Embedding Grammar in a Quantitative Framework:  
Case Studies from Phonology and Metrics

Class 2: Frequency Matching in Hungarian/More on Models

1. Last time
   - Gradience in general
   - Law of Frequency Matching—with the challenge of explaining it, and the deviations.
   - Basics of maxent — how to derive gradient outputs given constraint weights and violations

2. Today
   - A bigger case study of the Law of Frequency Matching — with one particular kind of deviation from it — conjectured to be UG.
   - Learning challenges for computational linguists
   - Other models of gradience

3. Suggested reading
   - Hayes and Londe (2006), on course website

   HUNGARIAN VOWEL HARMONY

4. Sources
     - Basic work establishing applicability of Law of Frequency Matching
   - Hayes/Zuraw/Siptár/Londe (submitted)
     - Further study emphasizing the UG problem (below)
   - Both on course website.

5. Background: the UG problem
   - UG = Universal Grammar
   - One long-standing research interest of generative linguistics is locate aspects of language that are grounded in human nature — at some level, genetically coded.
   - These aspects can be:
     - Characteristics specific to the language faculty (UG narrowly construed)
➢ Other innate properties of humans (cognitive, phonetic, etc.) that determine properties of language (UG broadly construed)

• The distinction is not crucial in this context.

6. Finding UG using just analysis is hard

• The usual method is to pursue typology: one formulates grammars for many languages, all pointing toward the same underlying theory (e.g. Hayes 1995).

• However: language typology may reflect factors other than UG; notably factors of language change


7. UG: the experimental program

• There has recently been a great upwelling of interest in experiments addressing UG.


8. Ernestus and Baayen (2003—course website): “Predicting the unpredictable”

• Not a UG experiment, but relevant here.

• Dutch has classical Final Devoicing, as in

➢ [−sonorant] → [−voice] / ___ ]word

<table>
<thead>
<tr>
<th>Final</th>
<th>Before vowel</th>
</tr>
</thead>
<tbody>
<tr>
<td>[v?r??it]</td>
<td>‘widen’</td>
</tr>
<tr>
<td>[v?r??it]</td>
<td>‘reproach’</td>
</tr>
</tbody>
</table>


• The alternation patterns are unevenly distributed in the Dutch lexicon:

➢ many cases of [p#] ~ [pV], few of [p#] ~ [bV]

➢ few cases of [f#] ~ [fV], many of [f#] ~ [vV]

10. Ernestus and Baayen (2003): the idea behind the study

• Could Dutch speakers use a knowledge of the statistical regularities in the lexicon to “undo Final Devoicing”, guessing the suffixed form from the isolation form for novel stems?

➢ This would be “predicting the unpredictable”.

- Give subjects imaginary base forms, ask them to identify the suffixed form.
  - Hear \[k d\up{up}\] “I daup”, reply with *dauben* or *daupen*
  - Hear \[k t\up{f}\] “I taf”, reply with *tafen* or *taven*


- Many cases of \([p\#] \sim [pV]\), few of \([p\#] \sim [bV]\), so most (but not all) subjects prefer *daup* ~ *daupen*
- Few cases of \([f\#] \sim [fV]\), many of \([f\#] \sim [vV]\), so most (but not all) subjects prefer *taf* ~ *taven*

13. Becker, Nevins, and Ketrez (ms.—course website)

- A study similar to Ernestus and Baayen’s, but for Turkish, which also has “undoable” Final Devoicing.
  - many cases of \([p\#] \sim [bV]\), few of \([p\#] \sim [pV]\) (Dutch has the opposite disparity.)
  - few cases of \([t\#] \sim [dV]\), many of \([t\#] \sim [tV]\)
- New element: they checked the lexicon for VC environments. Examples:
  - Voicing alternation is more common after *high vowels*
  - \([t?] \sim [d?]\) alternation is more common after *back vowels*

14. Becker et al.’s results

- C-internal environments, such as place of articulation: Turkish speakers acted just like Dutch speakers, obeying the Law of Frequency Matching.
- Environments based on preceding vowel: null result
  - no frequency matching
  - in fact, no evidence that speakers are aware of the lexical pattern at all

15. Becker et al.’s interpretation

- They claim a new kind of argument for UG.
- The information needed to pass the Wug test with VC environments was made available to Turkish speakers in childhood.
- But they cannot pass this test for these environments—why not?
- Reason: UG isn’t good enough to detect the crucial generalizations here.
- Specifically: roughly, they claim that in UG, the only permitted vowel-consonant interactions are those involving a *shared feature*, like nasality or backness.

