Final assignment: discovering generalizations

Our final topic in this class has been the use of large databases of words to discover phonologically and morphologically relevant generalizations. Your final assignment is to employ these techniques to test some hypotheses, of your choosing, about English or Spanish.

In the course of doing this assignment, you will need to do several things:

1. Pick a phenomenon to try to explain.
   - Location of stress in Spanish (recommended!)
   - Verb class membership (-ar, -er or -ir)
   - Tendencies of syllable shapes in English
   - Differences between high and low frequency words

2. Hypothesize some factors that might be involved
   - Particular segment features (place, manner, voicing, etc.)
   - Syllable structure (clusters, syllable weight, etc.)
   - Combinations of multiple features, or multiple segments

3. Use Excel to calculate your hypothesized factors
   - Augment the feature table, and calculate feature values of a target segment, or...
   - Determine the syllabic structure of relevant roots, or...
   - Or some other property (consult me if you aren’t sure how to go about calculating something that you’re curious about!)

4. Test whether your hypothesized factor actually makes a difference in predicting the pattern

If you choose to work on English or Spanish, you can use the same Excel workbooks that we have been using in class (SpanishVerbs.xls from last week, EnglishMonosyllables.xls or SpanishStress.xls from this week). These files have a lot of information pre-compiled for you (such as phonemic transcription, decomposition of onsets and codas into individual segments, and so on). In the course of testing hypotheses, you should add at least two new columns in which you calculate new potential factors.

Write-up: in addition to submitting to me your Excel file (by email), you will give me a write-up, in which you explain: (1) What phenomenon are you trying to explain, and what factors do you hypothesize might be involved in predicting it? (why?) (2) How did you calculate the relevant factors? what obstacles did you encounter, or what made it difficult to calculate these factors perfectly accurately? (3) To what extent did the factors turn out to be relevant? (i.e., are your hypotheses supported?) Be sure to include a rough estimate of expected values, in addition to the actual (observed) counts.

Be sure to consult me if you are having difficulty formulating hypotheses that are feasible to test with these databases, or if you would like to test a hypothesis but can’t figure out how to use Excel to calculate the relevant factor (such as syllable weight). Remember that the goal is not necessarily to find absolutely true generalizations; statistical tendencies are enough!

---

1If you prefer to look at a language other than English or Spanish, there are some other possibilities; please let me know immediately if you prefer a different language, and we can take a look to see whether a suitable database is available.