

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

Lisa Hilte

Reinhild Vandekerckhove
Walter Daelemans



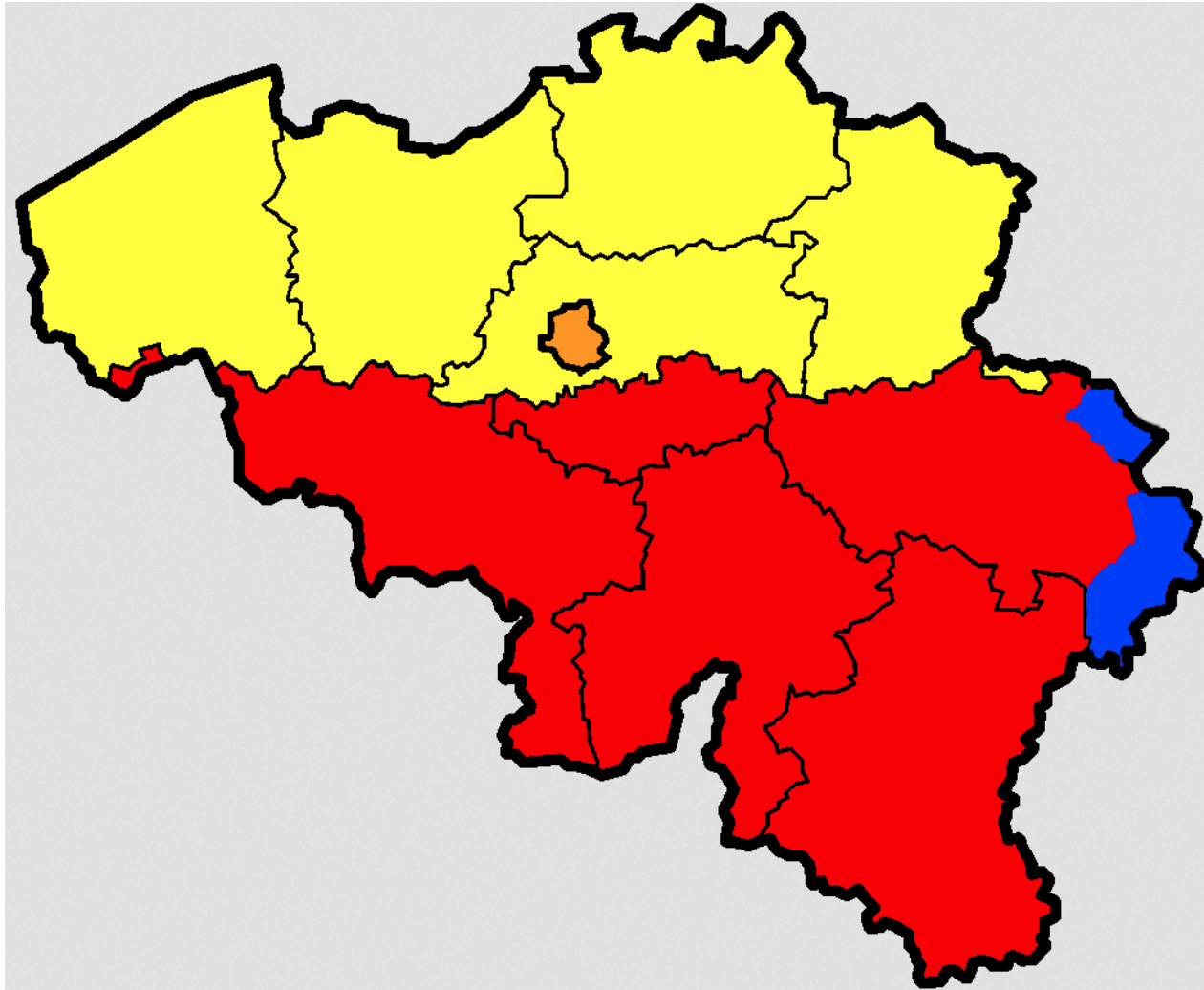
CLiPS

Computational Linguistics & Psycholinguistics
University of Antwerp

Regional and linguistic situation



Regional and linguistic situation



Dutch

Bilingual: French & Dutch

French

German

Topic

Social and medium-related variation in expressiveness

Topic

Social and medium-related variation in expressiveness

→ **dependent variables:** expressive markers (cf. later)

