Linguistic Characteristics of Dutch Computer-Mediated Communication: CMC and School Writing Compared

Lieke Verheijen
Introduction

**CMC**: communication via modern technologies

**Previous research** on written CMC:
English, German, French, Italian, Spanish, Portuguese, Finnish, Swedish, …

**Differences** from standard language conventions
(e.g. Thurlow & Brown 2003, Crystal 2008, Frehner 2008, Cougnon & Fairon 2014)

- **nonstandard orthography**: *fyi i’ll B @home l8er 2night, r u OK with that? :-)*
- **syntactic omissions**: *car broken down, mailed garage yesterday, haven’t responded yet*
YOU GUYS SERIOUSLY DON'T CHECK E-MAIL??

BTW, CAN YOU-

WAIT.

CAN YOU TWEET THE HOMEWORK @ US?

WAIT!

OR SNAP CHAT IT?

DID YOU JUST SAY “B-T-W” OUT LOUD?

WAIT!

INSTEAD OF “BY THE WAY”?

OH, YEAH.

IT ACTUALLY HAS MORE SYLLABLES!!

RLY?

OMG!

OBVS.
Research questions: What does the language used by Dutch youths in CMC actually look like? How does it differ from Standard Dutch?

Research goal: explore how Dutch youths’ informal written CMC linguistically differs from their more formal school writings
Previous studies on how CMC affects literacy

Mixed results...
Explanation: many differences in methodology & participants
(Verheijen 2013)
Materials

CMC writings

- Instant messaging with Internet application: **MSN**
- Text messaging (SMS)
- Microblogging: **Twitter**
- Instant messaging with mobile phone app: **WhatsApp**

Selected from existing corpus (SoNaR: *STEVIN Nederlandstalig Referentiecorpus*, Oostdijk et al. 2013)

- ages: 12-23

School writings

- lower & higher educational levels
- adolescents & young adults
### Materials: CMC writings

#### Corpus of CMC texts so far

<table>
<thead>
<tr>
<th>Genre</th>
<th>Year(s) of collection</th>
<th>Age group 1 (12-17)</th>
<th>Mean age</th>
<th># words</th>
<th># chats or contributors</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSN</td>
<td>2009-2010</td>
<td>16.2</td>
<td>45,051</td>
<td>106</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>18-23</td>
<td>19.5</td>
<td>4,056</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>total:</td>
<td>49,107</td>
<td>127</td>
<td></td>
</tr>
<tr>
<td>SMS</td>
<td>2011</td>
<td>15.4</td>
<td>1,009</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>18-23</td>
<td>20.4</td>
<td>23,790</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td></td>
<td>total:</td>
<td>24,799</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>Twitter</td>
<td>2011</td>
<td>15.9</td>
<td>22,968</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>18-23</td>
<td>20.6</td>
<td>99,296</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>total:</td>
<td>122,264</td>
<td>108</td>
<td></td>
</tr>
<tr>
<td>WhatsApp</td>
<td>2015</td>
<td>14.4</td>
<td>55,865</td>
<td>11 / 84</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>18-23</td>
<td>20.1</td>
<td>140,134</td>
<td>23 / 132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>total:</td>
<td>195,999</td>
<td>34 / 216</td>
<td></td>
</tr>
</tbody>
</table>

**grand total:** 392,169

*No. of chats: MSN, WhatsApp; no. of contributors: SMS, tweets, WhatsApp*
## Materials: school writings

### Corpus of school essays

<table>
<thead>
<tr>
<th>Educational level</th>
<th>Year(s) of production</th>
<th>Age group</th>
<th># words</th>
<th># texts</th>
</tr>
</thead>
<tbody>
<tr>
<td>lower secondary education ('vmbo')</td>
<td>2013–2014</td>
<td>± 14-15, 3rd grade</td>
<td>50,143</td>
<td>128</td>
</tr>
<tr>
<td>higher secondary education ('vwo')</td>
<td>2013–2014</td>
<td>± 14-15, 3rd grade</td>
<td>50,070</td>
<td>153</td>
</tr>
<tr>
<td>lower tertiary education ('mbo')</td>
<td>2012–2014</td>
<td>± 17-18, 2nd grade</td>
<td>39,793</td>
<td>137</td>
</tr>
<tr>
<td>higher tertiary education ('uni')</td>
<td>2012–2014</td>
<td>± 18-19, 1st grade</td>
<td>50,175</td>
<td>169</td>
</tr>
<tr>
<td><strong>total:</strong></td>
<td></td>
<td></td>
<td><strong>190,181</strong></td>
<td></td>
</tr>
</tbody>
</table>
Method

Register analysis: quantitative study of linguistic features

1) manual analysis: CMC writings

Orthographic features
- textisms
- misspellings
- typos
- emoticons
- symbols

Syntactic feature
- omissions

Lexical features
- borrowings
- interjections

2) automatic analysis: CMC writings vs. school writings, T-Scan (Pander Maat et al., 2014)

