Expressive markers in online teenage talk

A correlational analysis

Lisa Hilte, Reinhild Vandekerckhove and Walter Daelemans*

NT 23 (3): 293–323
DOI: 10.5117/NEDTAA2018.3.003.HILT

Abstract

This paper discusses the expression of emotional involvement in informal computer-mediated communication (CMC). While related research is quite fragmentary through its exclusive focus on a limited number of expressive markers or the inclusion of just one independent variable, the present study includes a wide range of expressive markers and three independent variables. The data reveal strikingly consistent age and gender correlates across all expressive markers and a strong correlation between the preferences of younger adolescents and girls. Furthermore, the study highlights a major impact of medium type. It calls for a refinement of the operationalization of the variable medium, as apart from its inherent characteristics (private/public, synchronous/asynchronous), the nature and goal of the interaction (which is also partly related to the type of social media that people use) trigger specific linguistic practices.

Keywords: CMC, youth language, expressive markers, sociolinguistics, gender, age, medium

* We thank Giovanni Cassani and Dominiek Sandra for their help and advice in the statistical aspect of the paper. We are also grateful towards Jens Vercammen for the manual data processing of the lexical variable. Finally, we thank the editorial board of *Nederlandse Taalkunde*/*Dutch Linguistics* and the two anonymous reviewers for their helpful and constructive feedback on an earlier draft of this paper. This work was supported by the FWO (Research Foundation Flanders).

1 Introduction

Since the rise of informal computer-mediated communication (CMC), both laymen and linguists have been fascinated by the prototypical features of several forms of digital writing (see Crystal 2001). Androutsopoulos (2011: 149) relates these features to three dimensions (also called maxims or principles): orality, compensation, and economy. While orality refers to the use of spoken language features in written discourse[1] and economy covers all strategies to shorten messages, the 'semiotics of compensation' 'includes any attempt to compensate for the absence of facial expressions or intonation patterns' (Baron 1984: 125 as cited in Androutsopoulos 2011: 149). De Decker & Vandekerckhove (2017: 278) stress the importance of making a distinction between economical and expressive chatspeak features in CMC research, as both groups of features appear to correlate differently with the variables of age, gender and medium. While they found age and medium correlates for several chatspeak features, they did not identify significant gender patterns, except for the only expressive variable that was part of their analyses. Consequently, the authors concluded that '[their] findings call for further refinement of the operationalization of emotional expressiveness in CMC and a broader selection of expressive markers'. The present paper, which focuses exclusively on that type of markers in Flemish online teenage talk, meets these requirements. It does not only include typographic features that are prototypically associated with the maxim of compensation, such as emoticons or the capitalization of words and utterances, but also a lexical and an onomatopoeic variable – namely the use of intensifiers and the onomatopoeic rendering of laughter (e.g., *haha*). The notion of expressiveness is thus used as a cover term for the expression of (strong) involvement, in most cases emotional involvement. The following example contains four of the eight features that function as the dependent variables in the present study (i.e. the onomatopoeic *hahaha*, the capitalization of *super*, repetition of the exclamation mark and the emoticon *:D*):

1 For a more elaborate and nuanced view on the dichotomy between written and spoken language, we refer to Koch & Oesterreicher (2001: 584-585; 2011: 3-4), who take both *Medium* ('realization': either phonic or graphic) and *Konzeption* ('register': spoken/informal register or written/formal language) into account to create four combinations on what they call the *continuum* between spoken and written language. In the case of informal CMC, the medium may be a written medium, but the discourse is often to a large extent conceptually oral (see also Schlobinski 2005).

(1) Hahaha SUPER!!! :D

Our main research question relates to the potential correlation between the use of the selected expressive markers and the sociolinguistic profile of the chatters. All informants are adolescents from Dutch-speaking northern Belgium, i.e. Flanders. The social variables operationalized in the present study are their age and gender. The main goal is to identify the most expressive subgroup: do women and younger adolescents outperform men and older adolescents respectively in the use of expressive markers, or do these groups show distinct preferences for specific expressive markers? These research questions are inspired by the related research that will be discussed in section 3. Apart from age and gender, section 3 also discusses the potential impact of different digital media. Since our data contain both largely public asynchronous and private synchronous online messages, this variable had to be included in the research design. Moreover, the combination of these three variables distinguishes the present study from much of the related research. Before discussing potential determining factors, we will present the expressive markers themselves and previous literature on each of them (section 2). Section 4 is devoted to the experimental setup: it describes the corpus, the participants and the methodology of the data extraction and processing. The following section (5) contains the results of the analyses and the final one (6) presents the conclusion.