Topic

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

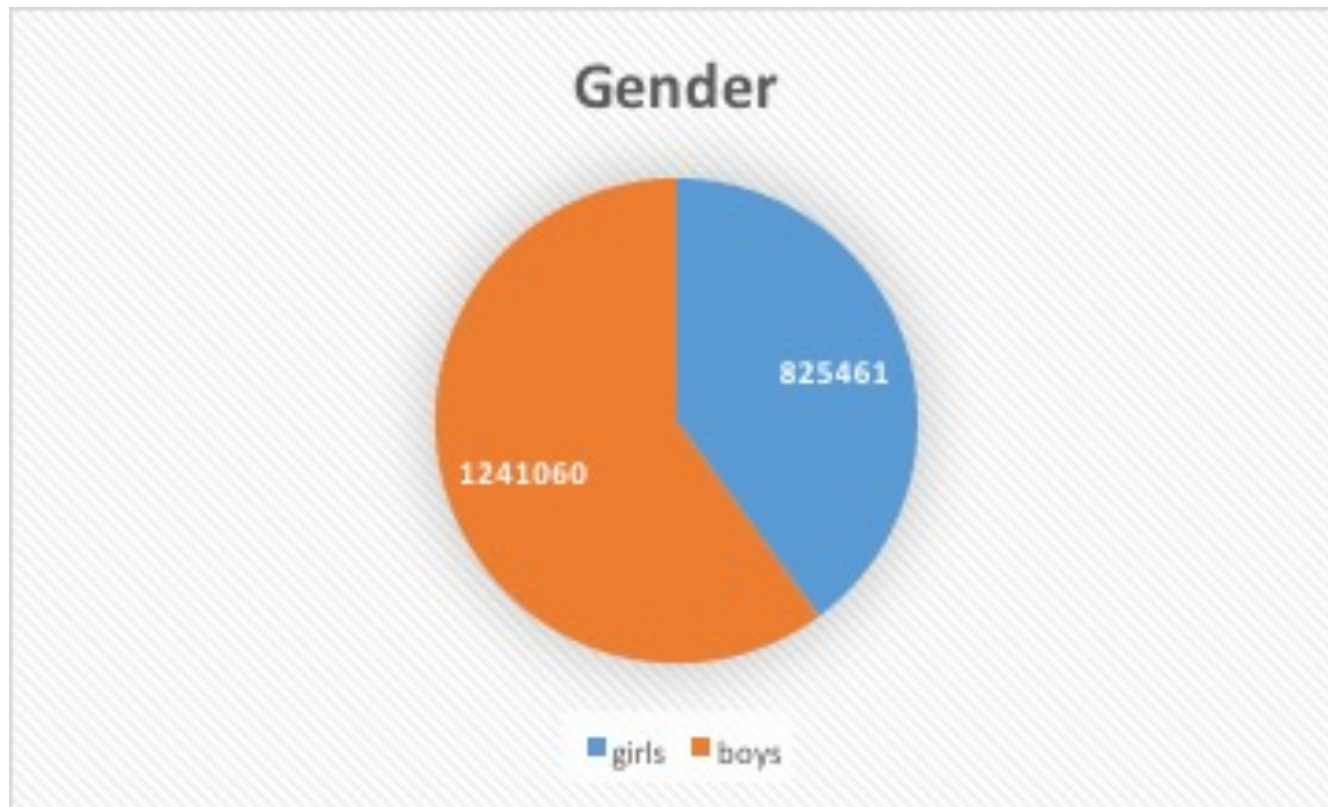
→ different hypotheses/results in related research

