#anorexia, #anarexia, #anarexyia: Characterizing Online Community Practices with Orthographic Variation

Ian Stewart, Stevie Chancellor, Munmun De Choudhury and Jacob Eisenstein
SocialNLP 2017
Language variation online

- Online language is subject to profound variation and rapid change over time. (Androutsopoulos 2011)
- An individual’s adoption of language change in a particular online community is related to their community membership. (Sebba 2009)
Orthographic change

- Prior studies in online language have focused mainly on **lexical** change. (Danescu-Niculescu-Mizil et al. 2013)
- Change at the level of **orthography** is also important but less well understood. (Herring 2012)

\[
\text{anorexia} \leftrightarrow \text{anarexia} \leftrightarrow \text{anarexyia}
\]
Research questions

RQ1: **Who** adopts new orthographic variants?

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

RQ3: Does a variant’s **depth** influence its **social reception**?
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. 2016)
Content ban

No photos or videos yet!
Pro-ED posts (Chancellor et al. 2016)

2.4 million pro-ED posts
17 source hashtags
673 variant hashtags
Variants grow more frequent, “deeper”
Who drives this change?

RQ1: Who adopts new orthographic variants?
Differentiating community members

- Membership attributes:
  - Age
  - Lifespan
RQ1: Who adopts new orthographic variants?

- **Newcomers** and **committed** (long-lifespan) community members.
RQ1: Who adopts new orthographic variants?
RQ1: Who adopts new orthographic variants?

- Regression results
  - Predicting appearance of any variant in a post.
RQ1: Who adopts new orthographic variants?

- Regression results
  - Age: $\beta = -0.00456^{***}$, effect size = -0.348
  - Lifespan: $\beta = 0.00294^{***}$, effect size = 0.654

*** = p < 0.001
RQ1: Who adopts new orthographic variants?

- Conclusion: variants adopted more often by newcomers and committed members.
Research questions

RQ1: Which community members adopt more variants?

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

RQ3: Does a variant’s depth influence the post’s social reception?
Compute depth: edit distance

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

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

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

```
thighgap → thyygapp
```

```plaintext
thighgap → thyghgap → thyyhgapp → thyy̬gap → thyygapp
```

1 2 3 4
RQ2: Does a variant’s depth influence its likelihood of adoption?
RQ2: Does a variant’s depth influence its likelihood of adoption?

- Deeper variants are adopted by newcomers and committed members.
RQ2: Does a variant’s depth influence its likelihood of adoption?

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

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

- Regression
  - Predicting appearance of shallow (distance 1) and deep (distance 4) variant.
RQ2: Does a variant’s depth influence its likelihood of adoption?

- **More extreme effects** for deeper variants.
- **Distance 1**
  - Age: $\beta = -0.00177^{***}$, effect size = -0.097
- **Distance 4**
  - Age: $\beta = -0.00450^{***}$, effect size = -0.416
RQ2: Does a variant’s depth influence its likelihood of adoption?

- Conclusion: deeper variants are more often adopted by newcomers and committed members.
RQ3: Does a variant’s depth influence the post’s social reception?

- Social reception: *likes* and *comments*. 
RQ3: Does a variant’s depth influence the post’s social reception?

• Regression results
  ○ Control for hashtags used, presence of any variant, and fixed-effect for each member.
RQ3: Does a variant’s depth influence the post’s social reception?

<table>
<thead>
<tr>
<th>COMMENTS</th>
<th>LIKES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edit distance: -3.72E-3</td>
<td>Edit distance: 0.0130***</td>
</tr>
</tbody>
</table>

*** = p < 0.001,
otherwise p > 0.05
RQ3: Does a variant’s depth influence the post’s social reception?

- Conclusion: greater orthographic depth implies more likes.
Implications

- Why do newcomers use more variants and deeper variation?
  - (1) Newcomers are over-compensating for perceived orthographic conventions.
  - (2) “Newcomers” are banned accounts whose behavior is worsening after returning from ban (Cheng, Danescu-Niculescu-Mizil, Leskovec 2014).
  - (3) Newcomers are generating new variants to distinguish from current community members (“flag-planting”).
Implications

- Why do newcomers use more variants and deeper variation?
  - (1) Newcomers are over-compensating for perceived orthographic conventions.
  - (2) “Newcomers” are banned accounts whose behavior is worsening after returning from ban (Cheng, Danescu-Niculescu-Mizil, Leskovec 2014).
  - (3) Newcomers are generating new variants to distinguish from current community members (“flag-planting”).

- Deeper variants $\rightarrow$ **stronger effects**.
  - Effect is **incremental**, not just binary.
  - Orthographic variation reveals hidden community dynamics.
References


Acknowledgments

Thanks to Brendan O’Connor for the Twitter data.