16. The general prediction made under Becker et al.’s approach

- The Law of Frequency Matching will hold true only for those phonological patterns that can be expressed in UG.
• For others, language learners are at a loss—there is no learning without UG help.

17. I’m skeptical of this view

• Earlier research gives strong evidence for productive phonological patterns that are totally arbitrary.
• Examples:
  ➢ Yidi? productively epenthesizes [u] after nasals, extending the pattern to new stems (Hayes 1999)
  ➢ English uses the regular past ending after every verb ending in a voiceless fricative; and wug-testing shows this is productive (Albright and Hayes 2003)

18. What is really going on: a conjecture

• People do have a strong (but not unlimited) inductive learning capacity for unnatural phonology.
• But there is a bias (Wilson 2006, Cognitive Science) for grammars that obey principles of UG.
  ➢ Wilson has formalized bias with maxent. More on this below.
• Becker et al.’s experiment was sensitive enough to pick up the UG-based patterns (like for place of articulation), but not the arbitrary VC ones.
• My colleagues and I haven’t worked on Turkish, but can offer evidence from another language.

19. Some Hungarian wug-testing

• Two papers:
  ➢ Hayes and Londe (2006)—establish frequency matching; other issues in phonology.
  ➢ Hayes, Zuraw, Siptár, and Londe (submitted): extend this result, take on the UG issue

20. The basics of Hungarian vowel harmony

• Hungarian vowel inventory:
  ➢ Back [u, ū, o, ơ;, ?, a:] “B”
  ➢ Front rounded [y, ų, ơ, ơ:] “F”
  ➢ “Neutral” [i, ĩ, e:, ?] “N”
• Most suffixes alternate, agreeing in backness with the nearby vowels of the stem.
• We’ll deal just with the dative suffix: [-n?k] ~ [-n?k].

21. The vowel harmony generalizations

• For a nice overview consult Siptár and Törkenczy (2000).
• If the closest stem vowel is back, then back suffixes:
• If the closest stem vowel is front rounded, then front suffixes:
  ➢ [ɔb?k-n?k] ‘window-dat.’
  ➢ [glyko¿-n?k] ‘glucose-dat.’

22. The vowel harmony generalizations (cont.)

• If all neutral (front unrounded), then generally front
  ➢ [k?rt-n?k] ‘garden-dat.’
  ➢ [tsi?m-n?k] ‘address-dat.’

23. Hungarian vowel harmony: the zones of variation

• If a back vowel is followed by one or more front unrounded vowels (…BN, …BNN), the lexicon takes over!
• Harmony is unpredictable, and the behavior of every stem must be memorized.
• But there are quantitative lexical generalizations, just like in Dutch and Turkish.

24. Hungarian vowel harmony in the zones of variation: quantitative generalizations

• Source of the counts
  ➢ Hayes and Londe’s (2006) study, based on Googling thousands of stems to obtain user frequencies.

25. Height effect

• … Back + [i, i?] stems mostly take back suffixes;
• … Back + [e?] stems take front more often;
• … Back + [?] stems (? is low) usually take front suffixes.
26. Count effect

- Stems with two neutral vowels after a back take front suffixes more often than stems with just one.

27. Hayes/Londe Wug test: speakers match lexical frequencies
28. Hungarian so far

- The vowel harmony system obeys a number of generalization, including statistical generalizations (Height effect, Count effect) governing the zones of variation.
- In wug testing, speakers respond to the statistical generalizations, in accord with the Law of Frequency Matching.

29. Hayes and Londe’s analysis I: UG principles assumed

- UG favors assimilation of single features (cf. Becker et al.)
- UG favors local triggers over distal (cf. [glyko2z-n?k])
- “Spread bad vowel” (Kaun 1995) Perceptually weak frontness → strong frontness harmony trigger (they need more help, hence trigger harmony more)
  - [y, y?, ø, ø?], with backness perceptually obscured by rounding, are stronger triggers.
  - Lower front vowels, which aren’t as front as higher ones, are stronger triggers.