Corpus

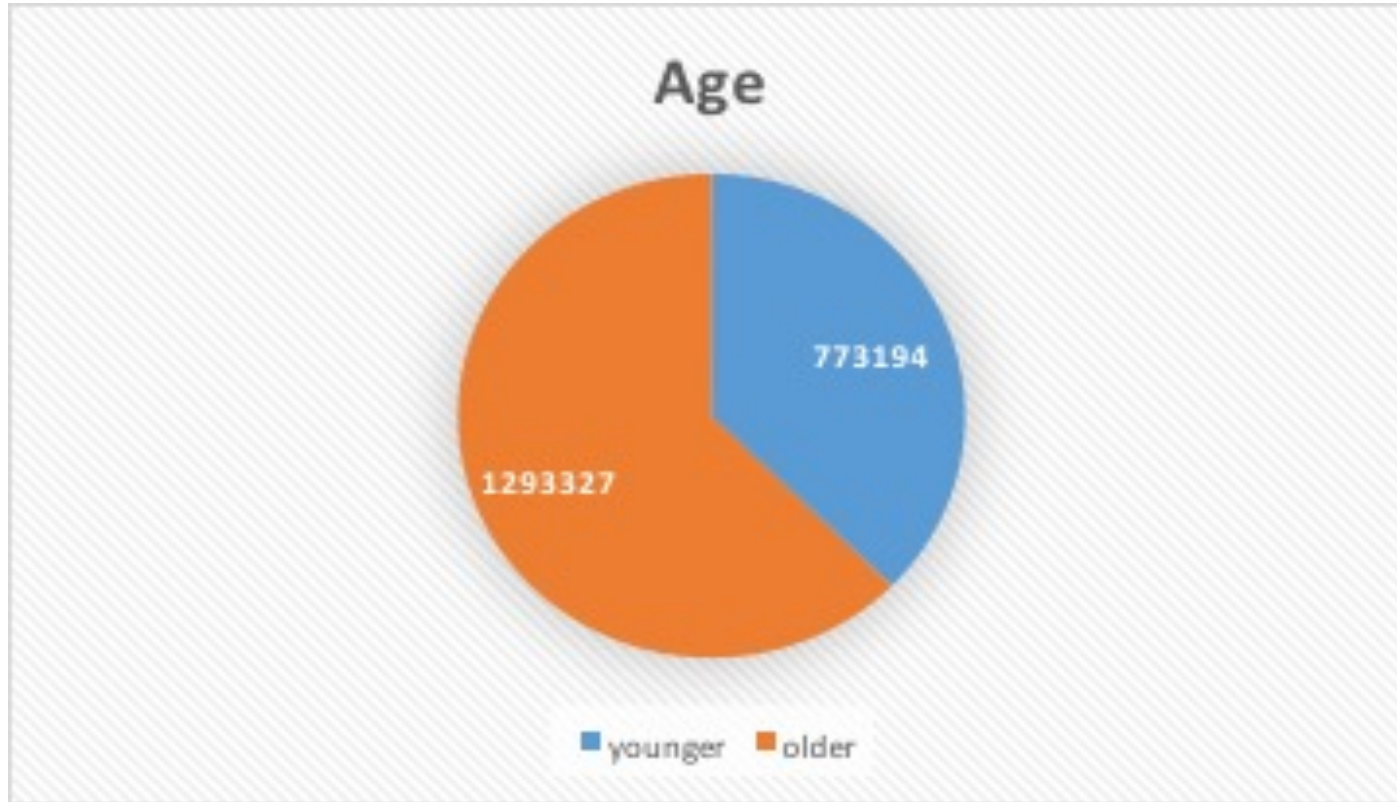
informal CMC produced by Flemish adolescents
2007-2013

	YOUNG GIRLS	OLDER GIRLS	YOUNG BOYS	OLDER BOYS	TOTAL
IM	118694	176233	29146	973061	1297134
SOC. MEDIA	463277	67257	162077	76776	769387
TOTAL	581971	243490	191223	1049837	2 066 521

