Syllabus: Quantitative and Computational Phonology

Course Description

In the grammar architecture of classical Optimality Theory, constraints are ranked and the grammar generates exactly one winner per input. Phonologists later proposed constraint-based grammar frameworks that employ mathematics in some way to generate multiple outputs, assigning probabilities to all members of the candidate set. Such quantitative grammars open up new research possibilities for constraint-based phonology, particularly in modeling free variation, gradient intuitions, and phonological learning.

However, much of the focus of the course actually will be rather traditional phonological study: we want to learn about the phonological grammar internalized by the native speaker by carrying out analysis, in the course of which we develop a theory that explicates the nature of phonological grammars. Part of what’s involved is making analysis more thorough and accurate—which implies various techniques: data corpora, machine-searching of constraint violations, stochastic modeling to capture variation in the data, statistical testing. So a major theme of the course is to teach the tools one would want to have in order to do effective phonological analysis by the standards of 2013.

The course will emphasize learning by doing. Participants will use software tools that embody the theories at hand and will examine and model data from a variety of digital corpora. The course will not cover computational phonology per se, but it will cover enough computation to give participants a good understanding of the tools they are using.

Requirements

- If you want course credit, then do a small project, written up as a short paper.
  - If you elect this option, you are also required to meet with me at least once to talk about what you are doing.
- There will be exercises in which you model data using software. You can hand these in to me if you wish and I will look at and comment on your work. You should do the exercises if you’re going to do a project.
- There will be readings, which you should read if you want to do a project.

Talking to me

- I will schedule office hours and can also make appointments (bhayes@humnet.ucla.edu)

Course web site

- The web site for this course is on the UCLA web server:
  > www.linguistics.ucla.edu/people/hayes/LSA2013
Software to be used

- Alas a couple of the software programs are Windows-only.
- See course website for information on workarounds for Mac users.

<table>
<thead>
<tr>
<th>Software</th>
<th>Platform</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>PhonologySearch</td>
<td>Windows</td>
<td><a href="http://www.linguistics.ucla.edu/people/hayes/EnglishPhonologySearch/">www.linguistics.ucla.edu/people/hayes/EnglishPhonologySearch/</a></td>
</tr>
<tr>
<td>OTSoft</td>
<td>Windows</td>
<td><a href="http://www.linguistics.ucla.edu/people/hayes/otsoft/">www.linguistics.ucla.edu/people/hayes/otsoft/</a></td>
</tr>
<tr>
<td>Maxent Grammar Tool</td>
<td>all platforms</td>
<td><a href="http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/">www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/</a></td>
</tr>
<tr>
<td>R</td>
<td>all platforms</td>
<td><a href="http://www.r-project.org">www.r-project.org</a></td>
</tr>
</tbody>
</table>

Tentative list of topics

**Tuesday, June 25: Basics**
- The role of computation in phonology
- Types of variation: sociolinguistic, lexical, subject-response
- The Law of Frequency Matching and its implications
- Plunging into modeling with a real example

**Thursday, June 27: Models**
- Constraint-based models of variation: freely-ranked strata, Stochastic OT, NHG, maxent
- Their affiliated learning algorithms
- Sample applications
- Exercise: the same data modeled in different frameworks

**Tuesday, July 2: Corpora**
- Assessing constraint violations in a corpus: Excel, searching software, other
- Differential phonotactics and its applications (finding environments; lexical strata)

**Friday, July 5 (note special day): Model evaluation**
- Logistic regression: maxent’s convenient little brother
- Logistic regression in R
- Model evaluation with statistical testing — how and why

**Tuesday, July 9: Bias — the P-map and its formalization**
- Bias as a gradient model of Universal Grammar
- Types of bias: conservativeness, simplicity, naturalness
- Modeling bias using maxent grammars (J. White)

**Thursday, July 9: More on bias: conservativity**
- The justification for keeping weights low
- Empirical patterns that result (Ryan, Martin)

**Tuesday, July 16: Style, frequency, and knobs**
- Empirically-observed effects
- Attaching “knobs” to models to interface with societal knowledge and frequency

**Tuesday, July 18: Testing the models**
- Case studies: constraint-family interaction in Tagalog, French, and Hungarian