Expressiveness in Flemish Online Teenage Talk:
A Corpus-Based Analysis of Social and Medium-Related Linguistic Variation

Lisa Hilte
Reinhild Vandekerckhove
Walter Daelemans
Regional and linguistic situation
Regional and linguistic situation

Bilingual: French & Dutch

French

German
Social and medium-related variation in expressiveness
Social and medium-related variation in expressiveness

→ dependent variables: expressive markers (cf. later)
Social and medium-related variation in expressiveness

→ dependent variables: expressive markers (cf. later)

→ independent variables:
  - GENDER
  - AGE
  - MEDIUM
Research questions

**Number and nature** of expressive markers: differences?

- **GENDER:**
  - girls vs boys

- **AGE:**
  - younger (13-16) vs older teenagers (17-20)

- **MEDIUM:**
  - social media posts (public + async.) vs instant messaging (private + sync.)
Hypotheses

**Number of expressive markers:**

- **GENDER:** girls $>$ boys
- **AGE:** younger (13-16) $>$ older teenagers (17-20)
- **MEDIUM:** social media posts (public + async.) ??? instant messaging (private + sync.)
Hypotheses

**Number of expressive markers:**

- **GENDER:** girls > boys
- **AGE:** younger (13-16) > older teenagers (17-20)
- **MEDIUM:** social media posts (public + async.) ??? instant messaging (private + sync.)

⇒ less investigated
⇒ different hypotheses/results in related research
Corpus
informal CMC produced by Flemish adolescents
2007-2013

<table>
<thead>
<tr>
<th></th>
<th>YOUNG GIRLS</th>
<th>OLDER GIRLS</th>
<th>YOUNG BOYS</th>
<th>OLDER BOYS</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>IM</td>
<td>118694</td>
<td>176233</td>
<td>29146</td>
<td>973061</td>
<td>1297134</td>
</tr>
<tr>
<td>SOC. MEDIA</td>
<td>463277</td>
<td>67257</td>
<td>162077</td>
<td>76776</td>
<td>769387</td>
</tr>
<tr>
<td>TOTAL</td>
<td>581971</td>
<td>243490</td>
<td>191223</td>
<td>1049837</td>
<td>2066521</td>
</tr>
</tbody>
</table>
The independent variables: sample size
The independent variables: sample size

![Pie chart showing age distribution with younger and older categories]
The independent variables: sample size

![Pie chart showing sample sizes]

- Asynchronous: 1297134
- Synchronous: 769387
The dependent variables: expressive markers

Wide range – different types of features:

- 6 typographic chatspeak features
- 1 onomatopoeic variable
- 1 lexical variable
Typographic markers:

1. emoticons:  
   *dude :P*

2. delib. repetition of letters  
   *suuuper*

3. delib. repetition of punctuation marks  
   *nice!!!*
Typographic markers:

4. Combination question and exclamation marks
   *wtf?!?

5. Allcaps
   *FAIL*

6. Kisses
   *Xxxx*
Onomatopoeic marker:

7. Onomatopoeic rendering of laughter

*hahahaha*
Lexical marker:

8. Intensifiers

Supermooie t-shirt  ‘super nice T-shirt’
Data processing

(1) automatic feature extraction with Python scripts

(2) evaluation of coverage:
- test set of 1000 random posts (5595 tokens)
- comparison software’s output vs. manual decisions ➡️ accurate!
Results: quantitative

General degree of expressiveness (%) = all expressive features/all tokens

<table>
<thead>
<tr>
<th>Medium</th>
<th>PublicAsync/Young/Girls</th>
<th>PrivateSync/Older/Boys</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>28.3</td>
<td>5.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>PublicAsync/Young/Girls</th>
<th>PrivateSync/Older/Boys</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25.2</td>
<td>7.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>PublicAsync/Young/Girls</th>
<th>PrivateSync/Older/Boys</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>21.8</td>
<td>9.3</td>
</tr>
</tbody>
</table>
Results: quantitative

General **degree of expressiveness (%)**
= all expressive features/all tokens

![Bar chart showing expressiveness by medium, age, and gender.](chart)

- **Gender**
  - PublicAsync/Young/Girls: 21.8%
  - PrivateSync/Older/Boys: 9.3%

- **Age**
  - PublicAsync/Young/Girls: 25.2%
  - PrivateSync/Older/Boys: 7.7%

- **Medium**
  - PublicAsync/Young/Girls: 28.3%
  - PrivateSync/Older/Boys: 5.9%

Significant!
**Results: quantitative**

**Cramer’s V: correlation strength** (between independent variables & expressive markers)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEDIUM</td>
<td>0.31</td>
</tr>
<tr>
<td>AGE</td>
<td>0.24</td>
</tr>
<tr>
<td>GENDER</td>
<td>0.17</td>
</tr>
</tbody>
</table>

