#thighgap to #thyghgapp: Incrementation of orthographic variation on Instagram

Ian Stewart and Jacob Eisenstein
Georgia Institute of Technology
Content warning: eating disorders
Pro-ED Instagram

- Community that “share[s] content, advice and provide[s] social support for disordered or unusual eating choices” (Chancellor et al. 2015)
Pro-ED Instagram

1. Today's total: 966 calories 😎 today was horrible 😞 Ana #anorexia #staystrong #fat ugly #disgusting #nothappy #sad #strong #ed #eatingdisorder #girl #cutting #losingweight #weight #wshweight #clean #horribleday #horrible #selfharm

2. Feeling a little better... if I want to reach my goal I have to be patient and work harder. I hope you are doing the same😊 #cardio #exercise #loseweight #workout #skinny #thin #thighgap #collarbones

3. Dinner today was some whole grain spelt bread with sunflower seeds (yummy) topped with hummus, tomatoes, dried tomato spread and some cucumber... I also had some leftover sauerkraut 😋 #ed #anorexia #bullsh*t #vegan

#dinner #veganrecovery #thisisohospital #edfamily #ranawko #fitgirl #thegap #therescue #redfighter

#edwarriors #edwarrior #edsoldier #edfree #eatingdisorderrecovery
Pro-ED Instagram

 Feeling a little better.. if I want to reach my goal I have to be patient and work harder. I hope you are doing the same :) #cardio #exercise #loseweight #workout #skinny #thin #thighgap #collarbones

#thighgap
Feeling a little better if I want to reach my goal I have to be patient and work harder. I hope you are doing the same :) #cardio #exercise #loseweight #workout #skinny #thighgap #collarbones

#thighgap

thighgap

(space between thighs)
Orthographic variation

- “Represent spoken and vernacular forms, simulate prosody or shorten the message” (Androutsopoulos 2011)
- Tied to social differentiation, identity marking (Sebba 2009)
  - Community may only allow certain variants (Herring 2012)

<table>
<thead>
<tr>
<th>Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonetic</td>
<td><code>&lt;just&gt;</code> → <code>&lt;jus&gt;</code></td>
</tr>
<tr>
<td>Typographic</td>
<td><code>&lt;leet speak&gt;</code> → <code>&lt;1337 5934K&gt;</code></td>
</tr>
<tr>
<td>Syllabograms</td>
<td><code>&lt;before&gt;</code> → <code>&lt;b4&gt;</code></td>
</tr>
</tbody>
</table>
Dynamics of variation

- Writing conventions evolve over time (Sebba 2009)
- Communities are dynamic
  - Language change mirrors social dynamics (Danescu-Nicelscu-Mizil et al. 2013)
  - Locally-defined social categories: newcomers vs. regulars
- Changing practices of pro-ED community
  - Community of practice: “aggregate of people who come together around mutual engagement in an endeavor” (Eckert & McConnell-Ginet 1992)
Community change: hashtag ban

**SOCIETY**

**Instagram Bans Thinspo Content**

Instagram is the latest social media platform to ban thinspiration content. But are these policies effective?

By Heba Haasen @Heba__H | April 26, 2012

Thinspo content will no longer be welcome on Instagram. Following in the footsteps of Pinterest and Tumblr, Instagram is the latest social media site to ban “thinspiration” photos — images that are meant to provide motivation for those who want to lose weight and which health experts say often contribute to eating disorders.