30. Hayes and Londe’s analysis II: constraints

- AGREE(back) with local back
- AGREE(back) with back
- AGREE(back) with local low front
- AGREE(back) with local nonhigh front
- AGREE(back) with local front
- AGREE(back) with local front + front
- AGREE(back) with front rounded
- AGREE(back) with local front rounded

31. The constraints reflect the UG theory

- Every constraint is an AGREE() constraint for backness.
- Those which are restricted to particular classes single out “bad vowels”, in Kaun’s sense.
- 6/8 constraints are restricted to local environment; i.e. penalize only disagreement with the local stem vowel.

32. Finding the right weights

- Hayes/Londe did a maxent analysis (as well as others) and got a good match to the observed data.
- Training data: use the Google data corpus frequencies for the training lexicon
33. Matchup to the wug test results, using this grammar

![Graph showing wug test results and model predictions]

34. Local summary

- Hayes and Londe offer a first-pass analysis of the Hungarian system, which
  - uses only UG-based constraints
  - is tuned by training on a data corpus
  - correctly predicts the behavior of Hungarian speakers when wug-tested—they obey the Law of Frequency Matching

35. Hungarian beyond UG: are there unnatural constraints?

- Hayes and Zuraw did an intensive search of the Hungarian lexical data (Excel + search programs), looking for consonant environments that favor front or back suffixes, within the zones of variation.
- These turned out to exist! We picked the best four.

36. Four unnatural vowel harmony constraints

- Prefer front suffixes when the stem ends in a bilabial noncontinuant ([p, b, m])
- Prefer front suffixes when the stem ends in a sibilant ([s, z, ?, ts, t?, d?])
- Prefer front suffixes when the stem ends in a coronal sonorant ([n, ?, l, r])
- Prefer front suffixes when the stem ends in two consonants
  - All have statistically significant effects in the lexicon
  - None has a particularly plausible UG basis.

37. One more unnatural vowel harmony constraint

- All-N words are (weakly) a zone of variation, too, since a few of them take idiosyncratic back suffixes. Examples:
  - [hiʔl-n?k] ‘bridge-dat.’
  - [ʔiʔp-n?k] ‘whistle-dat.’
38. Question: do Hungarian speakers internalize the unnatural constraints?

- Becker et al. would presumably predict, “no”.
  - This is a vowel-consonant interaction not based on shared features.
- We can find out by doing a new wug test.

39. Design of the new wug test

- All our wug stems were from the zones of variation: BN, BNN, N.
- To increase sensitivity, we tested 1703 wug stems — each subject got a separate batch, thus reducing the chance of unwanted factors about particular stems playing a role.
- Also to increase sensitivity, we used many (131) consultants

40. Choice of wug words

- They were designed to test the unnatural constraints, and otherwise statistically resemble Hungarian words in every possible respect.
- Each consultant got 13 words, as for example in the following batch.

<table>
<thead>
<tr>
<th>Number</th>
<th>bilabial</th>
<th>coronal sonorant</th>
<th>sibilant</th>
<th>CC</th>
<th>example</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>[kóåle]</td>
</tr>
<tr>
<td>2</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>[prítt]</td>
</tr>
<tr>
<td>2</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>[t??nd?nd??]</td>
</tr>
<tr>
<td>1</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>[s?lha?te???]</td>
</tr>
<tr>
<td>2</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>[n?n]</td>
</tr>
<tr>
<td>1</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>[vur?ld?m]</td>
</tr>
<tr>
<td>2</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>[ha?k?m]</td>
</tr>
<tr>
<td>1</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>[ke?bb]</td>
</tr>
</tbody>
</table>

41. Subject recruitment

- To obtain authentic data, we ran our test on the Web, using Google Ad Words to recruit subjects living in Hungary.
• This yielded 131 subjects, after discarding non-Hungarians, under-18’s, non-finishers.

