Linguistic markers of secrets and sensitive self-disclosure in Twitter

David J. Houghton\textsuperscript{1} & Adam, N. Joinson\textsuperscript{2}

\textsuperscript{1}University of Birmingham, England
\textsuperscript{2}University of Bath, England
Thanks

- Dr. Mina Vasalou
- Matthieu Roesch
- Dr. Etienne Roesch
- Dr. Dawn Eubanks
- Dr. Nigel Caldwell
- All those who rated the Tweets
Background

- Balance between privacy and the use of social media to communicate
- Lots on how privacy is violated in social media
  - by site
  - by self
  - by others on/off the site
  - by usability issues
  - by government
- AT THE TIME of research
  - Little control over personal groups
- Sites that offered settings were complicated (Bonneau & Preibusch, 2009)
Self-disclosure

• Much of information posted requires self-disclosure in some form
  • Directly disclose information
  • Disclosure through objects (pictures/videos)
  • Disclosure via others (requires initial disclosure)

• Definition

“that which occurs when A knowingly communicates to B information about A which is not generally known and is not otherwise available to B” Worthy et al. (1969)

• Self-disclosure measured in breadth and depth

• Measurement usually survey based e.g.: -
  • JSDQ (Jourard & Lasakow, 1958)
  • SDSS (Chelune, 1975)
Self-disclosure

- Many benefits of disclosure and SM use generally
  - Self-presentation
  - Need to connect
  - Professional/personal profile
  - Self-esteem and psy. well-being
  - Social capital
  - Reduce uncertainty
  - Development of relationships
  - Trust
  - Group Identity
Conflict of Costs and Benefits

- Want to use SM for social connection
  - Has many benefits
- BUT, need to have control over information
  - Privacy settings not always well understood
  - If controlling through content, easy to ‘slip-up’?
How to control the balance?

- Control privacy by not using site
- Control privacy through content alterations generally
- Control privacy by audience
  - Get an LCD effect (Context Collapse, or Conflicting Social Spheres)
- Utilise settings available
- Directly instruct others “Do not put that on online!”
What about researchers?

- Self-disclosure is a pain to measure....
- Breadth – word count? or just verbose?
  - Category/frequency
- Depth – rated – Altman & Taylor onion metaphor
  - How is something core?
- Which bits of text should we look at?
- Which linguistic categories could help identify each of these?
Aims

- Determine a set of linguistic markers to secrets and sensitive self-disclosure to:
  - Aid users in their decisions to share information on social media
  - Determine what is considered ‘sensitive’ in a specific CMC context to further self-disclosure & communication literature
  - Contribute to the methodological tool-kit available o researchers looking at sensitive disclosure in CMC
Study 1 – Secrets - Method

• Two sites chosen
  • Twitter
    • 140 Characters
    • Open accounts
    • “*a*” search used to generate ‘random’ tweets
  • Secret Tweet
    • 140 Characters
    • No accounts, one public stream
    • “Post your secrets to Twitter anonymously”
    • Random sample taken as all on one stream
I'm not afraid of fight you, I'm afraid that if I start, I won't stop. We have kids too, I wouldn't want them to see that.

I am a struggling Christian addicted to pornography.

Growing up my aunt was verbally & physically abusive to my mother and I. I hate my mother for not standing up to her.

I barely eat so I can afford to surprise my boyfriend by flying almost 5,000 miles to see him for his birthday... and it's worth it!!
Study 1 – Secrets - Method

- Tweets & Secret tweets collected at variety of times to capture US & UK waking hours
- Avoided ReTweets, quotes and obvious sales/marketing accounts
- 2 Raters rated tweets for ‘Sensitivity’
  - English, middle-class, educated, white females
  - Use own definition of sensitive to produce norm of expected sensitivity
  - Scale 1-10
  - Cronbach’s Alpha = .871
Study 1 – Secrets - Analysis

- Tweets were cleaned according to LIWC user manual
  - Advanced word counter
  - Uses word stems
  - Produces *percentage* of words in the tweet that belong to a particular category

- Entered cleaned data into LIWC (N=500, 250+250)
  - Used all 80 Linguistic Categories
  - Category examples
    - Noun, pronoun, sexual, swear, cognitive, family, you, they

- LIWC output entered into logistic regression (forward)
Study 1 – Secrets - Results

- T-test indicated secret tweets rated as more sensitive than normal tweets
  - (secret=5.54, normal=1.98, t=-31.728, \( p<.000 \))

- The full model sig. predicted tweet type
  - \( \chi^2=500.11, \text{df}=16, p<.001 \)
  - Variance = 62.3% (step 1) to 84.3% (step 16)
  - 91.8% overall prediction accuracy
Study 1 – Secrets - Results

- More = Sensitive information disclosure
  - Word Count,
  - Personal Pronouns,
  - Past Tense,
  - Work Words,
  - Family Words,
  - Human Words,
  - Inhibitions,
  - Sexual Words.