**Odds ratio: effect size** (size of the difference between the subgroups)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEDIUM</td>
<td>6.27</td>
</tr>
<tr>
<td>AGE</td>
<td>4.02</td>
</tr>
<tr>
<td>GENDER</td>
<td>2.71</td>
</tr>
</tbody>
</table>
**Results:** quantitative

Largest effect size (OR) + strongest correlation (CV):

*delib. letter repetition* and *kisses*

→ Most striking:

medium: kisses odds’ ratio **51.85**
Results: qualitative

Constants across gender, age groups and media:

- emoticons & delib. repetition of punctuation: most popular
- delib. letter repetition: mainly vowels, hardly ever plosives
Results: qualitative

Correlations between the social variables!
Results: qualitative

Correlations between the social variables!

GENDER: girls vs boys
AGE: younger vs older
MEDIUM: social media posts vs instant messaging
(public, async.) (private, sync.)
Results: qualitative

Correlations between the social variables!

GENDER: girls vs boys
AGE: younger vs older
MEDIUM: social media posts (public, async.) vs instant messaging (private, sync.)
Results: qualitative

Correlations between the social variables!

GENDER: girls vs boys
AGE: younger vs older
MEDIUM: social media posts (public, async.) vs instant messaging (private, sync.)

→ Much more expressive markers related to love and friendship

Top emoticons: <3
Top allcaps: LOVEYOU, BFF
Results: qualitative

Correlations between the social variables!

GENDER: girls vs boys
AGE: younger vs older
MEDIUM: social media posts (public, async.) vs instant messaging (private, sync.)

→ Much more expressive markers related to love and friendship

→ ← boys’ texts: heart-emoticons least popular variant!
Results: qualitative

Correlations between the social variables!

→ Possible explanations
Related research: *age* & *gender*

- **Style:** different use of function words
  - Men + older: articles, prepositions
  - Women + younger: pronouns, conjunctions, aux. verbs

- **Content:** different topics
  - Men + older: politics, religion, business
  - Women + younger: home, romance, fun

(Pennebaker 2011: 66; Argamon et al. 2007: n.pag.; Schwartz et al. 2013: 8-9)
Related research: **age & gender**

- **Style:** different use of function words
  - Men + older: articles, prepositions
  - Women + younger: **pronouns**, conjunctions, aux. verbs

- **Content:** different topics
  - Men + older: politics, religion, business
  - Women + younger: **home**, **romance**, fun

(Pennebaker 2011: 66; Argamon et al. 2007: n.pag.; Schwartz et al. 2013: 8-9)
Related research: age & gender & medium
Results: qualitative

Correlations between the social variables!

→ Possible explanations
  - Age & gender: related research
  - Medium: ?
Correlations between the social variables!

→ Possible explanations
- Age & gender: related research
- Medium: *nature and function of the texts/platforms*
Nature of the texts:

<table>
<thead>
<tr>
<th>Social media posts</th>
<th>Instant messaging</th>
</tr>
</thead>
<tbody>
<tr>
<td>public</td>
<td>private</td>
</tr>
<tr>
<td>asynchronous</td>
<td>synchronous</td>
</tr>
<tr>
<td>more neutral, formal</td>
<td>more personal, informal</td>
</tr>
<tr>
<td>largely personal character</td>
<td></td>
</tr>
<tr>
<td>➔ informing</td>
<td></td>
</tr>
<tr>
<td>➔ bonding, pleasing, flirting</td>
<td></td>
</tr>
</tbody>
</table>
Conclusion

**GENDER**: girls more expressive than boys

= to be expected

⇒ Expressive markers = tools for stressing emotional and social involvement
Conclusion

**AGE:** younger adolescents more expressive than older adolescents

= to be expected

Expressive markers = tools for identity construction, self-profiling, peer group communication ...
Conclusion

**MEDIUM**: more expressive writing in the social media posts

= to be expected??
Conclusion

**MEDIUM**: more expressive writing in the social media posts

Expressive markers = tools for pleasing and flirting, acquiring social prestige ...
Conclusion

**MEDIUM**: more expressive writing in the social media posts
+ strongest determinant

- Expressive markers = tools for pleasing and flirting, acquiring social prestige ...
- **Medium is important!**
Conclusion

**MEDIUM**: more expressive writing in the social media posts
+ strongest determinant
+ correlation gender and age

- Expressive markers = tools for pleasing and flirting, acquiring social prestige ...
- Medium is important!
- Goal/nature of the interaction is important!
References


Schwartz, A. et al. (2013): Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. PLoS ONE 8(9): e73791