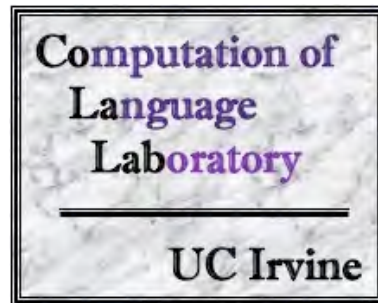


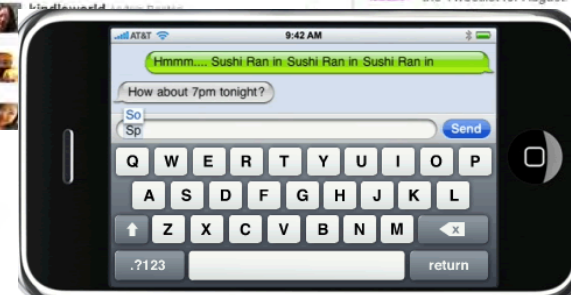
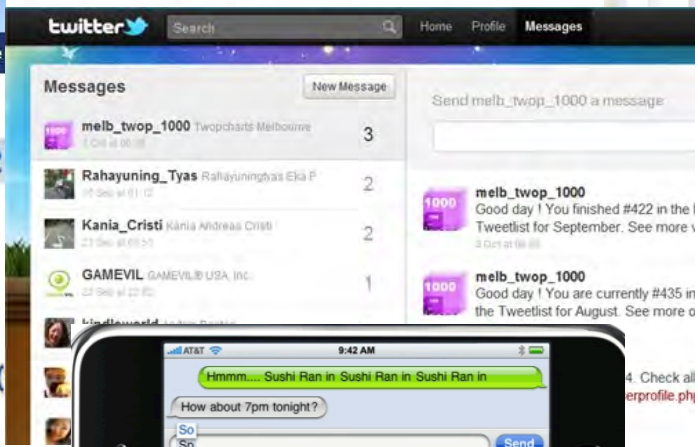
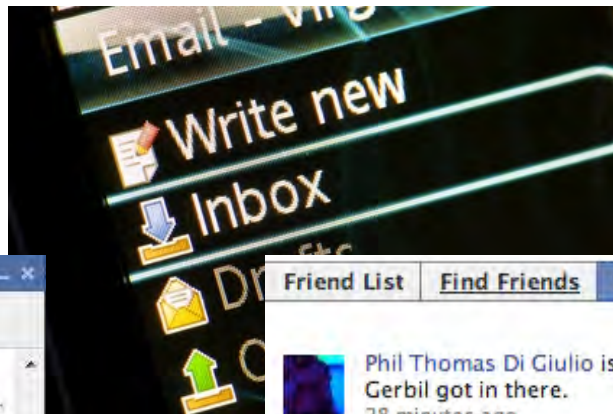
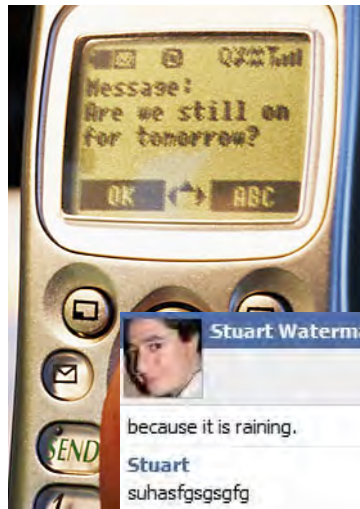
Making Sense of Information Overload in a Digital Age: Language & Identity

Lisa S. Pearl
lpearl@uci.edu



UC Irvine Expert Speakers Series
April 14, 2011

Got too many ways to communicate?



4. Check all erprofile.ph

Why this is a good thing...for computers

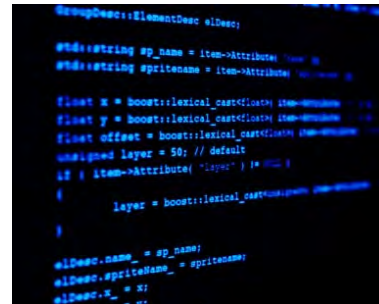
Computers can handle the details of large quantities of data.



fast processors



efficient software



lots of memory



Computers can learn interesting things **if there's enough data.**

Linguistic Data Consortium

The Linguistic Data Consortium supports language-related education, research and technology development by creating and sharing linguistic resources: data, tools and standards.