42. Where our subjects came from

Source: www.google.com/analytics/

43. Sample test page

(original version entirely in Hungarian)

Hálupem

Hálupem was a goddess worshipped by the early pagan Hungarians. It is believed that Hálupem was the goddess of weaving. Not just the Hungarians but also neighboring peoples celebrated ___-dat.’s divine powers. Please check one:

☐ Hálupennak
☐ Hálupennek

Please rate each one:

Hálupennak:

<table>
<thead>
<tr>
<th>worst</th>
<th>best</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5 6 7</td>
<td>○ ○ ○ ○ ○ ○ ○</td>
</tr>
</tbody>
</table>

Hálupennek:

<table>
<thead>
<tr>
<th>worst</th>
<th>best</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5 6 7</td>
<td>○ ○ ○ ○ ○ ○ ○</td>
</tr>
</tbody>
</table>
44. Results I: Hayes/Londe experiment is replicated

- Note “depolarization” effect—probably due to modeling bad forms to the subjects (ca. Albright and Hayes 2003).

45. Results II: 4/5 of the unnatural constraints had a significant effect

- Test: logistic regression; probability that the environment has no effect on outcome
  
<table>
<thead>
<tr>
<th>Environment</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monosyllable with [i:]</td>
<td>0.241</td>
</tr>
<tr>
<td>Final bilabial</td>
<td>0.000</td>
</tr>
<tr>
<td>Final coronal sonorant</td>
<td>0.016</td>
</tr>
<tr>
<td>Final sibilant</td>
<td>0.046</td>
</tr>
<tr>
<td>Final CC</td>
<td>0.000</td>
</tr>
</tbody>
</table>

- We suspect that the [i?] environment is also significant, but our test was not designed to check it and lacked enough forms.

46. Other statistical results

- The same test showed massive significance for all of the natural constraints.
- The effect size is smaller for the unnatural constraints, as expected.

47. Discussion

- Unlike Becker et al., we found a noticeable effect for unnatural constraints.
- Suggested conclusion: people do learn unnatural phonology, but
  
  ➢ it’s harder for them to learn
  
  ➢ it’s harder for us to detect — perhaps only experiments with many subjects and items will suffice

48. Maxent modeling: Goal

- Elucidate how Hungarian speakers, exposed to lexical data, learn the unnatural constraints and use them in forming their linguistic intuitions.
49. The basis for learning
   - 13 constraints, as given above
     - 8 natural, from Hayes/Londe
     - 5 unnatural, based on our discovered unnatural environments

50. Grammar L: train the weights against the lexicon
   - We used the same Google frequency data that Hayes and Londe used.
   - To assign weights, we used an early version of the Maxent Grammar Tool (course website).
   - Weights obtained:

<table>
<thead>
<tr>
<th>Constraint</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGREE(back) with local back</td>
<td>4.00</td>
</tr>
<tr>
<td>AGREE(back) with back</td>
<td>5.35</td>
</tr>
<tr>
<td>AGREE(back) with local low front</td>
<td>2.99</td>
</tr>
<tr>
<td>AGREE(back) with local nonhigh front</td>
<td>1.48</td>
</tr>
<tr>
<td>AGREE(back) with local front</td>
<td>1.64</td>
</tr>
<tr>
<td>AGREE(back) with local front + front</td>
<td>4.05</td>
</tr>
<tr>
<td>AGREE(back) with front rounded</td>
<td>1.72</td>
</tr>
<tr>
<td>AGREE(back) with local front rounded</td>
<td>3.74</td>
</tr>
<tr>
<td>Front suffix if final bilabial consonant</td>
<td>2.46</td>
</tr>
<tr>
<td>Front suffix if final coronal sonorant</td>
<td>1.08</td>
</tr>
<tr>
<td>Front suffix if final sibilant</td>
<td>0.91</td>
</tr>
<tr>
<td>Front suffix if final CC</td>
<td>1.75</td>
</tr>
<tr>
<td>Back suffixes if monosyllable with [i]</td>
<td>2.37</td>
</tr>
</tbody>
</table>

51. Basic performance of Grammar L (correlations)
   - Basic check: its predictions correlate well with the frequencies of the data from which it was trained: \( r = .992 \)
   - To check against the Wug test, we used a preference score: i.e. subject’s [-n?k] preference minus [-n?k] preference.
     - Correlation is not as good, but still substantial: \( r = .546 \) — confirming the Law of Frequency Matching
   - The unnatural constraints help: without them, we get a correlation of only \( r = .521 \)

52. What are native intuitions in the same terms? Grammar W
   - Grammar W is obtained by using the wug test data as the basis for constraint weighting.
   - This is not a learning simulation (children are not told their parents’ intuitions).
   - Rather, it is an assessment of the importance of the constraints in forming adult native speaker intuitions.
53. **Weight comparison: Grammar L vs. Grammar W**

