Pedagogical goals: help inspiration w/ your research projects.

Question: what makes 2 types of lang different?

| lecture content | pointers to resources | individualized feedback |

↑

generic to everybody in class

↑

R press to add.

Course webpals:
- quick link to today
- quick link to webspals

- let’s make it mandatory
- presumably we’ll discuss when you are w.r.t. what you promised to get done in a week, but that’s not
- primary goal: feed general specific feedback from me. you pick time. duration.

<as for generic stuff:

top of webspals

some tendrils, alphabetized.

"What differentiates 2 types of lang?"

Last time: no systematic Bayesian ID of distinguishing features
- inspire you to try it or improve on it
- this time: case study of hypothesis-driven approach

Interesting, inspiring features:
- adapt to yr own devices...

Treas: Attendance 12
Finishing up from last time:
- see if students can reconstruct!

... should get to: posterior \( p(\theta | I) \)

\[
\alpha_i' = \alpha_i + c_i
\]

- we recalled many things, and discussed choice points.
  - note: using the mode way easier, b/c you can talk about
    "the multinomial" that this Prichet likes best.

Ideas/variants:
- have enough data that you can just estimate the "null hypothesis" version of
  your test statistic \( t \) from it? (bootstrap sampling?)
- if you just want the prior for smoothing, you could "just" smooth (e.g., by add-8)
  instead of going thru all this machinery.
- should you have a separate prior for "Dens vs. Reg" or just use one?
  Is the "use concept of the same data" to get prior \( \alpha_i \) in terms of
  \( \theta \)?
  Really compelling a (non-predictive?) prior that \( D_i \)'s R's use
  the same language?
- instead of taking the \( \theta \) that is the MLE for the posterior, could you
  integrate out?
- better estimates of the variance of the log-odds ratio?
- how they picked \( c_i = \text{E} \theta_i \).
Case study:
<Haris: Nunkena '13> - hypothesis driven approach.

Pedagogical stance:
(a) go them paper line by line, stick to studying the many interesting choices, ideas.

(b) use paper as springboard to insert asides on related interesting ideas [that may be too complicated or less direct applicability].

Decision: you can should read the paper on your own.
I'm going to start.
* * it really bears line by line analysis.

and this allows me to take the more meandering approach.

[Quick]: question set-up, data setup & idea.

[Quick]: outline of features:
- visual (3,1)
- people-centric (3/6)
- beautiful unusual lang. (3,7)
- narrative? sourced? an interview? (3,8)
- appeals to emotions? (3/5)
- out of research context (3,6)
- readability (5, P2)
- good discourse coherence, transitions (5, P2)
- interesting by other prev. interestingness features
- bag of words (Bow)

Each is treated w/ a lot of care, and paper is pretty detailed. As inspiration, can be expanded to apply to you.
- show MRC, to show limits of data
- go thru unusual language
  - lang. models
  - ref. to smoothing
  - perplexity: requires explaining cross-entropy
    - explain entropy. (use 'surprised' idea - see Brandon's blog post)
      - use language unit to show:
        \[ \Theta_i \text{ fixed} \]
        \[ \sum p(x) \log q(x) \] min when \( q(x) = p(x) \)
        \[ \sum \Theta_i \text{ true or reference} \]
        \[ \sum \Theta_i \text{ fixed} \] min for \( \Theta_i = \Theta_i \text{ fixed} \)
  - perplexity: interpret as branching factor or # of choices