2 The expressive markers

There are many ways of expressing emotional involvement, both in speech and in written language. The most obvious way of doing so is by literally articulating emotions, e.g.: 'I feel sad'. In many cases, however, feelings and emotions are expressed in a more indirect way, for example through particular facial expressions. The absence of such facial expressions, but also of other forms of body language (e.g. hand gestures), of voice volume and pitch in textual computer-mediated communication leads to the compensatory strategies which we referred to above (see also Thurlow & Poff 2013: 176, who use the term *paralinguistic restitution*, and Kucukyilmaz, Cambazogly, Aykanat & Can 2006: 276). These compensatory typographic features represent the majority of the expressive markers that we selected for the present study. We refer to them as the *expressive chatspeak features* and discuss them in section 2.1. Section 2.2 concerns the onomatopoeic rendering of laughter, which is not a typographic feature but can be considered typical of chatspeak too. Section 2.3, finally, presents a lexical feature

which is not typical of CMC, but which certainly can be considered a marker of expressiveness that functions in much the same way as some of the typographic markers, i.e. the use of intensifiers.

2.1 Typographic expressive markers: expressive chatspeak features

Androutsopoulos (2011: 149) distinguishes several compensational features: 'emoticons, abbreviations that signify various types of laughter, simulation of expressive prosody by iteration of letters and punctuation'. All of these typographic markers are included in the present study, but we added two more: capitalization of entire words or utterances and the use of the letter(s) *x* or *xo*[2] (or several instances of both) to symbolize kisses versus hugs and kisses respectively. In the next paragraphs, we will briefly discuss each of these features.

The first marker in the present research design is so-called *flooding*:[3] the deliberate repetition of letters or punctuation marks (both are present in example 2).

(2) ik ben s**uuuuu**per hyper**!!!!**
'I am s**uuuuu**per hyper[active]**!!!!**'

Flooding can be interpreted as a way of symbolically emphasizing a word (letter flooding) (De Decker & Vandekerckhove 2017: 265) or an entire utterance (punctuation flooding). Parkins (2012: 52) states that letter flooding serves both expressiveness and creativity: 'The manipulation of letters, such as the repetition of a certain vowel or consonant, can be used creatively in many situations to represent emotional stances such as pondering, disappointment, doubt, frustration, sarcasm, and happiness'. As for punctuation flooding in particular, she adds that it is used 'to indicate a degree of intensity in what the author had to say' (2012: 50), rather than for grammatical purposes, as is the case for standard punctuation. Unlike De Decker & Vandekerckhove (2017: 265), we make a distinction between the repetition of letters and the repetition of punctuation marks. For letter

2 We note that single occurrences of *xo* are rather ambiguous, as they could be used (and perceived) as both a kiss and a hug, or as an emoticon representing a facial expression with an open mouth. We opted for the first interpretation, but that might be the wrong choice in some cases. However, as these occurrences are extremely rare in the corpus (only 0.4% of all kisses, and less than 0.01% of all tokens) and their impact consequently is negligible, we did not exclude them from the analyses.

3 Different terms are used to indicate the phenomenon of flooding, like *reduplication* (Verheijen 2015: 132), *additional letters* (Parkins 2012: 52) or *letter repetition* (Darics 2013).

flooding, we worked with a threshold of three or more[4] identical graphemes. Repetitions of the letter *x* were excluded, as they are generally used to render 'kisses' and thus serve a different function (see below). For punctuation flooding, we used a threshold of two or more repetitions and restricted the selection to question and exclamation marks.

