** (Australia, a focus of UCLA grad Margit Bowler, outstanding linguists have worked on it for decades; good online resources and book references).
- I also made extensive use of the excellent 1980 MIT dissertation by David Nash, which covers the phonology and particularly the medial clusters.
    - MIT: "DSpace" website

18. Grabbing the dictionary

- Download the whole online dictionary one initial letter at a time.
- Discard all but the entries:
    - In Word, replace every space with a tab.
    - Paste into Excel, and keep only the first column.
    - Sort that column and discard crud.

19. Forming a list of medial clusters with counts

- Harvest the medial clusters:
    - Paste first column of spreadsheet into Word, then
    - Replace the long vowel digraphs *aa*, *ii*, *uu* with single symbols.
    - Replace every vowel with *tab vowel tab*
    - Paste result back into Excel and intervocalic consonants and clusters are all in the same column! (no vowel initial words or hiatus)
- Reduce the medial clusters, original a list of tokens, to single counted types
    - I use my Typizer, toy software I can share.
    - In Excel a pivot table will do it.
- Discard the singletons (VCV)
- Starting with a list of the consonant phonemes, make a list in Excel of all logical combinations.
- Plug in the frequencies for the attested and zeros elsewhere.
- Now you are ready to analyse!

20. Maxent analysis of clusters on a spreadsheet

- Use columns that don't calculate but embody the **features** of C1 and C2.
  - E.g. Voice(C1)
- Use the logical formulas in Excel to calculate violations
  - = IF(condition, value if true, value if false)
  - = AND(this = this, this = this)
  - ditto OR()
  - These formulae can be references to feature values, or to segment identity.
- The rest is just plain phonology: use your brain/phonological experience/guile to find really good constraints, and watch the log likelihood go up.
- Don't let the bizarre ambiance (stats, spreadsheet) make you forget you are a phonologist! Find constraints that are
  - General
  - Typologically well-motivate (see below)

USEFUL INTUITIVE ANALYTIC METHODS FOR THIS KIND OF WORK

21. The n by n chart

	p	t	ʈ	c	k	m	n	ɳ	ɲ	ŋ	l	ɭ	ʎ	ɾ	ɽd	ɹ	w	j
p																		
t	3			2	1													
ʈ	1			1														
c	3																	
k																		
m	113																	
n	92	11			61	14				16								
ɳ	76		142	8	33	8				17								
ɲ	53				15	1												
ŋ					132													
l	14	7		52	77	1											5	
ɭ	85		14	16	6					2							1	
ʎ	89			2	33					2								
ɾ	171			26	117	22				18							8	1
ɽ																		
ɹ	4				1	1			1									
w																		
j	1																	

- Note that columns and rows are sorted in strict IPA order.

22. Methods of model evaluation

- A **scattergram** of observed/predicted shows fit, outliers clearly.

  - be sure to adjust the axes to go from 0 to 1
  - be sure to make the graph square
- It is useful to include a column that detects the biggest **overgeneration error** (higher predicted probability than observed probability).
  - Then you can try to find a constraint that penalizes the overgenerated clusters
- *Not recommended*: **correlation coefficient** of predicted and observed
  - Often contradicts the maximum likelihood criterion
  - Why? (my guess): correlation merely checks whether the model can be mapped with linear equation to the real data (y = mx + b). But we want complete identity.

23. Navigating the thicket of possible analyses

- Suppose you have invented a lot of constraints
- Some of them may overlap.
- This produces a messy thicket — how to extract a clean, effective phonological analysis?

24. Reference sources on this

- (2012) Bruce Hayes, Colin Wilson, and Anne Shisko, Maxent grammars for the metrics of Shakespeare and Milton. *Language* xx:xxx.
- Armed with 87 constraints they wanted to test, tried:
  - bottom up (add best constraint till no significant improvement)
  - top down (delete worst constraint until you would delete a significant one)

  with similar results.
- More ambitiously, a text: Kenneth Burnham and David R. Anderson (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*

25. Sharing the lore about cluster markedness

- Is there a really good reference somewhere?
- Many principles are *nonsyntagmatic*, simply restricting onsets and codas
  - *voiced obstruent in coda
  - *noncoronant in coda
  - *glide onset
- Many things assimilate (AGREE(feature))
  - voicing in obstruents
  - place, especially nasals before stops
  - secondary articulations
- Sonority sequencing: there is a literature on the Syllable Contact Law
  - To the extent that C1 is low sonority and C2 is high, assess violations
  - You can do this by assigning integer sonority values to C1 and C2, and subtracting