What we can learn about

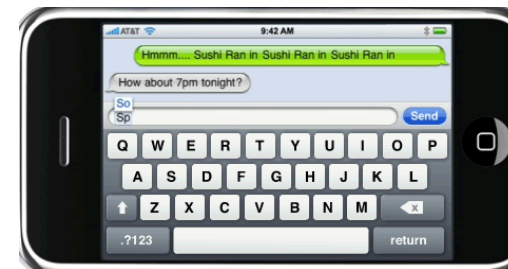
We all have our own way of communicating

Use non-verbal cues like voice pitch, eye gaze, and gesture (telephone, skype, in-person)



Use verbal cues

(text messages, email, twitter, blogs)



Verbal cues can be very distinctive.

Idea: We each have a “writeprint” (Iqbal et al. 2010) that captures our particular style.



So what's in a writeprint?

Some example **stylometric** and **content** features:

- punctuation (ex: ?, !, .)
- capitalization patterns (ex: i vs. I)
- self-reference (ex: me vs. you)
- lexical diversity (ex: lots of the same words)
- verbosity (ex: average sentence length, average word length)
- distinct synonym usage (ex: daddy vs. dad vs. papa vs. father)
- topics (ex: relationships vs. emails vs. sports)



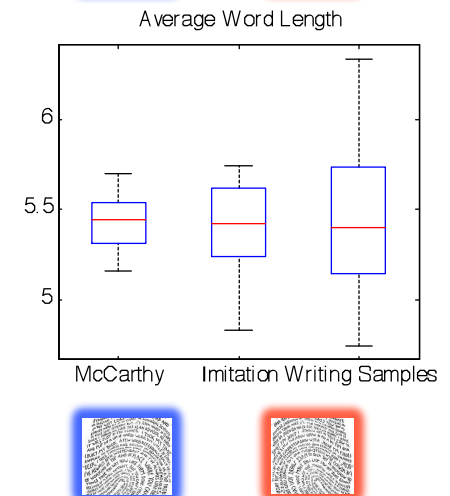
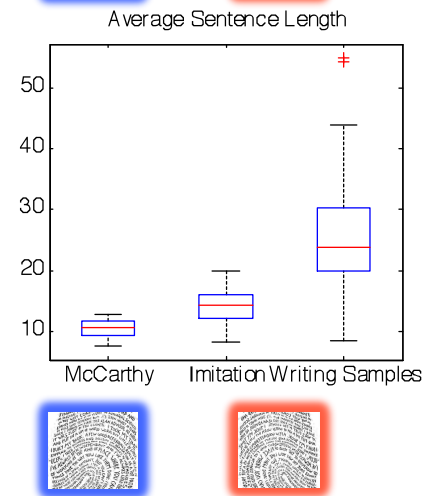
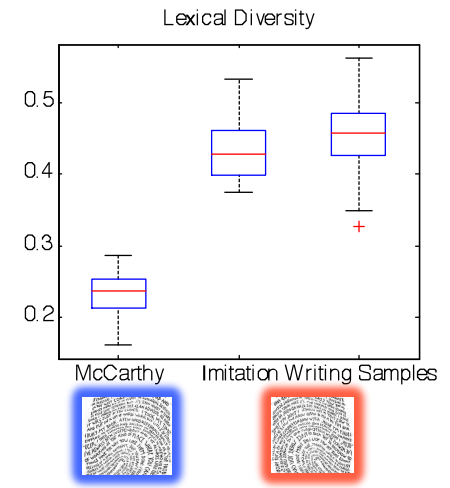
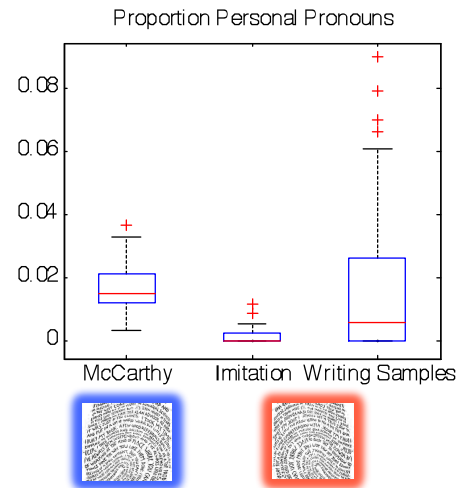
The use of writeprints

Writeprints in action for authorship deception

(Pearl & Steyvers, in prep)

Detecting imitation attacks where people attempted to imitate the writing of a particular author

100% detection, often using just a few features.