Apart from punctuation flooding, combinations of question and exclamation marks (example 3) are also included as a distinct variable:

(3) wat**?!?**
'what**?!?**'

Another way to express emotion or involvement in written CMC is the use of unconventional capitalization. The most common and probably most expressive application consists in writing entire words or utterances in capital letters (also called *allcaps*), which seems to be a visual, typographic representation of shouting. The following extract from a conversation between two Flemish chatters corroborates this interpretation:

(4) chatter A: NIE ZO RAP KAN NI VOLGEN
'not so fast, I can't keep up'
chatter B: nie schreeuwe
'don't shout'

Just like shouting in a face-to-face conversation, capitalizing entire words in an online conversation often is intuitively perceived as an expression of anger. However, it can just as well express other emotions, such as excitement and happiness (Parkins 2012: 51):

(5) *ik zal morgen **ALLES** vertellen*
'tomorrow, I will tell **EVERYTHING**'

(6) ***IT WAS SO GOOD THOUGH!** I'll have to show you so you can buy it:P*
(Parkins 2012: 51, emphasis added)

Finally, it can be used as a more neutral emphasizer, which draws attention to parts of the utterance:

(7) *wie gaat er nu **ZEKER** mee?*
'who is coming along **FOR SURE**?'

4 We note that there is no 'rule' that decides which number of repetitions is needed for character repetition to be counted as flooding, nor is there the certainty that some occurrences of flooding were not just typed by mistake.

These cases of allcaps were included in the present study, but other unconventional ways of capitalization, like alternating upper and lower case letters (e.g. *hElLo* instead of *hello*, see Herring 2012: 2) were not, because they seem to have a primarily fun-oriented and creative function, rather than a strictly (emotionally) expressive one. For the detection of allcaps, we only selected words that contain more than one letter, in order to reduce noise.

Furthermore, emoticons (short for 'emotional icons', Wolf 2000: 828) or smileys are quite explicit expressive markers, as many of them literally are (typo-)graphic representations of facial expressions. Emoticons are very popular in CMC (Wolf 2000: 828). Parkins (2012: 52) even states that '[they] are the most frequently used prosodic features to express emotion online'. Originally, typographic characters (mainly punctuation marks) were combined to create a stylized image of a human face. Among these original smileys, both Western and Asian (also called Japanese) variants can be distinguished. The main difference is that the Western ones (examples 8 and 9) are rotated – one must tilt one's head to the left or sometimes to the right to read them (Wolf 2000: 828) –, whereas the Asian ones, called 'kaomoji', are not (examples 10 and 11).

(8) :) (smiling face)
:-O (surprised face)
;-) (winking face)
XD (face laughing, eyes closed)

(9) *sgoe* ***:)*** 'alright ***:)***'

(10) ^^ (closed, smiling eyes)
T__T (crying face, tears streaming from eyes)
o_O (surprised face, confused)
-__- (unamused face, frustrated)

(11) *kvin een papier nimeer -_-*
'I can't find a sheet of paper -_-'

Punctuation marks, letters, numbers and other symbols can also be combined to create images other than human faces, like (rotated) hearts:

(12) *ik mis em ook <**3***
'I miss him too <**3**'

These manually composed smileys are the oldest, i.e. first-generation emoticons. More recent than these traditional smileys are all kinds of Unicode or ASCII encodings which today are called emoji. Instead of actually composing the desired emoticon, the chatter simply selects emoji (as proper

images) from a list. In the present paper, we use the term emoticon as a cover term for both classic emoticons and expressive emoji. Some examples of the latter can be found below: 13 contains one that was bound to the data produced on the synchronous chat platform, whereas 14 contains one typical of the asynchronous data produced on a social media site (see section 3.3).

(13) *kheb toch gratis smse* 
'I can send free text messages anyway  '

(14) *Mrciii* 
'Thank you  '

Emoticons can express a whole range of feelings. Wolf (2000: 830) distinguishes the following categories: 'teasing/sarcasm,[5] humor, sadness, despair, confusion, to offer an apology, a positive feeling or thanks, or to express solidarity/support'. She adds a separate category for emoticons with an unclear or no apparent purpose.