Instagram’s new policy doesn’t come as a surprise. The app came under scrutiny last week when celebrity and Instagram user Alexa Chung posted a photo of herself and was attacked by users for being too skinny.

newsfeed.time.com/2012/04/26/instagram-bans-thinspo-content
Ban effect

No photos or videos yet!
Ban response

Post frequency over time

Frequency (in thousands)

Weeks since ban

-20 0 20 40 60 80 100 120

All posts

Posts containing variants
Ban response

**Post frequency over time**

- All posts
- Posts containing variants

**Social media sites, 2012-2014**

% of online adults who use the following social media websites, by year

- **Facebook**: 2012, 67; 2013, 71; 2014, 71
- **LinkedIn**: 2012, 20; 2013, 22; 2014, 28
- **Pinterest**: 2012, 15; 2013, 21; 2014, 26
- **Instagram**: 2012, 13; 2013, 17; 2014, 18
- **Twitter**: 2012, 16; 2013, 18; 2014, 23

---


PEW RESEARCH CENTER

Research questions

RQ1: Which community members adopt more variants?
Example variants

thighgap
Example variants

thyghgap   thyghgapp   thyygap   thyygapp
thighgap   thghgap    thiigap   thighgaapp   thygaps
thightgap  thightgrap thightpag   thygsp
Example variants

thyghgap    thyghgapp    thyygap    thyygapp    thighgap    thghgap    thiigap    thighgaapp    thygaps    thightgap    thightgrap    thightpag    thygsp

Depth
Example variants

thighgap  thyghgap  thyghgapp  thyygap  thyygapp
thghgap  thyygap  thiigap  thighgaapp  thygaps
thightgap  thightgrap  thightpag  thygsp

Depth
Incrementation of variation

- Orthographic variation as continuum
- Similar to phonetic incrementation
  - “Successive cohorts and generations of children advance a change beyond the level of their caretakers and role models” (Labov 2001)
- Do community members adopt deeper variants differently than shallow ones?
Research questions

RQ1: Which community members adopt more variants?

RQ2: Does a variant’s depth influence its likelihood of adoption by these community members?
Methods

- Data collection
- Compute orthographic depth (language variables)
- Compute membership attributes (community variables)
- Building regression models
Methods

- **Data collection**
- Compute orthographic depth (language variables)
- Compute membership attributes (community variables)
- Building regression models
Data collection (Chancellor et al. 2015)

- Collected in November 2014
  - Ban in April 2012
- 2.4 million posts
  - January 2011 to November 2014
Data collection (Chancellor et al. 2015)
Data collection (Chancellor et al. 2015)

<table>
<thead>
<tr>
<th>Identify pro-ED seed terms (not banned), mine Instagram</th>
</tr>
</thead>
<tbody>
<tr>
<td>anorexia, ed, bulimia</td>
</tr>
</tbody>
</table>
Data collection (Chancellor et al. 2015)

Identify pro-ED seed terms (not banned), mine Instagram

- anorexia, ed, bulimia

Filter for pro-ED content, identify top 200 hashtags
Data collection (Chancellor et al. 2015)

Identify pro-ED seed terms (not banned), mine Instagram

- anorexia, ed, bulimia

Filter for pro-ED content, identify top 200 hashtags

Manually identify 17 banned source hashtags

- ana, thighgap, thinspo
Data collection (Chancellor et al. 2015)

Identify pro-ED seed terms (not banned), mine Instagram

- anorexia, ed, bulimia

Filter for pro-ED content, identify top 200 hashtags

Extract 673 variants with regular expressions

- th*nspo* => thynspoo

Manually identify 17 banned source hashtags

- ana, thighgap, thinspo
Data collection (Chancellor et al. 2015)

Identify pro-ED seed terms (not banned), mine Instagram

- anorexia, ed, bulimia

Filter for pro-ED content, identify top 200 hashtags

Extract 673 variants with regular expressions

- th*nspo* => thynspoo

Manually identify 17 banned source hashtags

- ana, thighgap, thinspo

2.4 million posts
176,000 users
51% variant posts
673 variants
17 sources
Methods

- Data collection
- **Compute orthographic depth (language variables)**
- Compute membership attributes (community variables)
- Building regression models
Compute depth: edit distance

- Operations needed to transform source → variant hashtag
  - Used in dialectology (Nerbonne, Heeringa & Kleiweg 1999)

\[
\text{thighgap} \xrightarrow{\text{variant}} \text{thyygapp}
\]
Compute depth: edit distance

