Course Business

1) Friday's Workshop (July 29th)
   a) Auditors welcome of course
   b) 2 Groups of 4
   c) 5-10 minute spiels: hypothesis, methodology, where I’m stuck
   d) 20 minutes of workshop time per person

2) Tuesday's Presentations (August 2nd)
   a) Auditors welcome of course
   b) 10-minutes strictly timed plus 2-3 minutes for questions
   c) Any media is acceptable:
      i) If handouts: please email us the relevant file as .pdf the night before so we can arrange for copies
      ii) If PowerPoint: you will need to use your own laptop (no computer in room); please confirm that the connections will work before next Tuesday

3) Paper
   a) Due date: Wednesday, August 5th @ 5:00 PM (emailed pdf file)
   b) Format
      i) An experiment proposal related to issues we have discussed
      ii) Less than 8pp, double-spaced, 1-inch margins, no smaller than 12 point true-type font
      iii) Prefer APA style for Methods Section at the very least; sample papers and general overview can be found here: http://www.apastyle.org/index.aspx

Overview

4) Last Time
   a) Hammond
   b) Ohala
   c) Methodological Points
      i) Corpora
      ii) Spontaneous Errors

5) Today
   a) PCFGs (from last time)
   b) Berent et al 07 (new)
   c) Daland et al 11 (new)
   d) Summary & General Discussion
PCFGs

6) What are they? Probabilistic context-free grammars: phrase structure rules with probabilities, e.g.
   a) $A \rightarrow B \ C$ and $p(A \rightarrow B \ C)$
   b) Probability distribution requirement
   
   Probabilities of all rules expanding some symbol A in a grammar must sum to 1

7) A very simple syllable-based phrase-structure grammar
   a) Word
      i) $w \rightarrow s$
      ii) $w \rightarrow s \ w$
   b) Syllable
      i) $s \rightarrow o \ r$
   c) Onset
      i) $o \rightarrow \emptyset$
      ii) $o \rightarrow obs$
      iii) $o \rightarrow liq$
      iv) $o \rightarrow nas$
   d) Rhyme
      i) $r \rightarrow V$
      ii) $r \rightarrow V \ obs$
      iii) $r \rightarrow V \ liq$
      iv) $r \rightarrow V \ nas$

8) Let’s test how consonant sonority affects medial syllabification
   a) VC.V vs. V.CV

9) Train the grammar on monosyllables only
   a) Same source data as last time (in source file on website: "newdic")
   b) All R code on website

10) Onset counts (how many times do each of these occur?)
    a) $o \rightarrow \emptyset$ 87512
    b) $o \rightarrow obs$ 116665
    c) $o \rightarrow liq$ 12681
    d) $o \rightarrow nas$ 27639

11) Onset probabilities (divide counts by the total)
    a) $p(o \rightarrow \emptyset) = .36 = 87512/(87512+116665+12681+27639)$
    b) $p(o \rightarrow obs) = .48 = 116665/(87512+116665+12681+27639)$
    c) $p(o \rightarrow liq) = .05 = 12681/(87512+116665+12681+27639)$
    d) $p(o \rightarrow nas) = .11 = 27639/(87512+116665+12681+27639)$
12) Rhyme counts (how many times do each of these occur?)
   a) $r \rightarrow V$ 126728
   b) $r \rightarrow V_{obs}$ 120301
   c) $r \rightarrow V_{liq}$ 46276
   d) $r \rightarrow V_{nas}$ 67350

13) Rhyme probabilities (divide counts by totals)
   a) $p(R \rightarrow V) = 0.35 = \frac{126728}{126728+120301+46276+67350}$
   b) $p(R \rightarrow V_{obs}) = 0.33 = \frac{120301}{126728+120301+46276+67350}$
   c) $p(R \rightarrow V_{liq}) = 0.13 = \frac{46276}{126728+120301+46276+67350}$
   d) $p(R \rightarrow V_{nas}) = 0.19 = \frac{67350}{126728+120301+46276+67350}$

14) V Obs V sequences
   a) $V_{0}.V = p(r \rightarrow V_{0}) \times p(o \rightarrow \emptyset)$
      $= 0.12 (41\%) = 0.33 \times 0.36$
   b) $V.V_{0} = p(r \rightarrow V) \times p(o \rightarrow \emptyset)$
      $= 0.17 (59\%) = 0.35 \times 0.48$