- More = Normal information disclosure
  - Articles,
  - You,
  - She/He,
  - Swear Words,
  - Question Marks,
  - Exclamation Marks,
  - All Punctuation,
  - Fillers.
Study 1 – Secrets - Issues

- Started off as a *warm up* to further research
- Raters confused over *Level of Disclosure*
- Only 2 raters
- Logistic regression
  - Better utilisation of GLR through rating structure
- Secrets not key to understanding “*sensitive*”
- Subtle site differences
- But, data may be indicative within a secretive context
Study 2 – Sensitive Self-disclosure (SSD)

- Just used Twitter
  - Used same ‘cleaned’ sample from study 1 (N=250)
  - 6 Raters
    - 3 Male, 3 Female
    - Just normative ‘sensitivity’ no LoD
    - Use one site, one data set, rate within

- GLR used on scale ratings of tweets

- To reduce error and probability of type 1 error:
  - 30 Linguistic Process categories entered in Step 1 to control for general language use
  - Selected **16** variables of LIWC
    - Based on privacy, disclosure and comms literature
    - Also some from study 1 that were interesting
## Study 2 – SSD - Results

10+1 Linguistic Markers

<table>
<thead>
<tr>
<th>Category Type</th>
<th>Word Category</th>
<th>+/-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linguistic Processes</td>
<td>3\textsuperscript{rd} Person Singular (she/he)</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>3\textsuperscript{rd} Person Plural (they · p=.059)</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Verbs</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Present Tense</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Future Tense</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Prepositions</td>
<td>+</td>
</tr>
<tr>
<td>Psychological Processes</td>
<td>Family</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Negative Emotions</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Discrepancies</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Sexual</td>
<td>+</td>
</tr>
<tr>
<td>Personal Concerns</td>
<td>Death</td>
<td>+</td>
</tr>
</tbody>
</table>
Study 2 – SSD - Interpretation

- Markers tally with Altman & Taylor’s Onion
  - More sensitive markers tended to be more ‘core’ aspects of the self (study 1 & 2)

- Potential application to software tools to aid/alert users of information sensitivity / if it may go to undesired recipients
  - Pop-ups
  - Automated grouping
  - General sensitivity app

- Contribute to researchers’ ability to identify self-disclosure depth within public CMC communication contexts
Study 2 – SSD - Interpretation

- Disclosure of some information may be perceived negatively by recipients (such as the raters)
  - Over/under disclosure?

- With *combined* presence of 10 linguistic markers, one may be able to identify over disclosure that may cause negative consequences
  - Stress
  - Relationship degradation
Forewarned is forearmed

Me

Mundane Post

Relative

Partner

Friend

Work mate
Forewarned is forearmed

Me

‘Sensitive’ Post

Relative

Partner

Friend

Work mate
**SensiTweet**

- Thanks to Matthieu Roesch (matthieu.roesch@gmail.com)
- Preliminary, proof of concept tool
- Markers from Study 2 “The main study”
- Scraped 84,000 tweets from open accounts (inherent bias)
- Mean score for each of the linguistic markers was calculated with LIWC for all 84,000 tweets to get a baseline
- php script compares up to 2000 tweets for an individual user & compares it to the baseline
Have you ever wondered whether the information you write on Twitter is too sensitive? Our diagnostic tool assesses the sensitivity of your tweets compared to the average Twitter user. Just enter a Twitter handle to get a temperature reading of a profile.

http://interactionslab.net/sensitweet

David J. Houghton
d.j.houghton@bham.ac.uk
Have you ever wondered whether the information you write on Twitter is too sensitive? Our diagnostic tool assesses the sensitivity of your tweets compared to the average Twitter user. Just enter a Twitter handle to get a temperature reading of a profile.

Enter twitter handle: [input] Analyze tweets

4604 words were analysed in total (359 tweets). The profiledjhoughton is 12.69 % more sensitive than the average twitter user.

http://interactionslab.net/sensitweet

David J. Houghton
d.j.houghton@bham.ac.uk
Thanks

Please, feel free to contact me: -

Email: d.j.houghton@bham.ac.uk

WWW: davidhoughton.info

Twitter: djhoughton

More of our work: interactionslab.net