Linguistics 219 Phonological Theory III
B. Hayes Spring 2019

Class 2, 3/4/2018: Interpreting Results; Cluster Analysis; The Framework Bazaar

1. Assignments

- Read Zuraw and Hayes (2017)
  - Covers the framework bazaar and makes an argument about who wins!
  - On web site.
- Start on your homework — medial clusters.
  - This is probably the biggest homework and is due in 12 days, Monday April 16.

2. What have we got so far?

- Linguistics extends its goals:
  - from its ur-homeland in providing satisfying accounts of patterns in a data corpus
  - … to attempts at prediction
- We are seeking a good framework of constraint-based linguistics, hoping to make predictions.
- Criteria:
  - Should be grounded in what mathematically-qualified people have found about valid inductive reasoning and predictive models — hence mostly likely probability theory.
  - Should account for variation in output and ambivalence in judgment, likewise, therefore probability.
  - Ambivalence in judgment = incomplete basis for belief, which fits with the Jaynesian view of what probability is.
- Maxent has sound mathematical foundations and seems a good thing to try.
  - It combines evidence from multiple sources (in linguistics, called “ganging”)
  - It demands more evidence to approach certainty.
  - We can borrow work by computer scientists\(^1\): we can use algorithms to match quantitative data optimally — cf. last time, hand-setting vs. machine-setting the weights for Tapping.

3. How to cite maxent framework

- My personal practice is to cite:

---


4. Return to our /f/ Voicing example

- We seek a model that includes constraints embodying various factors:
  - Why should [v] be favored in general?
  - Why should [f] be favored in general?
  - What circumstances totally rule out [v]?
  - What circumstances make [v] especially likely?
- We can now easily implement this with our data file, obtaining predictions about every word — existing, or wug words like heaf.
  - These attempt to be a model of the native speakers tacit degree of belief that a noun ending in [f] should take a [v] plural.

5. Looking at the output of the grammar

- Do a probability sort within categories and plot.
  - with more refined data, a scattergram is appropriate
- Are the forms predicted to be impossible, impossible?
- What are the most likely [f] plurals to be pronounced innovatively with [v]?
- What of Berko’s heaf form, where we already have a modest real probability value?
- What distinctions are made among the existing forms?

6. Significance testing for maxent models

- There are many ways to do this.
- The simplest is the likelihood ratio test.
- Double the improvement a new constraint makes in likelihood, use chidist(x, y), where
  - x is the doubled improvement
  - y is usually 1; but if you’re interested in the improvement from adding a batch of constraints, y is the number you are adding.

7. Logistic regression

- Maxent when there are just two viable candidates is called logistic regression.
- This opens up many further options.
  - Software, e.g. R.

---

2 A fine program for maxent in general is the Maxent Grammar Tool by Wilson and George, on my web site.
Different significance tests
Above all, random effects — Jesse Zymet’s dissertation topic (“lexical propensity”)

PREPARING FOR THE HOMEWORK: MEDIAL CLUSTER ANALYSIS

8. A recent personal experience

- 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?

9. The “Markedness Only” approach to phonotactics

- To my knowledge this was invented by Hayes and Wilson (2008), though the idea is pretty obvious.
- 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.

10. 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: ?[pɔɪk] is mildly aberrant to me, but I have no inclination to repair it (e.g. to [pɑɪk]).

11. 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: \[VCCV\] 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.

12. 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 sometime have to steal the data one letter at a time.

13. Sample solution

• I did Warlpiri (Australia, a focus of colleague 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.

14. 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.

15. 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!

16. Maxent analysis of clusters on a spreadsheet

• Add a lot of columns with feature values needed for both C1 and C2.
• Then use the formula = IF(AND( …. ), 1, 0) to assign constraint violations.
… can be references to feature values, or to segment identity.

- The rest is just plain phonology: use your brain/guile to find really good constraints, and watch the log likelihood go up.
- A scattergram of observed/predicted can lead to increased analytical excitement.
- It is useful to include a column that detects the biggest overgeneration error (higher predicted probability than observed probability).