Lexical measures
- lexical diversity (MTLD)
- density of ‘special’ words
- lexical density
- density of ellipses

Syntactic complexity measures
- average of all dependency lengths
- average no. of subordinate clauses
- average sentence length
- D-level
Statistics

independent t-tests

(one-tailed probability values reported)
Results + discussion:

automatic analysis with T-Scan, comparing CMC texts to school essays
MTLD: measure of textual lexical diversity
= avg. length of sequential word strings in a text that maintain a TTR above a specified threshold (McCarthy & Jarvis 2010)

TTR = type-token ratio = no. of types [different words] / no. of tokens [total number of words]

assumption: higher MTLD value → more lexical diversity → more different(ly spelled) words

result: CMC texts ($M = 119.62$, $SE = 14.39$) > school essays ($M = 76.10$, $SE = 2.23$)
$t(10) = -2.08$, $p < 0.05$

explanation: textisms, typos, misspellings in CMC

hypothesis: confirmed

Density of ‘special words’
= no. of ‘special words’ (character strings T-Scan cannot recognize as words) per 1,000 words

assumption: higher density of ‘special words’ → more unrecognizable words → more differently spelled words + non-words

result: CMC texts ($M = 140.77$, $SE = 33.20$) > school essays ($M = 28.58$, $SE = 4.02$)
$t(10) = -3.35$, $p < .01$

explanation: textisms, typos, misspellings, URLs in CMC

hypothesis: confirmed
Lexical measures cont’d

**Lexical density**
= no. of content words (nouns, verbs, adjectives, adverbs) per 1,000 words (e.g. Johansson 2008)

**assumption:** higher lexical density → more content words → fewer function words

**Density of ellipses**
= no. of finite verbs without a subject per 1,000 words

**assumption:** higher density of ellipses → fewer grammatical subjects

**result:** CMC texts \((M = 531.70, \ SE = 9.28) >\) school essays \((M = 481.31, \ SE = 2.68)\)

\[ t(10) = -3.71, \ p < .01 \]

**explanation:** omission of function words in CMC

**hypothesis:** confirmed

**result:** CMC texts \((M = 25.86, \ SE = 3.17) >\) school essays \((M = 8.60, \ SE = 1.18)\)

\[ t(10) = -5.10, \ p < .001 \]

**explanation:** omission of subjects in CMC

**hypothesis:** confirmed
Syntactic measures

Average of all dependency lengths
= avg. no. of words that need to be skipped from head to dependent per sentence
dependency length = distance between head (of sentence/phrase) and its dependent

assumption: lower avg. of all dependency lengths → fewer discontinuous structures → less syntactic complexity (Gibson, 2000)

result: CMC texts ($M = 0.63$, $SE = 0.06$) < school essays ($M = 1.59$, $SE = 0.10$)
$t(10) = 9.04$, $p < .001$

hypothesis: confirmed

Average number of subordinate clauses
= avg. no. of subclauses (relative clauses, adverbial clauses, complement clauses, infinitival subclauses) per sentence

assumption: lower avg. no. of subclauses → less syntactic complexity

result: CMC texts ($M = 0.14$, $SE = 0.02$) < school essays ($M = 0.80$, $SE = 0.06$)
$t(10) = 10.21$, $p < .001$

hypothesis: confirmed
Syntactic measures cont’d

**Average sentence length**
= avg. no. of words per sentence

**assumption:** lower avg. sentence length → less syntactic complexity

**result:** CMC texts ($M = 6.55$, $SE = 0.28$) < school essays ($M = 16.33$, $SE = 0.79$)
$t(10) = 14.76$, $p < .001$

**hypothesis:** confirmed

**D-level: developmental level**
= based on classification and rank order of sentence types in eight increasingly complex developmental levels

(Rosenberg & Abbeduto 1987, Covington 2006)

**assumption:** lower D-level → less syntactic complexity

**result:** CMC texts ($M = 0.88$, $SE = 0.08$) < school essays ($M = 2.87$, $SE = 0.10$)
$t(10) = 15.51$, $p < .001$

**hypothesis:** confirmed
Conclusion

- Compared to school essays, written CMC:
  - **lexis** > is more diverse, different, dense
  - **syntax** > contains more omissions; is less complex

- Different **registers** → informal CMC vs. more formal school writing

- **Hopeful results**: no great interference of CMC with youths’ traditional writing skills after all...?
Future work

Corpus analysis
Next steps:
- Analyzing Facebook posts

Correlational study

RQ: Does youths’ CMC use (intensity/manner) correlate with writing proficiency?
✓ Conducting surveys on CMC use at secondary & tertiary schools
✓ Collecting school writings of the same students
Next steps:
- Analysing school writings qualitatively/quantitatively
- Computing correlations between answers on surveys & school writings

Experimental study

RQ: Is there a causal connection between CMC use and literacy?
References


Thanks for your attention

Questions or comments?

lieke.verheijen@let.ru.nl