<table>
<thead>
<tr>
<th>Natural constraints:</th>
<th>W</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGREE(back, local)</td>
<td>NA</td>
<td>4.00</td>
</tr>
<tr>
<td>AGREE(back, nonlocal)</td>
<td>3.58</td>
<td>5.35</td>
</tr>
<tr>
<td>AGREE(low front, local)</td>
<td>1.05</td>
<td>2.99</td>
</tr>
<tr>
<td>AGREE(non-high front, local)</td>
<td>.95</td>
<td>1.48</td>
</tr>
<tr>
<td>AGREE(front, local)</td>
<td>2.22</td>
<td>1.64</td>
</tr>
<tr>
<td>AGREE(double front, local)</td>
<td>1.94</td>
<td>4.05</td>
</tr>
<tr>
<td>AGREE(front rounded, nonlocal)</td>
<td>NA</td>
<td>1.72</td>
</tr>
<tr>
<td>AGREE(front rounded, local)</td>
<td>NA</td>
<td>3.74</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unnatural constraints:</th>
<th>W</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>USE FRONT / bilabial ___</td>
<td>1.04</td>
<td>2.46</td>
</tr>
<tr>
<td>USE FRONT / [+cor,+son] ___</td>
<td>.43</td>
<td>1.08</td>
</tr>
<tr>
<td>USE FRONT / sibilant ___</td>
<td>.37</td>
<td>.91</td>
</tr>
<tr>
<td>USE FRONT / CC ___</td>
<td>.69</td>
<td>1.75</td>
</tr>
<tr>
<td>USE BACK / [C₀iC₀] ___</td>
<td>.80</td>
<td>2.37</td>
</tr>
</tbody>
</table>

54. **Analyzing the weight comparison**

- Observe: a bigger “hit” (.34 vs. .63) for the unnaturals.

55. **What’s going on?**

- Depolarization: weights lower overall.
- Natural vs. unnatural: language learners can pick up unnatural environments, but give them less credence.
- Further details (constraint by constraint) can be accounted for, perhaps, with a simplicity bias.
56. More modeling results

- The best fitting model we can find uses two parameters: overall weakening, and unnaturalness bias. These are used to multiply all the constraints, and just the unnatural constraints (values: .56, .70)
- This achieves a correlation of $r = .569$ to the wug test data.

57. In progress

- We think we can show that the alleged “weakness” of the unnatural constraints can be demonstrated statistically—we’re doing a Monte Carlo simulation to show this.

58. Summary

- Both the earlier Hayes/Londe experiment and our subsequent web experiment found confirmation for the Law of Frequency Matching.
- We got opposite results from Becker et al., concerning whether unnatural environments can have effects in phonology, and don’t know why (though we have guesses).
- But the unnatural constraint seem to be “weaker”, tantalizing evidence that there may be UG effects in learning.

TWO LEARNABILITY PUZZLES ARISING FROM THIS WORK

59. Learnability puzzles I: handling the listed items

- The maxent grammars given were “tailor made”—learn from lexicon, and test on wug data.
- But real life is more complicated—more tasks for the grammar to do.
- Forms in the zones of variation get the outcome specified in the lexicon.
- I think the simplest and best supported theory for this is Zuraw’s: they’re listed.
- But suppose they are listed. We must make sure that the listed form gets used, and not the one created by the phonological grammar.
- So: the weight of USE LISTED (Zuraw 2000) is very high.
- I’ve been able to get results vaguely in the right direction, but not what is really needed: reliable rendering of listed forms, reliable frequency-matching in novel forms.

60. Learnability puzzles II: how to hobble?

- We don’t know how to “hobble” a constraint for its unnaturalness.

61. One potential way to hobble: the Gaussian prior

- He biases his constraints using a Gaussian prior, with lower sigma for a priori less likely constraints.
• Gaussian prior: maximize probability of learning data, but adding a penalty for positive weights:

\[
\log P_{\mu, \sigma}(y|x) = \sum_{i=1}^{m} \frac{\left( w_i - \mu_i \right)^2}{2\sigma_i^2}
\]

➢ If \( \mu \) is zero, then a constraint gets punished for having a large weight, especially when \( \sigma \) is small.

