# Dimensions of Self-Expression in Facebook Status Updates

Adam D. I. Kramer, Ph.D.
Data Scientist
Facebook, Inc.
akramer@fb.com

Cindy K. Chung, Ph.D.
Department of Psychology
University of Texas at Austin
cindyk.chung@mail.utexas.edu

Presented 2011-07-20, ICWSM 11, Barcelona, Spain

# Text Analysis

- Natural language (text) is how we express *who* we are and *how* we are to the world.

# Word Counts

- "Word count" approaches to text analysis rely on the *lexical hypothesis,* that the most *important* terms will become encoded as *single lexical items* as language evolves (c.f. Goldberg, 1990)

Fred Zeleny hates passive-aggressive Facebook updates, but loves irony.
3 hours ago • Like • Comment

# Word Counts

- "Word count" approaches to text analysis rely on the *lexical hypothesis,* that the most *important* terms will become encoded as *single lexical items* as language evolves (c.f. Goldberg, 1990)

Fred Zeleny hates passive-aggressive Facebook updates, but loves irony.
3 hours ago • Like • Comment

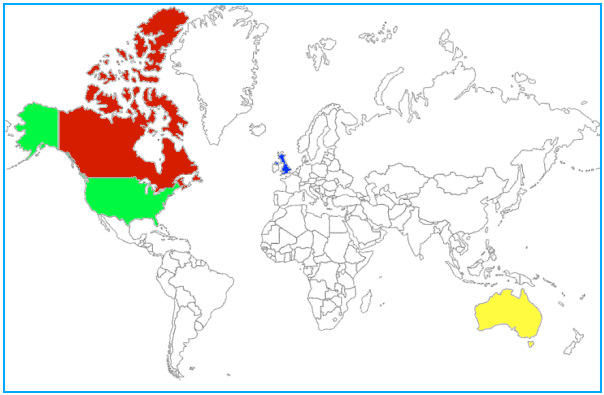
# The Meaning Extraction Method (MEM)

- Beyond our words: Our vocabularies in some context
  - Which words co-occur?

Fred Zeleny hates passive-aggressive Facebook updates, but loves irony.
3 hours ago • Like • Comment

# Our Goal

- Our corpus is Facebook “status updates”
  - Which words co-occur for users, across their updates?
  - Sampled separately for four countries: Australia, Canada, UK, US

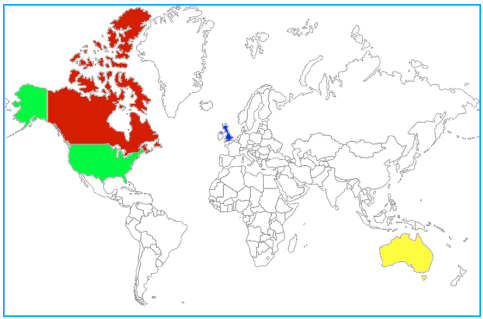


# Method

- We randomly selected 1m users from each country (N=4m)
  - Corpus: the set of all status updates from these users
- We parsed and sorted the 1,000 most frequently used words into a set of 760 stemmed content words, examined each update, conducted principle components analysis
  - PCAs computed both within and across countries
  - These PCAs extract “meaning:” *dimensions* along which users differ in usage of the Status Update product
- At no point were any raw status updates read by researchers, nor were aggregates viewed in any manner connected to an identifying quality of the user as part of this research.

# Results: Overall matrix

- These results should represent cross-cultural systems by which status updates were used
  - One large first factor emerged, representing the natural separation of the use of language among countries



# Results: Overall matrix

- These results should represent cross-cultural systems by which status updates were used
  - One large first factor emerged, representing the natural separation of the use of language among countries
  - Loadings for, e.g., “mom,” “vacation,” and “laundry,” opposed loadings for “pub,” “mate,” and “arse.”

# Results: Individual Countries

- Countries' individual correlation matrices were quite similar, $\alpha=0.97$, but differed significantly $\chi^2(3) = 40{,}202$; we examine the US first.

- Component 1 included loadings for "informal speech," including internet neologisms (haha, lol) and profanity

# Results: Individual Countries

- Countries' individual correlation matrices were quite similar, $\alpha$=0.97, but differed significantly $\chi^2(3) = 40{,}202$; we examine the US first.

- Component 1 included loadings for "informal speech," including internet neologisms (haha, lol) and profanity
- Component 2 contained loadings for positive words (good, like, love) and for time words (day, time). We call this "reminiscing" words

# Results: Individual Countries

- Countries' individual correlation matrices were quite similar, $\alpha=0.97$, but differed significantly $\chi^2(3) = 40{,}202$; we examine the US first.

- Component 1 included loadings for "informal speech," including internet neologisms (haha, lol) and profanity
- Component 2 contained loadings for positive words (good, like, love) and for time words (day, time). We call this "reminiscing" words
- Component 3 contained mostly school-related words (homework, study, essay)

# Results: Australia, Canada, the UK...

- ...overall the same. But differences exist in the specifics:
  - Australians showed negative loadings for profanity in the “school” component; *hubby* and *daughter* in informal speech
  - The UK used more neologisms for informality, but showed “Family” instead as an opposite to school

# Discussion

Informal Speech

Reminiscing

School

# Caveats and Future Directions

# Special Thanks!


- James W. Pennebaker
- Moira Burke
- The Facebook Data Team
  facebook.com/data
- Anonymous ICWSM reviewers
- ...and Fred Zeleny.


Questions?