

### CS1110 Analysis of “lies” using recursion 08 Oct

Have your iClickers out and set to frequency AB

Today: recursion in the context of an interdisciplinary application (computer science/computational linguistics, psychology, history/politics)

Reading for next lecture (casting about): Secs 4.2 & 4.3

#### Reminders:

- Prelim, 7:30-9pm today
  - Last name A-K: go to Olin 255
  - Last name L-Z: go to Upson B17
- A4 due Friday Oct. 16
- No labs Tue-Wed Oct 13-14 due to fall break including Tuesday
- Office/consulting hour changes between now and next Friday (A4 due date): see website
- A3 regrade requests: on CMS only; provide clear statement explaining what you think was graded unfairly or incorrectly

1

### Lies, damned lies, and statistics

James Pennebaker et al., “Lying words: predicting deception from linguistic styles”, 2003:

Claim: deceptive communication is characterized by (among others):

- fewer 1<sup>st</sup>-person singular and 3rd-person pronouns (“I”, “they”)
- more negative emotion words (“hate”, “enemy”)
- fewer “complex/exclusive” words (“but”, “except”, “without”)

**Research question (1):** What really are the best cues? (or models)

More realistic, convenient source of “lies”?

The “Iraq War Card False Statements Database”

<http://projects.publicintegrity.org/WarCard/Search/Default.aspx>

2

### Where we are, and why (besides automatic lie detection being inherently cool)

**Research question 1:** What are the best linguistic lie cues? (or models)

**Research question 2:** given statements regarding Iraq by top Bush administration officials, are the “true” and “false” ones distinguishable? This would imply something about their beliefs.

- Demonstration of interdisciplinary research involving computer science, psychology/linguistics, politics and history
- Demonstration of methodology in approaching a programming problem
  - stepwise refinement, writing and reading specs carefully, String manipulation, recursion, testing, etc.

(Lecture loosely based on joint work with CS grad student Cristian Danescu-Niculescu-Mizil and CS undergrad Haden Lee, in consultation with Comm. Prof. Jeff Hancock.)

3

### Formulating the task: might something be a lie cue?

#### Given

- a file containing “lie span mark-up”:

... with respect to iraq, the problem is quite simple. <font color="red">we suspect they are developing weapons of mass destruction. we more than suspect it; we know it.</font> there's an easy they ...

- a target word w (e.g., “they”)

**Consider** these two statistics for w:

- number of hits: occurrences of w in a “lie” span
- number of misses: occurrences of w not in a “lie” span

4

1. Get data from source file into Java manageable format
2. Get target word w
3. Process each occurrence of w in the source file's text
  - Is it in or not in a span?
4. Report relevant statistics

5

Is w in a span?

spanStart ... spanEnd ... w ... (no spanEnds) ... spanEnd ...

iW iEnd

6