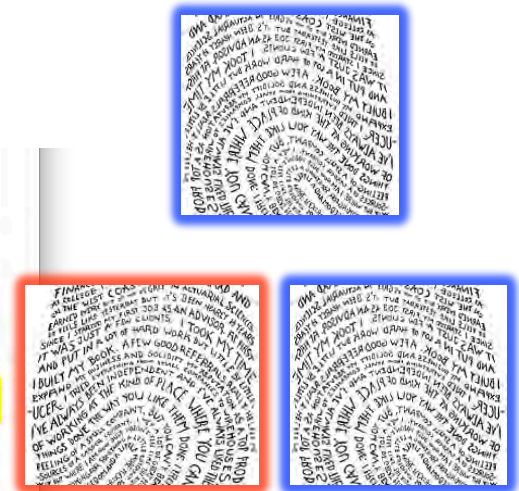
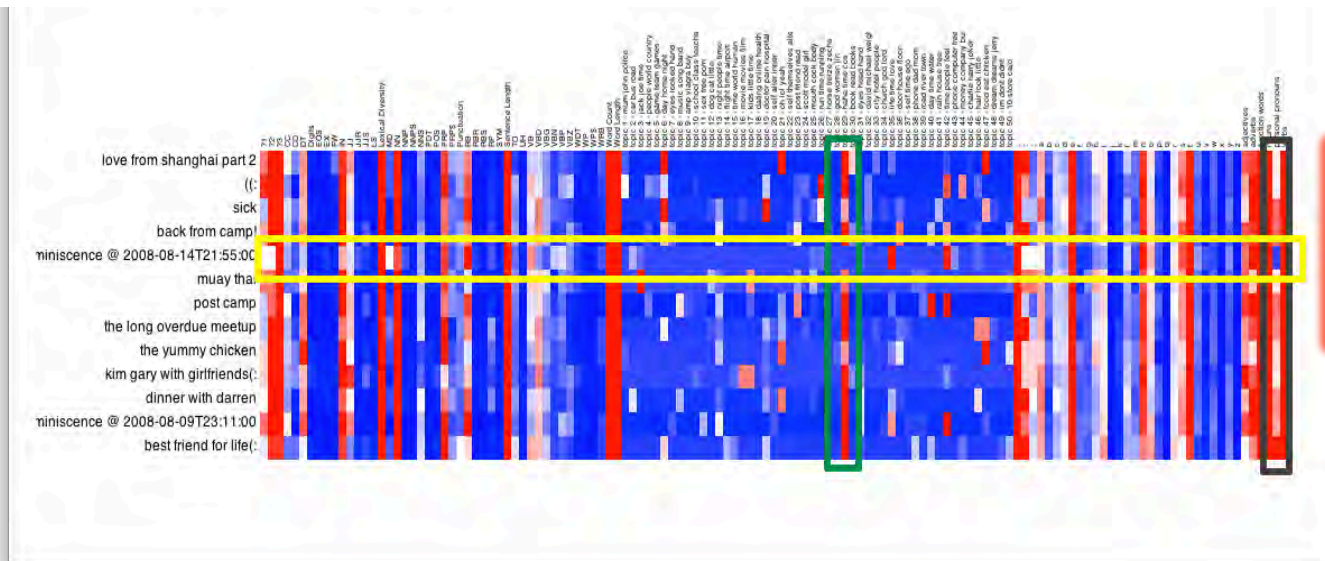
The use of writeprints

Authorship attribution: Who really wrote this?

Someone has anonymously posted a vitriolic entry on a public forum.

How could someone tell who wrote the entry?

Whose writeprint does the entry match?



The use of writeprints

Writeprints in action for authorship attribution

(Pearl & Steyvers, in prep)

Detecting whether a blog entry
of unknown authorship was
written by a given author.

81% attribution accuracy, using
a mix of content and stylometric
features.

Example content feature

Topic key words:

game, team, games, play,
win, poker, casino

Most likely authors:

andykelly	0.086
LouDiFalco	0.036
poker_star	0.035
ben	0.024
TheOminpotentQ	0.021
gamblingnet	0.021
OMmetsfan07	0.018

...had a rockin' win last night...

...poker's definitely my game...

...the games that are the best...



home_cookin

poker_star

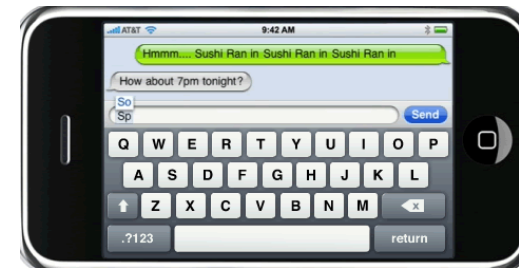


Writeprints: Detecting social cues

You may infer certain information if someone changes their writeprint.