Finally, we added a typographic feature which dates back to pre-digital times, but which, judging from the Flemish chat conversations, seems to enjoy a renewed and intense popularity nowadays: the use of one (or more) instances of the letter *x* (sometimes capitalized) to symbolize a kiss (or several kisses). Many adolescents do not only use this symbol at the end of their conversations, by way of greeting, but insert the *x*'s in their discourse continuously or quite frequently. For the sake of completeness, we also included the sequence *xoxo* (and variants: *xoxoxo, …*), which stands for 'hugs and kisses'. Examples are shown below.

(15) *hey snelle cv metj* ***xx***
'hey handsome, everything alright **xx**'

(16) *hey!!!* ***xoxo***

Summing up, these are the six typographic expressive markers that function as variables in the present study: (1) flooding of letters, (2) flooding of punctuation, (3) combinations of exclamation mark and question mark, (4) capitalization of words or entire utterances, (5) emoticons, (6) rendering of kisses or hugs and kisses.

5 Wolf (2000: 832) points out that 'whether [sarcasm and teasing] constitute an emotion is debatable'. We will not focus on that debate here, as the expression of sarcasm or teasing increases the overall expressiveness in CMC just as well as the expression of 'unambiguous' emotions does.

2.2 Onomatopoeic expressive marker: onomatopoeic rendering of laughter

An alternative for one of the most common emoticons, i.e. the smiling face, are the onomatopoeic utterances *haha* and *hihi* (and variants: *hahaha*, *whaha, hihihihi,* …). These utterances may not be prototypical chatspeak features, but for two reasons we decided to include them: first of all, they seem to be the equivalent of smileys that express laughter (see example 17). Secondly, they are fairly frequent in the Flemish corpus. Therefore, it seemed somewhat incongruent to include laughing smileys but exclude their onomatopoeic equivalents, so we chose to include both.

(17) ***Haha*** *Grappig profiel*
'**Haha** funny profile'

2.3 Lexical expressive marker: intensifiers

The concept of intensifiers is quite ambiguous. Symptomatic in this respect is the fact that there is no real consensus among linguists concerning the appropriate terminology. Some of the names and terms used in previous research are *intensives* (Stoffel 1901), *amplifiers* (Quirk, Greenbaum, Leech, Svartvik & Crystal 1985: 590), *maximizers* and *boosters* (Quirk et al. 1985: 591). We adopt both the terminology and the definition used by Stenström, Andersen & Hasund (2002: 139), and see intensifiers as 'items that amplify and emphasize the meaning of an adjective or adverb'. This definition captures both their function and their grammatical 'compatibility'. In Dutch, intensifiers can either be adverbs, as illustrated in example 18, or intensifying prefixes, as shown in example 19.

(18) *Auwtch daswel* ***heel*** *vroeg*
'Ouch that is **very** early'

(19) ***kei****schattig!!*
'**very** cute!!'

Intensifiers are not typical of computer-mediated communication. However, they can be considered markers of expressiveness and they often function in much the same way as the other expressive features. According to Peters (1994: 271), people mainly use intensifiers to captivate the interlocutor or reader by displaying linguistic creativity, and to express emotional involvement. Both functions apply to most of the other expressive features as well. Compare, for instance, an utterance like *you are BEAUTIFUL* with *you are so beautiful*. In the former utterance, the speaker stresses his involvement through the capitalization of the adjective, in the latter through

the insertion of the intensifier *so*. By using an intensifier, the speaker shows that his enthusiasm, disappointment, happiness, appreciation, etc. is not just moderate or mediocre, but intense. Typographic features like flooding and capitalization generally have the same effect.

Since we are focusing on the correlation between the frequency of intensifier use and authors' age, gender and medium, we will not be dealing with the actual appearance of the many variants, but we note that they are fascinating objects of linguistic study for several reasons, one of them being that they are very dynamic and marked by constant renewal and change (Quirk et al. 1985: 590, Pyles & Algeo 1993: 250, Peters 1994: 271, Méndez-Naya 2003: 372, Tagliamonte 2008: 391 and references therein). Moreover, they are often subject to delexicalization or grammaticalization, i.e. the process in which a word gradually loses lexical content but gains grammatical functionality (Partington 1993: 183, Lorenz 2002: 144).