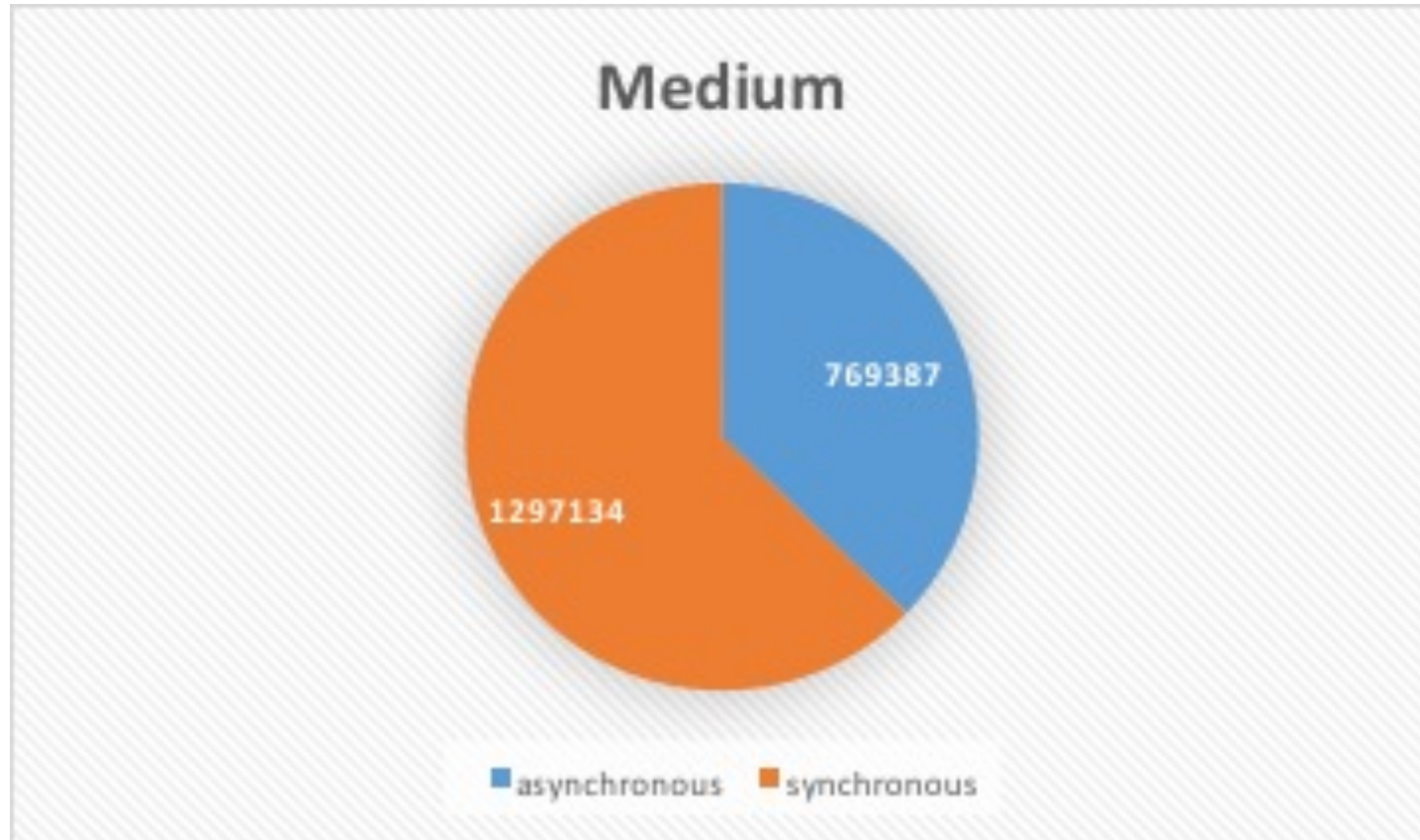
The independent variables: sample size



The independent variables: sample size



The independent variables: sample size



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'