- Operations needed to transform source → variant hashtag
  - Used in dialectology (Nerbonne, Heeringa & Kleiweg 1999)

\[
\text{thighgap} \xrightarrow[\text{thyygapp}]{}
\]

thighgap
Compute depth: edit distance

- Operations needed to transform source $\rightarrow$ variant hashtag
  - Used in dialectology (Nerbonne, Heeringa & Kleiweg 1999)

$\text{thighgap} \leftrightarrow \text{thyygapp}$

$\text{thighgap} \rightarrow \text{thyghgap}$
**Compute depth: edit distance**

- Operations needed to transform source → variant hashtag
  - Used in dialectology (Nerbonne, Heeringa & Kleiweg 1999)

\[
\text{thighgap} \rightarrow \text{thyygapp}
\]

\[
1 \text{ thighgap} \rightarrow \text{thyghgap} \rightarrow 2 \text{ thyyhgap}
\]
Compute depth: edit distance

- Operations needed to transform source → variant hashtag
  - Used in dialectology (Nerbonne, Heeringa & Kleiweg 1999)
Compute depth: edit distance

- Operations needed to transform source → variant hashtag
  - Used in dialectology (Nerbonne, Heeringa & Kleiweg 1999)
Compute depth: edit distance

- Operations needed to transform source → variant hashtag
  - Used in dialectology (Nerbonne, Heeringa & Kleiweg 1999)
## Edit distance: Distribution of variants

<table>
<thead>
<tr>
<th>Edit distance</th>
<th>Variants</th>
<th>Top 3 variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>253</td>
<td>anarexia, bulimic, eatingdisorders</td>
</tr>
<tr>
<td>2</td>
<td>221</td>
<td>anarexyia, thinspoooo, thynspoo</td>
</tr>
<tr>
<td>3</td>
<td>108</td>
<td>secretsociety123, thinspoooo, thygap</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
<td>secret_society123, secretsociety_123, thinspoooooo</td>
</tr>
</tbody>
</table>
Edit distance: Adoption over time

![Edit distance frequencies graph](image)

- Post frequency vs. Weeks since ban
- Lines represent different DIST categories
- Chart shows frequency changes over time
Language variables

- per post:
  - TAGS, VARIANT, MAX_EDIT, DIST_1, DIST_4
Language variables

● per post:
  ○ TAGS, VARIANT, MAX_EDIT, DIST_1, DIST_4
Language variables

• per post:
  ○ TAGS, VARIANT, MAX_EDIT, DIST_1, DIST_4

TAGS=3
VARIANT=1
MAX_EDIT=2
DIST_1=1
DIST_4=0
Methods

● Data collection
● **Compute orthographic depth (language variables)**
● **Compute membership attributes (community variables)**
● Building regression models
Community data: membership attributes

- Locally-defined variables (within pro-ED community):
  - relative age
  - lifespan
Community data: membership attributes

- per post:
  - SINCE_START, TILL_END

- per user:
  - DATE_RANGE
Community data: membership attributes

- per post:
  - SINCE_START, TILL_END

- per user:
  - DATE_RANGE
Community data: membership attributes

- newcomer = low SINCE_START (< 10 weeks)
- committed user = high DATE_RANGE (≥ 10 weeks)
Community data: membership attributes

- newcomer = low SINCE_START (< 10 weeks)
- committed user = high DATE_RANGE (≥ 10 weeks)
Community data: membership attributes

- newcomer = low SINCE_START (< 10 weeks)
- committed user = high DATE_RANGE (≥ 10 weeks)

User A

transient

User B

committed
Recap: all variables

- per post:
  - VARIANT, DIST_1, DIST_4, MAX_EDIT, TAGS, SINCE_START, TILL_END, DATE

- per user:
  - DATE_RANGE
Methods

- Data collection
- **Compute orthographic depth (language variables)**
- Compute membership attributes (community variables)
- **Building regression models**
Questions

RQ1: Which community members adopt variants?