62. But hobbling with a constraint-specific prior is very tricky

• If you hobble, other constraints will get different values too.
• E.g., two constraints, one hobbled, 50/50 learning data: the other constraint weakens itself so as still to derive 50/50 outputs.
• We suspect that Wilson succeeded because of particular, unusual properties of his grammar; we cannot get the same results in ours.

OTHER CONSTRAINT-BASED THEORIES OF GRADIENCE

63. OT with free-variation strata

• Anttila (1997a, 1997b)
• Group the constraints into strata; rank the strata, but rank at random within strata.
• This predicts a specific distribution of outputs.
• Very tightly constrained model; our Hungarian is an example it seems unable to deal with.

64. Stochastic OT

• Invented by Paul Boersma (1997); applied to phonology by Boersma and Hayes (2001).
• Give every constraint a “ranking value”.
• When you run the grammar, jiggle the weights by adding to each ranking value a small random quantity. Then sort them and apply good-old OT to the result.

65. The learnability situation for stochastic OT

• Boersma invented an algorithm (“Gradual Learning Algorithm”) for stochastic OT.
• It works pretty well for many simulations—though without maxent’s uncanny accuracy.
• Behaves very strangely for others (my experience)
• and (ouch!) was found to fail to find the solution in a well-defined class of cases—Pater (2008), course web site

- Magri (ms.), course web site, has a beautiful demonstration of why GLA fails: sometimes the right answer isn’t even in its search space! (= grammars obtainable by legal ranking values adjustments.
- Magri has a better GLA, which he proves to converge, but only for non-stochastic grammar.

66. **Noisy Harmonic Grammar**

- Paper by Boersma and Pater (course web site).
- This is like the simple Harmonic Grammar described last time (lowest penalty score wins), but as with Stochastic OT you add a bit of noise to each constraint weight when you “apply” the grammar.

67. **The learnability situation for stochastic OT**

- Same as for stochastic OT: there is a learnability proof, but only for the non-stochastic applications

68. **Where maxent differs sharply from these models**

- **Harmonically bounded** candidates can semi-win (i.e. have more than zero probability)
- A candidate is harmonically bounded if some other candidate has a strict subset of its violations.
- Scholars differ in whether harmonically bounded candidates should ever win. Keller and Asudeh (*Linguistic Inquiry* 2002) thinks they should; I’ve found slightly better performance in textsetting.\(^1\) I’d say not letting them win is the majority current view.

69. **Model-shopping: my own feelings**

- Once burned, twice shy, re. using algorithms that don’t have a convergence proof.
- Some empirical worries re.
  - Constraint ganging (all versions of Harmonic Grammar)
  - Harmonically bounded semi-winners (maxent)

---

A QUICK OVERVIEW OF HOW LEARNING IN MAXENT WORKS

70. Source

- This discussion follows the attempted layman’s explanation in Hayes and Wilson (2008) (course website).

71. Core idea: “Objective function”

- Defines the “goal” of learning.
- This is separated from the (varying) computational algorithms can be used to achieve it.
- *Maximize the predicted probability of the observed forms*
- hence, minimizes the predicted probability of the unobserved forms
- Predicted probability of observed forms is quite calculable: calculate each one as given last time, then multiply them all together.

72. Metaphor: the objective function is a mountain

- If we have just two constraints, let North-South be the axis for Constraint 1’s weight, and East-West be the axis for Constraint 2’s weight, and height be the predicted probability of the observed data under any weight assignment.
- Climb the mountain, and you will be standing at the point of optimum weights.

73. Two beautiful theorems

- The mountain has but one peak (=is convex; has no local maxima)
- The slope along any axis (if height expressed as a log) is *Observed Violations – Expected Violations*, a calculable quantity.
- So you can always reach the top, simply by persistently climbing uphill.
- This may sound trivial but remember that the mountain actually exists in \( n \)-dimensional space, where \( n \) is the number of constraints.

74. The rest is implementation

- Ascending gradients efficiently is a popular challenge for computer scientists; both Goldwater and Johnson (2003) and the Maxent Grammar Tool adopt the “Conjugate Gradient” algorithm.

75. Next time

Phonological well-formedness (*blick - ?bloick - *bnick*) and how to predict it with maxent.