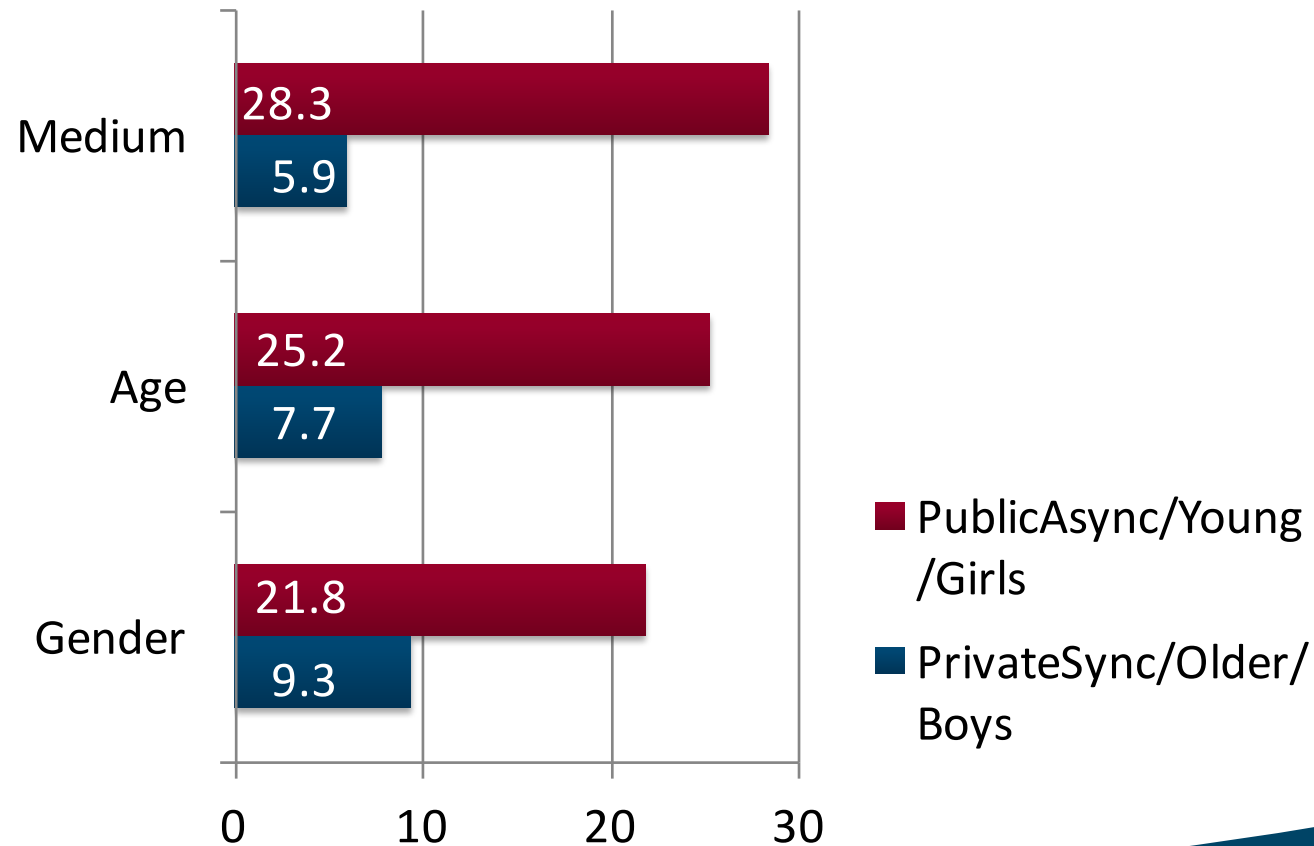
Data processing

- (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 (%)