15) V Nas V sequences
   a) $V_{n}.V = p(r \rightarrow V_{n}) \times p(o \rightarrow \emptyset)$
      $= 0.07 (58\%) = 0.19 \times 0.36$
   b) $V.V_{n} = p(r \rightarrow V) \times p(o \rightarrow N)$
      $= 0.05 (42\%) = 0.35 \times 0.11$

16) V Liq V sequences
   a) $V_{l}.V = p(r \rightarrow V_{l}) \times p(o \rightarrow \emptyset)$
      $= 0.05 (71\%) = 0.13 \times 0.36$
   b) $V.V_{l} = p(r \rightarrow V) \times p(o \rightarrow L)$
      $= 0.02 (29\%) = 0.35 \times 0.05$

17) PCFG Conclusions
   a) A pcfg-trained on monosyllables gets the same results as Treiman et al.
   b) This does not explain why these patterns exist
   c) This does suggest that lexical distributions mirror judgments

Berent et al 07

18) Main point: to test whether speakers are sensitive to the well-formedness of structures not present in the native language
   a) As in Hammond, Ohala papers, focus is on sonority in non-native clusters, specifically the sonority profile of onsets
19) Six Experiments with English & Russian speakers
   a) Experiments 1-2: judge how many syllables in aurally-presented nonsense stimuli
   b) Experiments 3-4: identity task; are monosyllables identical to epenthetic versions?
   c) Experiments 5-6: priming task; do epenthetic versions identity prime?

20) Experiments 1-2: bnif vs. bdif vs. lbif: how many syllables?
   a) As markedness increases (as in lbif), more likely to be perceived as polysyllabic
   b) As markedness increases, more likely to be primed by disyllabic counterpart (lebif)
   c) Unlike English speakers, Russian speakers generally get monosyllables correct
      i) ...showing that the clusters are not “invariably imperceptible” and therefore not reducible to a phonetic explanation (must be grammatical)

21) Experiments 3-4: bnif=benif, etc: are these identical?
   a) For both, errors increase for more marked clusters

22) Experiments 5-6: Does bnif get identity priming with benif?
   a) More likely as markedness increases: suggests that lbif is treated as lebif
      i) But in Exp 5, epenthetic primes equally likely as monosyllables
   b) In Exp 6, skewed epenthetic primes with yeses
      i) English speakers can perceive difference when attention to phonetic detail thus encouraged

23) Final points
   a) Different tasks from previous [MH & DKO papers], so a more general result but manipulation of epenthesis consistent with DKO
   b) Unlike MH, don’t include 2 vs. 3 consonant onsets
   c) Gist: "simple" statistical experience won’t account for these results; some knowledge of universals is necessary
**Daland et al 11**

24) Main point: to show via statistical modeling that (23c) is wrong

25) Claim: sonority effect follows from feature-based statistical experience/learning plus something like syllable structure

26) First: judgment task for occurring, marginal, and non-occurring clusters
   a) Experimental results correlate with experience

27) Statistical learning (machine learning; ML) successful with features and structure only
   a) But ML can learn ANYTHING
   i) What if a language exhibited REVERSED sonority? ML techniques could learn that without problem.
   b) Where do the features come from? In ML, typically built in. What does that imply?

28) So where does the sonority pattern come from?
   a) ML plus features and SOME structure doesn’t tell us why sonority exhibits the patterns it does

29) Evolutionary phonology doesn’t answer the question either.

**Sum up**

30) Lots of effects of syllables in many different types of tasks
   a) There is a syllable
   b) Likely the same in perception as in production (cf. Berent et al with Ohala)

31) But what kind of syllable?
   a) Probabilistic grammar of some sort seems likely
   i) What sort? See son.txt file on website for pcfg examples

32) A probabilistic grammar of any sort will get
   a) well-formedness (Coleman & Pierrehumnbert, Hammond, Daland et al)
   b) epenthesis (Berent et al, Ohala)
   c) cluster reduction (Ohala)
   d) coda production (Zamuner et al)

33) Has to be syllable structure to get Treiman, Zamuner & Ohala effects

34) Initial expectations
   a) Form of grammar
   b) Which features
   c) Initial weights