3 The independent variables: gender, age and medium

In this section, we discuss the results of previous research on the linguistic impact of gender (3.1), age (3.2) and medium (3.3). We will focus on expressiveness and include both sociolinguistic and stylometric[6] research. Following the discussion of the related research, we will present our hypotheses (3.4).

3.1 Gender

Sociolinguistic and stylometric research reveal parallel tendencies with respect to patterns in male and female language[7] related to expressiveness. Female discourse is said to be more expressive and emotional, in offline (i.e. face-to-face) as well as in online communication (Jespersen 1922: 251, Wolf 2000: 831, Kucukyilmaz et al. 2006: 282, Parkins 2012: 48, 50, 53). These findings contradict hypotheses about 'online gender swapping', i.e. women and men adopting different roles in online communication than in face-to-face interaction and thus possibly communicating in new, non-stereotypical ways

6 Stylometry is a subdiscipline of computational linguistics: 'The basic research question for computational stylometry seems then to describe and *explain* the causal relations between psychological and sociological properties of authors on the one hand, and their writing style on the other' (Daelemans 2013: 1, emphasis in original).

7 Gender is generally reduced to a binary variable (male vs. female). For criticism of this approach and for alternative views, see Bing & Bergvall (1996) and Coates (1993).

(Wolf 2000: 827). While women are found to use more emotional language or language expressing social involvement – talking and writing more about personal, social and emotional processes like feelings and thoughts –, men appear to use more informative language – focusing more on specific facts, objects and events (Jespersen 1922: 251, Argamon, Koppel, Fine & Shimoni 2003: 323, 334, Baron 2008: 51, Newman, Groom, Handelman & Pennebaker 2008: 223, 229, 232-233, Argamon, Koppel, Pennebaker & Schler 2009, Schwartz et al. 2013: 9).

With respect to the expressive markers that are subject of the present study, women (or girls) have been found to use significantly more intensifiers than men (or boys) (Stenström et al. 2002: 142 and references therein). Apart from this quantitative discrepancy, a qualitative difference has been found as well, with men and women preferring different intensifiers (Tagliamonte & Roberts 2005: 289, Xiao & Tao 2007: 251, Tagliamonte 2008: 388). While teenage girls may be more expressive quantitatively, from a qualitative perspective, the teenage boys seem to outperform the girls as they opt more often for strong intensifiers (e.g. *extremely*) and taboo words (e.g. *fucking*) (Stenström et al. 2002: 139, 143).

Furthermore, CMC research generally reveals a higher frequency of emoticons in female utterances (Baron 2004: 415, Herring & Martinson 2004: 436, Kucukyilmaz et al. 2006: 282, Parkins 2012: 52, Schwartz et al. 2013: 8). Moreover, Wolf (2000: 833) points to a functional expansion of smileys in female discourse: 'Females have expanded on the male definition of emoticons and their use, adding other dimensions including solidarity, support, assertion of positive feelings, and thanks'. Huffaker & Calvert (2005), on the one hand, and Wolf (2000), on the other, however, challenge and nuance the findings concerning the gender-dependent rate of emoticon use. Huffaker & Calvert (2005: n.p.) report the opposite effect among adolescent chatters, i.e. boys using more emoticons than girls. Wolf's nuance concerns the interlocutors: she found that in mixed-gender conversations, 'both males and females display an increase in emoticon use', resulting in an insignificant gender difference (2000: 831-832). Moreover, her findings also reveal convergence with respect to the communicative function of the emoticons. According to Wolf, women mostly use smileys for humorous purposes, while men deploy them more for teasing or expressing sarcasm. In mixed-sex conversations, this difference is levelled out to some extent (Wolf 2000: 832). However, while the corpus for the present case study contains both mixed-sex and single-sex conversations, this variable was not included in the research design.

Finally, Parkins (2012: 48, 50-53) reports a higher frequency in online female communication for several of the expressive markers that are subject

of the present study: letter and punctuation flooding, capitalized text, emoticons and expressions of laughter. Varnhagen et al. (2010: 729) and Baron (2004) also report a higher frequency of typical chatspeak features and markers of emotional involvement in girls' CMC.

