Homework 4: Maxent Phonotactics

Due in class February 16th.

1. **Find an electronic dictionary**
   - You want one with phonetic entries, or with a phonemic orthography.
   - Wiktionary seems to be getting better …

2. **Find relevant pairs and count them**
   - Two choices:
     - You can do medial consonant clusters: All CC / V ___V
     - You can do all consecutive vowel pairs (separated by zero or more C).\(^1\)
     - Something else, if you clear it with me first.

3. **Ways to find all vowels**
   - In a word processor, turn every consonant into a tab.
   - Put a tab after every letter.
   - Turn every tab sequence into a tab (repeat)
   - Paste into spreadsheet and keep first two columns.

4. **Form a GEN**
   - All logical possibilities, given the vowel or consonant inventory.
     - This will be \(n^2\) if there are \(n\) consonants/vowels.
   - Give frequencies for each, including zero for absent ones.

5. **Add the frequencies of existing items**
   - At worst, sort your spreadsheet of data and count by hand.
   - You can use the COUNTIF() function of Excel.

6. **Form baseline MaxEnt model in spreadsheet**
   - The key idea here is to give precedent to any explanation other then sequencing constraints, so use a crude, detailed theory.
   - Specifically, penalize each vowel or consonant in each position (2\(n\) constraints, where \(n\) is the number of vowels or consonants.

\(^1\) For simplicity I would suggest just doing disyllabic words, or (if the phonology fits) the first two vowels of every polysyllabic word.
7. **Try plausible new constraints to improve your model**
   - A method that will help you find them is to arrange your counts in an $n \times n$ matrix.
   - Test each added constraint with likelihood ratio test, only add it to the grammar if it does well.
   - Plot your model’s performance in a scatterplot.
   - Look for outliers in an attempt to improve your model.

8. **What kind of constraints**
   - When it’s your model, you want to be principled and elegant
   - See typology below
   - Feel free to use negative weights (licenses) if you like.

9. **Using features**
   - Many constraints will be easier to state if you add feature columns.

10. **Write up what you did and what you learned**
    - Describe your procedure
    - Insert suitable tables and graphs to back up your exposition.
    - Include results of Likelihood Ratio Test
    - Please email me the spreadsheet you used.

11. **Help**
    - I will post my Turkish spreadsheet.
    - Come see me for advice if you like

12. **Typology of consonant sequences**
    - Place agreement, particular for nasal + C
    - No geminates
    - Coronals after noncoronals
    - Fricatives after stops
    - No laterals after coronals
    - Various generalizations about sonority (hard to implement without syllable structure)

13. **Typology of vowel sequences**
    - Often richer inventory in initial syllable, or under stress.
- Harmony: backness, rounding, high (but seldom low!), tenseness/ATR
- Sometimes a tendency to favor identical vowel sequences