A Sociolinguistic Analysis of Emotives in Irish English

Martin Schweinberger
www.martinschweinberger.de

New Perspectives on Irish English 5
Introduction

- Communicating one’s emotional state is a core element of language.
- Almost no corpus-based research on how speakers convey their emotions to others or on sociolinguistic differences in emotional language.
- Existing research on the verbalization of emotion is typically related to
  - subjectivity (Langacker 1985; Lyons 1981);
  - expression of stance (cf. e.g. Goodwin et al. 2012);
  - conceptual metaphor theory (cf. e.g. Kövecses 2003; Meier and Robinson 2005);
  - componential analysis (Wierzbicka 1992)
Introduction

- Synchronic quantitative corpus–based study of emotives in Irish English based on the Irish component of ICE Ireland 1.2.2 (Kirk and Kallen 2008).

- Emotives are words associated with one of eight core emotions (ANGER, ANTICIPATION, DISGUST, FEAR, JOY, SADNESS, SURPRISE, TRUST) (cf. Ekman 1992; Plutchik 1980) (cf. (1))

  (1) a. SADNESS: cry, tragedy, etc.
  b. JOY: happy, beautiful, etc.
  c. ANGER: fit, burst, etc.

- The present study analyses correlations between emotive use and extra-linguistic factors.
Sentiment Analysis

- Application of a sentiment analysis to code for emotional language use of socially distinct groups
- Widely used by companies to track the public reception of products/to check for potential issues or problems
- Core technique in computational linguistics (Feldman 2013; Liu and Zhang 2012; Younis 2015)
- First core linguistic studies utilizing this method (Hoffmann 2017; Joyce and Deng 2018)
- Inter-subjective and replicable coding of emotion (advantage)
Emotions and language

- Women talk more emotional than men (Aldrich and Tenenbaum 2006; Coates 2015; Lakoff 1973; Tannen 1994; Wolf 2000)
- Women talk more about emotions (Goldshmidt and Weller 2000)
- Women report being more emotionally expressive (Bronstein et al. 1996; Brebner 2003)
- Gender specific effects
  - Women being and expressing sadness more than men (Grossman and Wood 1993)
  - Girls report more frequent sadness compared to male peers (Brody 1984; Stapley and Haviland 1989).
  - Boys express anger more readily than girls (Brody 1984).
Hypotheses

- $H_A$
  Women show the highest frequencies of emotives compared with men (Aldrich and Tenenbaum 2006; Coates 2015)

- $H_B$
  Women show the highest frequencies of overall and positive emotives (Joy) compared with men

- $H_C$
  Men use higher frequencies of negative emotives (Anger) compared with women (Brody 1984)

- $H_D$
  Men adapt to women in mixed-gender conversations (cf. Wolf 2000)
Data (Processing)

The Irish component of the *International Corpus of English* (ICE) (Kirk and Kallen 2008)

- Split spoken dialogue data into utterances
- Removal of meta information
- Part–of-speech tagging
  - PoS–tagging via tidytext library in R
  - Problem: Accuracy of PoS–tagging not yet evaluated
- Split spoken data data into words
- Removing stop words (non-lexical function words)
- Apply Sentiment Analysis (determine emotional value of adjectives)
- Add speaker information (age, sex, etc.) (Kallen and Kirk 2008).
## Data

<table>
<thead>
<tr>
<th>Age</th>
<th>Gender</th>
<th>Speakers</th>
<th>Words</th>
<th>Anger</th>
<th>Anticipation</th>
<th>Disgust</th>
<th>Fear</th>
<th>Joy</th>
<th>Sadness</th>
<th>Surprise</th>
<th>Trust</th>
</tr>
</thead>
<tbody>
<tr>
<td>19-25</td>
<td>female</td>
<td>151</td>
<td>30,794</td>
<td>550</td>
<td>1368</td>
<td>431</td>
<td>732</td>
<td>1,132</td>
<td>736</td>
<td>580</td>
<td>1,357</td>
</tr>
<tr>
<td>19-25</td>
<td>male</td>
<td>61</td>
<td>9,451</td>
<td>169</td>
<td>406</td>
<td>130</td>
<td>221</td>
<td>325</td>
<td>215</td>
<td>196</td>
<td>413</td>
</tr>
<tr>
<td>26-33</td>
<td>female</td>
<td>71</td>
<td>13,279</td>
<td>212</td>
<td>604</td>
<td>184</td>
<td>311</td>
<td>488</td>
<td>276</td>
<td>225</td>
<td>551</td>
</tr>
<tr>
<td>26-33</td>
<td>male</td>
<td>19</td>
<td>4,873</td>
<td>138</td>
<td>221</td>
<td>68</td>
<td>224</td>
<td>134</td>
<td>182</td>
<td>89</td>
<td>308</td>
</tr>
<tr>
<td>34-49</td>
<td>female</td>
<td>48</td>
<td>8,700</td>
<td>138</td>
<td>369</td>
<td>103</td>
<td>265</td>
<td>341</td>
<td>193</td>
<td>165</td>
<td>511</td>
</tr>
<tr>
<td>34-49</td>
<td>male</td>
<td>74</td>
<td>14,156</td>
<td>328</td>
<td>649</td>
<td>183</td>
<td>532</td>
<td>399</td>
<td>396</td>
<td>238</td>
<td>1,031</td>
</tr>
<tr>
<td>50+</td>
<td>female</td>
<td>54</td>
<td>13,355</td>
<td>184</td>
<td>641</td>
<td>142</td>
<td>313</td>
<td>569</td>
<td>294</td>
<td>274</td>
<td>686</td>
</tr>
<tr>
<td>50+</td>
<td>male</td>
<td>86</td>
<td>16,608</td>
<td>427</td>
<td>748</td>
<td>236</td>
<td>566</td>
<td>425</td>
<td>450</td>
<td>271</td>
<td>1,056</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>570</td>
<td>111,216</td>
<td>2,146</td>
<td>5,006</td>
<td>1,477</td>
<td>3,164</td>
<td>3,813</td>
<td>2,742</td>
<td>2,038</td>
<td>5,913</td>
</tr>
</tbody>
</table>
Statistical analysis

