

Homework #1: Modeling of Medial Clusters in Maxent

Due Monday April 16 in class

The goal here is to do an analysis of the legal medial clusters of some language, using the rigor provided by a formalized maxent grammar (i.e., you end up responsible for every single logical possibility implied by the phonemic system). If it turns out that you can successfully model cluster frequencies as well, that's great.

1. Find a language on the Internet or other electronic resource where you can download a phonetic dictionary (it doesn't have to be IPA, just some text material you can manipulate and trust for pronunciation). A language you already know something about might be a good idea. The language must have medial CC clusters.

2. Find what you can on the phonotactics of this language in the library or on line.

If you're handy at coding and would like to do all or part of the assignment by writing a program, feel free; just make sure it's a full replication of what I describe below. This holds for any other technique you might want to use. Directions continue in spreadsheet mode.

3. Using manipulations of the kind taught in Class #2, gather your data into a spreadsheet, whose rows represent all logically possible two-consonant medial clusters. Let Columns 1 and 2 represent consonants, column 3 frequency.

4. Add to your spreadsheet new columns, holding feature values and (deduced from the feature values) constraint violations for constraint you obtain from your reference source or from your own thinking.

5. Do the normal columns (Harmony, eHarmony, Z, p, ln(p), Likelihood) for doing maxent on a spreadsheet. My own spreadsheet is on the course website as a model in case this is useful. Assume the Markedness-only theory of phonotactics, where there is one abstract input and the clusters are candidates. Find the best weights using the solver.

6. Convert the observed frequencies into percentages of the total and use the result to make a scattergram of Observed and Expected for your analysis.

7. For constraints with low weights, test their statistical significance with the likelihood ratio test (see sample spreadsheet for how).

8. Check out whether the Syllable Contact Law helps: give every segment a sensible integer sonority value, subtract to get violations, and incorporate into the system. If weight is low, be sure to test. How does this come out?

9. Write up a reasonably terse prose description of what you did and what you found out about the phonology of the language you covered. Submit in hard copy, email me your spreadsheet.

Please feel free to consult with me on any point that arises.