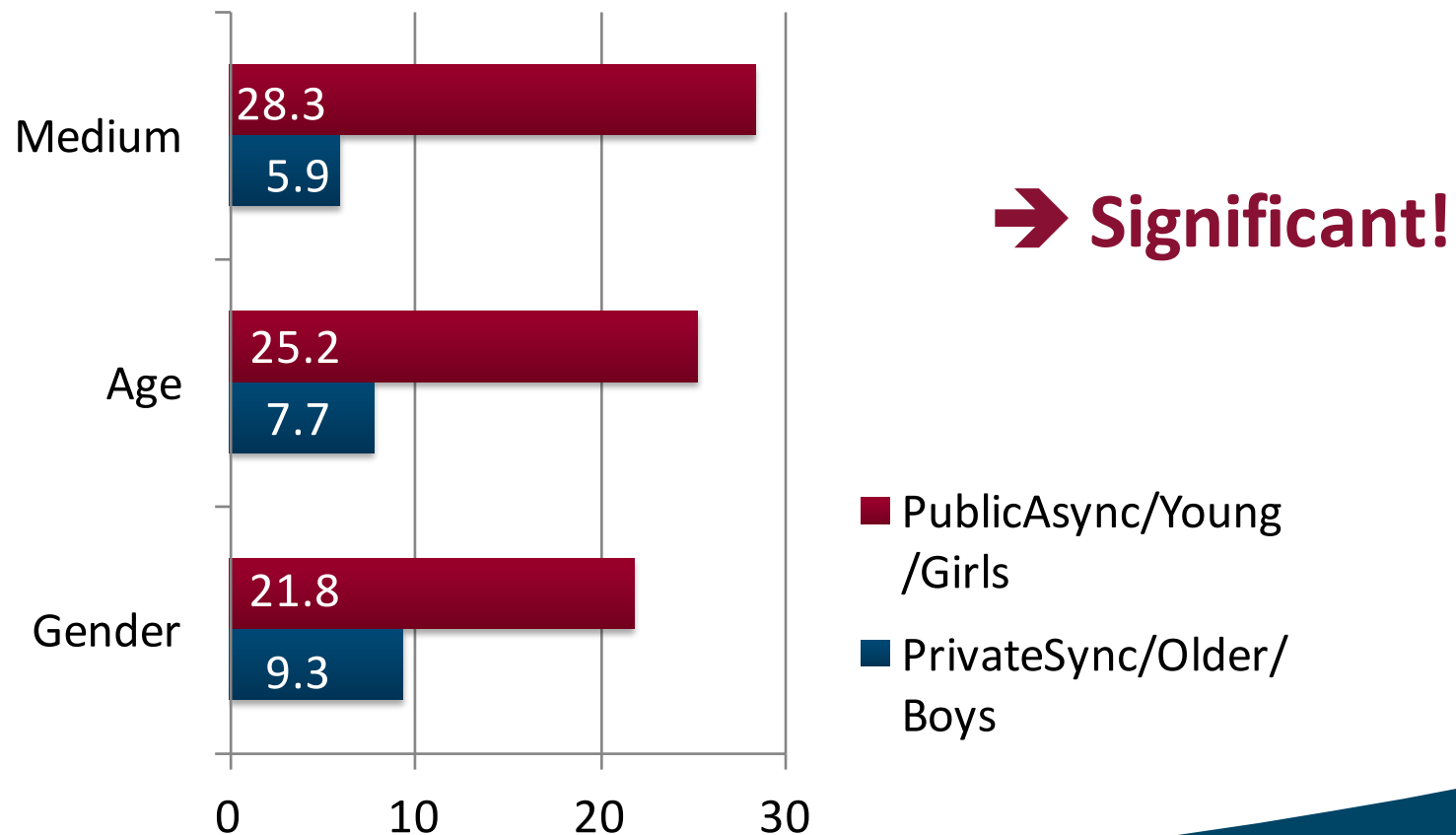
= all expressive features/all tokens



Results: quantitative

General degree of expressiveness (%)

= all expressive features/all tokens



Results: quantitative

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

MEDIUM	0.31
AGE	0.24
GENDER	0.17

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

MEDIUM	6.27
AGE	4.02
GENDER	2.71

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 (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.)

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: ?

Results: qualitative

Correlations between the social variables!

→ Possible explanations

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

Nature of the texts:

Social media posts

public

asynchronous

~~more neutral, formal~~

largely personal character

→ **informing**

→ **bonding, pleasing, flirting**

Instant messaging

private

synchronous

more personal, informal

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

Argamon, S., Koppel, M., Pennebaker, J.W., & Schler, J. (2007). Mining the Blogosphere: Age, Gender and the Varieties of Self-Expression. *First Monday*, 12(9), n.pag.

Pennebaker, J.W. (2011). *The Secret Life of Pronouns. What Our Words Say About Us*. New York: Bloomsbury Press.

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