Mixed effects binomial logistic regression
(step-wise step-up fitted)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Scale</th>
<th>Levels</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>txtspk</td>
<td>categorical</td>
<td>S1A-001A : S1B − 080P</td>
<td>Individual speaker</td>
</tr>
<tr>
<td>pos</td>
<td>categorical</td>
<td>adj (adjective); nnp (noun); vbf (verb); other</td>
<td>Part-of-speech</td>
</tr>
<tr>
<td>pos</td>
<td>nominal</td>
<td>19-25; 26-33; 34-49; 50+</td>
<td>Age of speaker</td>
</tr>
<tr>
<td>sex</td>
<td>nominal</td>
<td>female; male</td>
<td>Gender of speaker</td>
</tr>
<tr>
<td>smsx</td>
<td>nominal</td>
<td>same; mixed</td>
<td>Same- or mixed-sex conversation</td>
</tr>
<tr>
<td>pripub</td>
<td>nominal</td>
<td>private; public</td>
<td>Conversation setting</td>
</tr>
<tr>
<td>ints</td>
<td>categorical</td>
<td>1, 2-3, 4-5, 6+</td>
<td>Audience size</td>
</tr>
<tr>
<td>pos</td>
<td>categorical</td>
<td>adj (adjective); nnp (noun); vbf (verb); other</td>
<td>Part-of-speech</td>
</tr>
</tbody>
</table>
Statistical analysis

Mixed effects binomial logistic regression
(step-wise step-up fitted)

<table>
<thead>
<tr>
<th>Dependent variables</th>
<th>Nominal</th>
<th>0;1</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>emosum</td>
<td>nominal</td>
<td>0;1</td>
<td>Word is associated with at least one basic emotion</td>
</tr>
<tr>
<td>anger</td>
<td>nominal</td>
<td>0;1</td>
<td>Word is associated with anger</td>
</tr>
<tr>
<td>anticipation</td>
<td>nominal</td>
<td>0;1</td>
<td>Word is associated with anticipation</td>
</tr>
<tr>
<td>disgust</td>
<td>nominal</td>
<td>0;1</td>
<td>Word is associated with disgust</td>
</tr>
<tr>
<td>fear</td>
<td>nominal</td>
<td>0;1</td>
<td>Word is associated with fear</td>
</tr>
<tr>
<td>joy</td>
<td>nominal</td>
<td>0;1</td>
<td>Word is associated with joy</td>
</tr>
<tr>
<td>sadness</td>
<td>nominal</td>
<td>0;1</td>
<td>Word is associated with sadness</td>
</tr>
<tr>
<td>surprise</td>
<td>nominal</td>
<td>0;1</td>
<td>Word is associated with surprise</td>
</tr>
<tr>
<td>trust</td>
<td>nominal</td>
<td>0;1</td>
<td>Word is associated with trust</td>
</tr>
</tbody>
</table>
Results

(stat. sig.)
Emotives in Irish English

Results
## Overview

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Emosum</th>
<th>Anger</th>
<th>Anticipation</th>
<th>Disgust</th>
<th>Fear</th>
<th>Joy</th>
<th>Sadness</th>
<th>Surprise</th>
<th>Trust</th>
</tr>
</thead>
<tbody>
<tr>
<td>pos</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>pripub</td>
<td>✓</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>sex</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ints</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>age</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>smsx</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>interactions</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>✓</td>
</tr>
</tbody>
</table>
Findings

- Adjectives, in particular, and nouns used to express emotionality
- Men use more emotives compared with women (contrary to $H_A$)
- People use more emotives in public settings
- Previous research substantiated
  - Boys express anger more readily than girls (Brody 1984) → Men use more anger emotives ($H_C$ substantiated)
- Findings at odds with previous research
  - Women express sadness more readily than men (Goodwin et al. 2012) → Men use more sadness emotives
Findings

- Women use more joy-emotives while men use more anger- and sadness-emotives ($H_B$ and $H_C$ substantiated)
- Speakers use more trust emotives in same-gender conversations regardless of sex
- The older speakers become, the more trust emotives they use
- While overall emotionality is associated with public rather than private settings, surprise emotives are used more in private settings
- No significant interactions (contrary to $H_D$)
Discussion

Problems

- Sentiment analysis very coarse-grained
- Focus on lexemes while ignoring/neglecting...
  - Negation
  - Context
  - Fixed expressions

Advantages

- Results and studies are replicable
- It works on large scale data sets

Still, one of the first applications of sentiment analysis to an investigation of social stratification of linguistic phenomena.
Thanks 😊😊

www.martinschweinberger.de