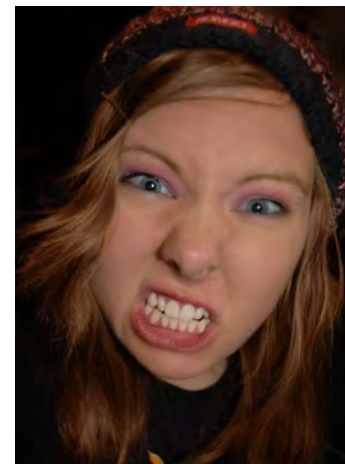
Text message example:

You're in a rush, so when your friend texts you asking you to meet her later on, you text back a quick "Sure". She fires back a text asking if you're mad at her.



What happened?

Friend's inference: Terse message = angry

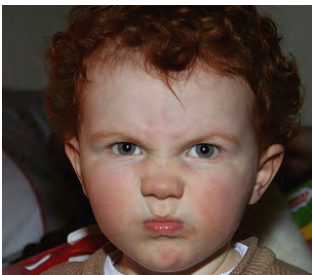


Writeprints: Detecting social cues


What if email & text software came with a **tone-checker**?

To develop this software, **we need lots of data about the intended tone of a message.**

Idea: Looking for the writeprint of **emotions** like anger and embarrassment, **attitudes** like confidence and disbelief, and **intentions** like persuasion or deception.



Using a game-with-a-purpose

Word Sleuth 
Test your social language intelligence

<http://gwap.ss.uci.edu>

A game-with-a-purpose (**GWAP**) (Pearl & Steyvers 2010)

- encourages people to generate messages with a specific tone
- evaluates how these messages are perceived by others



Word Sleuth

Test your social language intelligence



Write a message that more clearly expresses **embarrassment** than any other tag.

My message is complete!

Word Sleuth

Test your social language intelligence

Current score for lisa_test1

Expressive: 405 Receptive: 630 E-IQ: N/A R-IQ: 80 Activity Points: 252



You are playing on hard difficulty. You will earn 3x the base number of points when people guess correctly. Additionally, you will earn 10 points just for creating!

Write a message that more clearly expresses **embarrassment** than any other tag.

All tags: confidence, deception, disbelief, embarrassment, formality, persuading, politeness, rudeness

Don't use any of these taboo words:
embarrassed, embarrassment, embarrass, turning, slipped, wet, aww, tripped, conscious, mins

Please write quality messages. Items judged as bad may be removed; if so, points will be taken away. :(

My message is complete!

[Click to Skip](#)

Word Sleuth

Test your social language intelligence

Current score for lisa_test1


Expressive: 415 Receptive: 630 E-IQ: N/A R-IQ: 80 Activity Points: 253

Your message expressing embarrassment was:

Ugh, what's that smell? Oh no...I think you need to change your shirt...I'm really sorry about that...they don't usually overflow their diapers this fast...

Check back in later to see if other players could guess what you were trying to express with this message! If they could, your **expressive score** will increase.



Play with same settings 

Change play settings 

Word Sleuth

Test your social language intelligence

Current score for lisa_test1

Expressive: 415 Receptive: 630 E-IQ: N/A R-IQ: 80 Activity Points: 253

Figure out what this message is trying to express:

Hey did you want to dance? I'm really good at it.



You are guessing a medium sentence. You will earn 2x the base number of points.

<input type="radio"/> being deceptive	<input type="radio"/> formality
<input type="radio"/> politeness	<input type="radio"/> rudeness
<input type="radio"/> embarrassment	<input type="radio"/> confidence
<input type="radio"/> persuading	<input type="radio"/> disbelief

[I made my choice!](#)

[Click to Skip](#)

Word Sleuth

Test your social language intelligence

Current score for lisa_test1

Expressive: 415 Receptive: 660 E-IQ: N/A R-IQ: 80 Activity Points: 254


You guessed **confidence**, and the correct answer is **confidence**.
The message was: **Hey did you want to dance? I'm really good at it.**


Your guess was **correct**.

When you correctly identify what a message is trying to express, your **receptive score** will increase!

[Click here](#) to flag this sentence.



Play with same settings 

Change play settings 

The writeprint of social cues: Some examples

- deception:** - less use of 1st person pronouns like “me” and “my”
- more use of phrases like “of course” and “actually”
- persuasion:** - phrases like “have to” and “you should”
- rudeness:** - negative words like “ugly”, “annoying”, and “mean”
- politeness:** - words and phrases like “please”, “so sorry”, and “would you like”
- formality:** - less use of exclamation points and contractions like “don’t”
- disbelief:** - more question marks and words like “never” and “no way”
- confidence:** - more use of “I know”, “I believe”, and “positive”
- embarrassment:** - more use of “can’t believe” and “oh”

Writeprints: Where we can go

Who really wrote this?

- Protect against email identity theft
- Help recover authorship information



What did they really mean?

- tone-checker
- understand human perception of social cues
- help diagnose and train people who have difficulty detecting social cues in messages