This research was supported by Air Force Office of Scientific Research award FA9550-14-1-0379, National Institutes of Health award R01-GM112697, and National Science Foundation award 1452443.
Questions?

istewart6@gatech.edu

@alethioguy
Backup slides
Pro-ED Instagram

1. Today's total: 996 calories 😞 today was horrible 😔 😞 #Anorexia #staystrong #fat #ugly #disgusting #nothappy #sad #strong #ed #eatingdisorder #girl #cutting #losingweight #weight #wishingweight #clean #horribleday #horrible #selfiarm

2. Feeling a little better...if I want to reach my goal I have to be patient and work harder. I hope you are also 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 #bulimia #vegan #dinner #veganrecovery #thissorhospitl #edfamily #ranauho #fab #ana #recover #redfighter #eawarriors #edwarrior #edsoldier #erfree #eatingdisorderrecovery
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 #excercise #loseweight #workout #skinny #thin #thighgap #collarbones #thighgap
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

thighgap

(space between thighs)
Ban effect
All source hashtags

● ana
● anorexia
● anorexianervosa
● bonespo
● bulimia
● eatingdisorder
● mia
● proana
● proanorexia

● probulimia
● promia
● secretsociety123
● skinny
● thighgap
● thin
● thinspiration
● thinspo
## Distribution of variants

<table>
<thead>
<tr>
<th>Edit distance</th>
<th>Top 3 variants</th>
<th>Source hashtags</th>
<th>Variants</th>
<th>% posts with at least one variant from group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>anarexia, bulimic, eatingdisorders</td>
<td>17</td>
<td>253</td>
<td>41.1%</td>
</tr>
<tr>
<td>2</td>
<td>anarexyia, thinspoo0, thyospoo</td>
<td>15</td>
<td>221</td>
<td>2.07%</td>
</tr>
<tr>
<td>3</td>
<td>secretsociety123, thinspoo000, thygap</td>
<td>15</td>
<td>108</td>
<td>9.60%</td>
</tr>
<tr>
<td>4</td>
<td>secret_society123, secretsociety_123, thinspoo000</td>
<td>10</td>
<td>50</td>
<td>10.4%</td>
</tr>
</tbody>
</table>
Distribution of variants

Distribution of posts by edit distance

Distribution of variants by edit distance
Differentiating community members

- Per post:
  - SINCE_START, TILL_END

- Per member:
  - DATE_RANGE
Variable distributions
Regression: predictors

<table>
<thead>
<tr>
<th>RQ1</th>
<th>RQ2</th>
<th>RQ3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>Logistic</td>
<td>Logistic</td>
</tr>
<tr>
<td>Predicted</td>
<td>VARIANT</td>
<td>DIST_1, DIST_4</td>
</tr>
<tr>
<td>Predictors</td>
<td>SINCE_START,</td>
<td>SINCE_START, DATE_RANGE,</td>
</tr>
<tr>
<td></td>
<td>TILL_END,</td>
<td>DATE</td>
</tr>
<tr>
<td></td>
<td>DATE_RANGE,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DATE</td>
<td></td>
</tr>
</tbody>
</table>
RQ1: Who adopts new orthographic variants?

- Regression results

<table>
<thead>
<tr>
<th></th>
<th>β</th>
<th>SE</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>SINCE_START</td>
<td>-0.00456***</td>
<td>2.97E-4</td>
<td>-0.348</td>
</tr>
<tr>
<td>TILL_END</td>
<td>0.00294***</td>
<td>2.88E-4</td>
<td>0.654</td>
</tr>
<tr>
<td>DATE</td>
<td>0.00529***</td>
<td>1.77E-4</td>
<td>0.746</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>β</th>
<th>SE</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATE_RANGE</td>
<td>0.00294***</td>
<td>2.89E-4</td>
<td>0.654</td>
</tr>
<tr>
<td>DATE</td>
<td>0.00541***</td>
<td>1.77E-4</td>
<td>0.746</td>
</tr>
</tbody>
</table>

*** = p < 0.001
RQ3: Does a variant’s depth influence the post’s social reception?

<table>
<thead>
<tr>
<th>COMMENTS</th>
<th>β</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SINCE_START</td>
<td>5.27E-3*</td>
<td>1.57E-3</td>
</tr>
<tr>
<td>TAGS</td>
<td>0.110***</td>
<td>2.57E-3</td>
</tr>
<tr>
<td>VARIANT</td>
<td>-7.89E-3</td>
<td>3.44E-3</td>
</tr>
<tr>
<td>MAX_POP</td>
<td>-2.33E-3</td>
<td>1.26E-3</td>
</tr>
<tr>
<td>MAX_EDIT</td>
<td>-3.72E-3</td>
<td>5.51E-3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LIKES</th>
<th>β</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SINCE_START</td>
<td>-0.0319***</td>
<td>9.03E-4</td>
</tr>
<tr>
<td>TAGS</td>
<td>0.224***</td>
<td>1.47E-3</td>
</tr>
<tr>
<td>VARIANT</td>
<td>-0.00149***</td>
<td>1.98E-3</td>
</tr>
<tr>
<td>MAX_POP</td>
<td>-3.89E-3***</td>
<td>7.25E-4</td>
</tr>
<tr>
<td>MAX_EDIT</td>
<td><strong>0.0130</strong>*</td>
<td>3.16E-3</td>
</tr>
</tbody>
</table>