Identifying Personality and Psychological States in Words

Cindy K. Chung, Ph.D.
Senior Research Scientist
Intel Labs

In Collaboration With:
James W. Pennebaker, The University of Texas at Austin
Yla R. Tausczik, Carnegie Mellon University
Nairan Ramirez-Esparza, University of Connecticut
Adam D. I. Kramer, Facebook Immersive Computing Lab, Intel Labs
Markus Wolf and Hans Kordy, U of Heidelberg
Matthias R. Mehl, U of Arizona
Jeffrey T. Hancock, Cornell University
Arthur C. Graesser, U of Memphis
Two Research Tracks

How do the words we use in our everyday lives reflect our personalities, relationships, and wellbeing?

1. Function words & social relationships

2. Content words & dimensions of communication

Chung & Pennebaker
McCarthy & Boonthum (Eds.), *Applied Natural Language Processing*, 2012
Function Words and Social Relationships

Chung & Pennebaker
Fiedler (Ed.), Social Communication, 2007

Hancock, Beaver, Chung, Frazee, Pennebaker, Cai, & Graesser
Behavioral Sciences in Terrorism and Political Aggression, 2010

Pennebaker & Chung
Krippendorf & Bock (Eds.), “A Content Analysis Reader”, 2008
Do you have a big secret?

Participate in an email study for $100

A federally funded research project is exploring how people use emails when they have been living with a very big secret.

Requirements:
- Age 18 or older
- Your email program must be Gmail.
- You must have started keeping this secret sometime in the past 6-7 years (i.e. 2004 or later)
- If you eventually started telling others about the secret later, that’s OK.

You will not have to tell us exactly what the secret is. In addition to completing some questionnaires, we will want to analyze some of your emails. The procedure will guarantee anonymity to you and your email correspondents.

www.texasarchiveproject.org

This study is conducted at the University of Texas at Austin. DRB # 2009-12-0037

Tausczik, Chung, & Pennebaker
Under review
The (Galton) Lexical Hypothesis of Personality

1. Personality traits that are important to people’s lives will be represented in language.

2. The most important traits will be encoded as a single word in language.

Chung & Pennebaker
Journal of Research in Personality, 2008
Content Words and Dimensions of Communication

The Big Five

Openness
Conscientiousness
Extraversion
Agreeableness
Neuroticism
### Content Words and Dimensions of Communication

<table>
<thead>
<tr>
<th>NEGATIVITY</th>
<th>SOCIABILITY</th>
<th>MATURITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mad</td>
<td>Quiet</td>
<td>Mature</td>
</tr>
<tr>
<td>Bad</td>
<td>Shy</td>
<td>Successful</td>
</tr>
<tr>
<td>Hurt</td>
<td>Outgoing</td>
<td>Caring</td>
</tr>
<tr>
<td>Horrible</td>
<td>Reserved</td>
<td>Capable</td>
</tr>
<tr>
<td>Scared</td>
<td>Comfortable</td>
<td>Accomplished</td>
</tr>
<tr>
<td><strong>+N</strong></td>
<td><strong>-E</strong></td>
<td><strong>+E</strong></td>
</tr>
</tbody>
</table>

Chung & Pennebaker  
*Journal of Research in Personality, 2008*
### Content Words and Dimensions of Communication

<table>
<thead>
<tr>
<th>SOCIABILITY</th>
<th>HOBBIES</th>
<th>SIMPATIA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hablar</td>
<td>Musica</td>
<td>Carinosa</td>
</tr>
<tr>
<td>Speak</td>
<td>Music</td>
<td>Affectionate</td>
</tr>
<tr>
<td>Prefiero</td>
<td>Cine</td>
<td>Rencororsa</td>
</tr>
<tr>
<td>Prefer</td>
<td>Movies</td>
<td>Rancorous</td>
</tr>
<tr>
<td>Miedo</td>
<td>Bailar</td>
<td>Honesta</td>
</tr>
<tr>
<td>Fear</td>
<td>Dance</td>
<td>Honest</td>
</tr>
<tr>
<td>Expresar</td>
<td>Leer</td>
<td>Celosa</td>
</tr>
<tr>
<td>Express</td>
<td>Read</td>
<td>Jealous</td>
</tr>
<tr>
<td>Confianza</td>
<td>Escuchar</td>
<td>Sensible</td>
</tr>
<tr>
<td>Trust</td>
<td>Listen</td>
<td>Sensible</td>
</tr>
</tbody>
</table>
### Content Words and Dimensions of Communication

<table>
<thead>
<tr>
<th>INFORMAL SPEECH</th>
<th>POSITIVE EVENTS</th>
<th>SCHOOL</th>
</tr>
</thead>
<tbody>
<tr>
<td>fucked</td>
<td>day</td>
<td>homework</td>
</tr>
<tr>
<td>shitty</td>
<td>time</td>
<td>study</td>
</tr>
<tr>
<td>bitch</td>
<td>good</td>
<td>essay</td>
</tr>
<tr>
<td>ass</td>
<td>love</td>
<td>English</td>
</tr>
<tr>
<td>haha</td>
<td>likes</td>
<td>exam</td>
</tr>
</tbody>
</table>

Kramer & Chung

*Proceedings of the International Conference on Weblogs and Social Media, 2011*
### Content Words and Dimensions of Communication

<table>
<thead>
<tr>
<th>SYMPTOM</th>
<th>HOME</th>
<th>TREATMENT</th>
<th>WORK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain</td>
<td>Husband</td>
<td>Therapist</td>
<td>Company</td>
</tr>
<tr>
<td>Sad</td>
<td>Daughter</td>
<td>Therapy</td>
<td>Colleague</td>
</tr>
<tr>
<td>Body</td>
<td>Son</td>
<td>Physicians</td>
<td>Boss</td>
</tr>
</tbody>
</table>

Wolf, Chung, & Kordy
*Psychotherapy Research, 2010a; 2010b*
Thanks to

Jamie Pennebaker, UT
The Pennebaker Lab

Art Graesser, U of Memphis

Jeff Hancock, Cornell U

David I. Beaver, UT

Immersive Computing Lab, Intel Labs

Matthias R. Mehl, U of Arizona

Nairan Ramirez-Esparza, Uconn

Yla Tausczik, CMU

Adam D. I. Kramer, Facebook

Markus Wolf, U of Heidelberg

Hans Kordy, U of Heidelberg
Cindy.K.Chung@Intel.Com