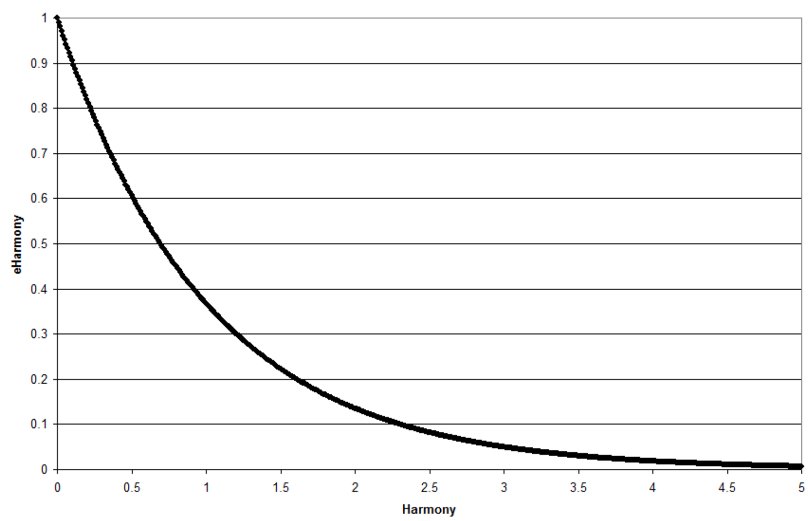
MAXENT AND INTUITIVENESS

26. Key idea

- Think of constraints as *evidence* in favor of or against particular candidates.
- Then we can compare the MaxEnt decision procedure to our own internalized sense of rational decision making.

27. The maxent procedure thus assessed

- Every constraint has a weight, a non-negative number.
  - ➢ This is intuitive; reasons differ in cogency.
- Every candidate is given a **harmony score**.
  - ➢ = weighted sum of its violations
  - ➢ i.e. pairwise multiply weights and violation counts, and sum up
  - ➢ This is intuitive: weigh **all** the evidence. ☞ what theory does not?
- Every harmony score is converted to an eHarmony[1] score, by negating it and taking *e* to that power. *e* is about 2.718.



  - ➢ Note that Harmony is a badness score (penalty), but eHarmony is a goodness score (virtue)
  - ➢ The conversion of Harmony to eHarmony is intuitive, because once Harmony is very big, only very small gains in eHarmony are made when Harmony increases (and, just below, probability will be derived from eHarmony).
- Take all the eHarmony scores and add them up. By tradition this number is called **Z**.
- The probability assigned to a candidate is the share of its eHarmony in Z; in other words, divide the eHarmony value by the Z value.
  - ➢ So, probabilities will sum to 1.

---

[1] Caution: I really like this term, invented by Colin Wilson. But it is tricky to use in writing, since it is a joke (eHarmony is a dating site on the internet).

- This is intuitive, since a choice is less appealing when there are strong alternatives.

THE FRAMEWORK BAZAAR

28. Noisy harmonic grammar

- References:
  - Boersma, Paul, and Joe Pater. 2008/2016. Convergence properties of a gradual learning algorithm for Harmonic Grammar. Amsterdam and Amherst, MA: University of Amsterdam and University of Massachusetts ms. Rutgers Optimality Archive. Published 2016 in John McCarthy and Joe Pater, *Harmonic Grammar and Harmonic Serialism*
  - Potts, Christopher, Joe Pater, Karen Jesney, Rajesh Bhatt & Michael Becker (2010). Harmonic Grammar with linear programming: From linear systems to linguistic typology. *Phonology* 27, 77-117. (non-stochastic version)
- This is a lot like maxent; again you calculate a Harmony score for every candidate.
- But you jiggle the harmony scores stochastically, deriving a winner for each evaluation time, just like in Stochastic OT.

29. Many varieties exist

- See
  - Bruce Hayes (2017) Varieties of Noisy Harmonic Grammar. *Proceedings of the 2016 Annual Meeting in Phonology*, USC.
- E.g., where do you put the noise?
  - On the constraint weights (= classical version)
  - In the tableau cells
  - On the harmony values (behaves amazingly like maxent)
- If there are several violations of a constraint, does this cause the noise to be similarly multiplied?

30. Assessment

- I personally feel this framework is in contention:
  - Performs about as well in practice (I suspect) as maxent.
  - No proof of convergence for learning algorithm, but I have never seen it misbehave.
  - Combines evidence from multiple sources in making predictions (in the very same way as maxent, its partner in stochastic Harmonic Grammar).

31. Calculating NHG analyses

- Online OT-specific software
  - Praat (Boersma and Weenink)

    - OTHelp (UMass)
    - OTSoft (Hayes website)
- Excel is easy, but only for problems where there are just two viable candidates.
    - I can do the math and demo if time.