THE FRAMEWORK BAZAAR

17. Historical theme

- As the “predictionist” approach increases (maybe) in influence, phonologists have gradually moved from models that are straightforward extensions of OT to models that are borrowed from probability and statistics.
- Perhaps this is right?
  - Why should we necessarily invent the best ways to go from data to prediction, when this is an issue addressed throughout science?
  - … admitting that, of course, language might be special…

18. Anttilean strata

- Formally:
  - Assume strata of constraints, each ranked collective above the next one down, but with free ranking within strata.
  - Construct all grammars compatible with the strata.
  - Assume they are equiprobable
  - This generates probabilities.
- This idea was proposed by Arto Anttila, e.g. in
- BH editorial opinion [caution: possible rant approaching!]: I’m not a fan:
  - Probabilities form a coarse set, with just a few values.
  - Very poor at combining evidence from multiple sources.
  - Of all the frameworks tried by Zuraw and Hayes (2017), this performed by far the worst on their data — primarily, for reason just given.
  - Why assume that human children are such crummy learners, when empirical work shows a remarkable ability to match frequencies in ambient data?
  - (For frequency-matching citations, see Zuraw/Hayes readings.)

---

3 It’s very elegant work, and I used to be a fan! My objections are only empirical.
• Anttila (and his colleague Paul Kiparsky) repeatedly have emphasized the restrictiveness of this approach; but excess-power arguments are trumped, I think, by insufficient-power arguments.

19. Boersmian Stochastic OT

• Every constraint has a number that represents its strength.

\[ C_1 \quad \text{(high ranked)} \quad C_2 \quad \text{(low ranked)} \]

• Jiggle each such number whenever you use the grammar, assigning a bit of Gaussian noise.

\[ \sigma \quad \mu \quad \mu - \sigma \]

• Once you have jigged, you get a complete constraint ranking, which generates a winner as in standard OT.

20. Part of a spreadsheet

http://www.linguistics.ucla.edu/people/hayes/GLA/RankingValuesToProbabilities.xls

<table>
<thead>
<tr>
<th>Difference in ranking value</th>
<th>Probability higher outranks lower</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>0.1</td>
<td>0.51</td>
</tr>
<tr>
<td>0.5</td>
<td>0.57</td>
</tr>
<tr>
<td>1</td>
<td>0.64</td>
</tr>
<tr>
<td>5</td>
<td>0.96</td>
</tr>
<tr>
<td>10</td>
<td>0.9998</td>
</tr>
<tr>
<td>50</td>
<td>1.00000000</td>
</tr>
</tbody>
</table>
21. A personal evaluation of this

- It can combine evidence from multiple sources, but only in limited ways and circumstances.
  - See Zuraw/Hayes paper for detailed diagnosis of where/how it fails to do this.
  - In essence, free combination often requires a constraint to be in two places at once on the Boersmian scale.
- Setting the weights (a.k.a. ranking values) has proven to be very problematic.
  - No provably convergent algorithm exists.
  - The leading one, the Gradual Learning Algorithm, behave erratically; fails on simple grammars (Pater (2008, *LI*)), and often sends the weights off toward infinity as no convergence occurs. Plenty of frustration in my own personal history as a user.
- It still has defenders, notably the stalwart Giorgio Magri, who has tried to improve the Gradual Learning Algorithm.

22. Noisy harmonic grammar

- References:

- 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.

### 23. Many varieties exist

- See
- 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)

### 24. 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).

RETURN FROM THE BAZAAR TO PONDER: WHAT ARE THE ISSUES IN FRAMEWORK CHOICE?

### 25. Ganging

- I’ve portrayed ganging as deeply rational and wholesome, but what are the linguistic facts?
- Perhaps one could say that ganging is increasingly noticed in phonology as people look for it.
- Some feel (unpublished work of Edward Flemming) that ganging is a property of optional phonology, and that “crystallized”, obligatory phonology doesn’t gang.
  - A tall order to explain, and worth pondering.

### 26. Harmonic bounding

- A harmonically bounded candidate in OT has a strict superset of the violations of a rival candidate.
- In classical OT, it can never win.
- In stochastic OT, it can never win.
- In maxent, it can, but never with the highest probability.
• Noisy Harmonic Grammar: usually it can (see Hayes paper), but there is one little-explored variant (Exponential Noisy Harmonic Grammar), in which it cannot.

27. Implication I

• Be very careful when you do analysis in maxent, because you must include harmonically bounded candidates in the candidate set.
  ➢ We lose a luxury that we had in classical OT analysis.

28. Implication II

• It becomes empirically important whether harmonically bounded candidates win in real life.
• I think they can be found in:
  ➢ Phonotactics (the Markedness-only approach)
  ➢ Metrics (see Hayes and Moore-Cantwell 2012 *Phonology*, paper with Russ Schuh under revision)
  ➢ Syntax-phonology interface: multiple phrasings from one syntactic structure.
• Harmonic bounding currently has Mom-and-apple-pie status (restrictiveness, ease of analysis) and it will take a lot of empirical argument for it to lose this status.