RQ2: Does a variant’s depth influence its likelihood of adoption by these community members?
# Regression: predictors

<table>
<thead>
<tr>
<th>RQ1</th>
<th>RQ2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>Logistic</td>
</tr>
<tr>
<td>Predicted</td>
<td>VARIANT</td>
</tr>
<tr>
<td>Predictors</td>
<td>SINCE_START</td>
</tr>
<tr>
<td></td>
<td>TILL_END</td>
</tr>
<tr>
<td></td>
<td>DATE_RANGE</td>
</tr>
<tr>
<td></td>
<td>DIST_1</td>
</tr>
<tr>
<td></td>
<td>DIST_4</td>
</tr>
</tbody>
</table>
Results
RQ1: Which community members adopt variants?
RQ1: Which community members adopt variants?

- Newcomers and committed (long-lifespan) users
RQ1: Which community members adopt variants?
RQ1: Which community members adopt variants?

- Regression results
RQ1: Which community members adopt variants?

- Regression results
- Predicting VARIANT
  - SINCE_START negatively correlated ($\beta = -0.00456$, $p < 0.001$)
  - TILL_END positively correlated ($\beta = 0.00294$, $p < 0.001$)
  - DATE_RANGE positively correlated ($\beta = 0.00294$, $p < 0.001$)
RQ1: Which community members adopt variants?

- Regression results
- Predicting VARIANT
  - SINCE_START negatively correlated ($\beta = -0.00456$, $p < 0.001$)
  - TILL_END positively correlated ($\beta = 0.00294$, $p < 0.001$)
  - DATE_RANGE positively correlated ($\beta = 0.00294$, $p<0.001$)
- Conclusion: variants adopted more often by newcomers and committed members
RQ2: Does a variant’s depth influence its likelihood of adoption?
RQ2: Does a variant’s depth influence its likelihood of adoption?

- Deeper variants associated with newcomers and committed members
RQ2: Does a variant’s depth influence its likelihood of adoption?

Newcomers versus regulars
RQ2: Does a variant’s depth influence its likelihood of adoption?

Committed versus transient
RQ2: Does a variant’s depth influence its likelihood of adoption?

- Regression results
RQ2: Does a variant’s depth influence its likelihood of adoption?

- Regression results
- Predicting DIST_1
  - SINCE_START $\beta = -0.00177$, (p < 0.001)
  - TILL_END $\beta = 0.00311$ (p < 0.001)
RQ2: Does a variant’s depth influence its likelihood of adoption?

- Regression results
- Predicting DIST_1
  - SINCE_START $\beta = -0.00177$, (p < 0.001)
  - TILL_END $\beta = 0.00311$ (p < 0.001)
- Predicting DIST_4
  - SINCE_START $\beta = -0.00450$ (p < 0.001)
  - TILL_END $\beta = 0.0133$ (p < 0.001)
RQ2: Does a variant’s depth influence its likelihood of adoption?

- Regression results
- Predicting DIST_1
  - SINCE_START $\beta = -0.00177$, (p < 0.001)
  - TILL_END $\beta = 0.00311$ (p < 0.001)
- Predicting DIST_4
  - SINCE_START $\beta = -0.00450$ (p < 0.001)
  - TILL_END $\beta = 0.0133$ (p < 0.001)
- Conclusion: depth of variation correlates more strongly with adoption by newcomers and committed members
Summary of findings

● Newcomers use more variants, deeper variation
  ○ Supports prior findings (Danescu-Niculescu-Mizil et al. 2013)
● Committed members also use more/deeper variants
● Deeper variants $\rightarrow$ stronger effects
  ○ Depth may influence orthographic perception in pro-ED community
● Additional: unclear social reception
  ○ Mixed results (likes $\neq$ comments)
Implications and future work

● Implications
  ○ Online communities provide useful setting to study large-scale, long-term language variation
  ○ Orthographic variation as incrementation
  ○ Sociotechnical effect on language variation

● Future work
  ○ Different processes of orthographic variation: deletion, lengthening, metathesis
References


Thank you!

Special thanks: Stevie Chancellor and Munmun De Choudhury (Georgia Institute of Technology)

Questions?