3.2 Age

As for the linguistic impact of age and adolescence, it is widely accepted that creativity, language innovation and non-standard language use peak during puberty (Eckert 1997: 163, Androutsopoulos 2005: 1499, De Decker 2014: 44, Peersman, Daelemans, R. Vandekerckhove, B. Vandekerckhove & Van Vaerenbergh 2016: 16-17). However, adolescence is no homogeneous linguistic period, since the so-called 'adolescence peak' tends to be situated at the ages of 15 and 16. The use of non-standard language is supposed to culminate at that age and to decrease as youngsters age (Wolfram & Fasold 1974 as mentioned in Eisikovits 2006: 42, Holmes 1992: 184, Coates 1993: 94, De Decker & Vandekerckhove 2017: 277). As for CMC specifically, younger teenagers are said to use more typical chatspeak features in their online messages than older adolescents (Tagliamonte & Denis 2008: 13). A possible explanation could lie in changing attitudes concerning deviations from the linguistic standard: whereas adolescents seem to consider them as cool and use them for 'belonging' as well as for identity construction (Verheijen 2015: 129, De Decker & Vandekerckhove 2017: 278), young adults might see these deviations as 'somewhat childish' (Verheijen 2015: 135).

In general, younger people's and particularly teenagers' language use is considered to be more expressive and emotionally loaded than that of the older generations: many of the (stylistic) innovations typical of adolescent talk are hypothesized to 'primarily serve expressive and interactive purposes' (Androutsopoulos 2005: 1499). Pennebaker (2011: 61-63) adds that younger people use more negative and fewer positive emotion words than older people. On a content-based level, teenagers often talk and write about how they feel (Argamon et al. 2009: n.p.). Quite surprisingly, however, adolescent speech is generally found to contain fewer intensifiers than adult language (Paradis 2000: 154, Stenström et al. 2002: 141, Pertejo & Palacios Martínez 2014: 218). Stenström et al. even report that in their corpus, 'the adults use intensifiers almost twice as frequently as the teenagers' (2002: 141). Paradis (2000: 154) ascribes this quantitative difference to a different choice of intensifying strategies. Yet intensifiers often function as a *groups binder* in adolescent peer groups: the use of a specific (set of) variant(s) can serve not only speaker but also *group* identification and signal in-group membership, at least until the variant becomes more widely popular

and gets picked up by other groups (Peters 1994: 271; Lorenz 1999: 24-25). Furthermore, research indicates qualitative differences between adolescents' and adults' use of intensifiers (Tagliamonte 2008: 388), with the former showing a greater preference for new, informal, regional and non-standard variants (Eckert 2003: 116; Androutsopoulos 2005: 1497).

CMC research suggests that teenagers generally use more stylistic (chatspeak) features than older chatters (Argamon et al. 2009: n.p., Goswami, Sarkar & Rustagi 2009: 215, Schwartz et al. 2013: 9). This also holds for some of the expressive markers included in the present study: they appear to be more frequent in teenagers' CMC than in older people's chat messages. Youngsters have been found to use more emoticons than adults (Argamon et al. 2009: n.p.; Schwartz et al. 2013: 9), while young adolescents apply more flooding than adolescents at the end of their teens (De Decker & Vandekerckhove 2017: 265).

Verheijen (2015, 2016) distinguishes two age groups: younger adolescents versus older adolescents or young adults. She reports that in instant messages, emoticons and unconventional spelling forms were used much more often by teenagers than by young adults (Verheijen 2015: 135-136; Verheijen 2016: 283, 285). Strikingly, the opposite effect was noted for emoticons in (telephone) text messages: young adults used more emoticons than adolescents (Verheijen 2016: 285).

3.3 Medium

The final independent variable relates to the medium on which the online communication took place. We distinguish four main types of CMC based on (the possible combinations of) two parameters: synchronicity of the medium and number of interlocutors[8] (see Table 1 for an overview).

Table 1 Different types of CMC (De Decker 2014: 3)

	One-to-one	One-to-many
Synchronous	Instant messaging with two interlocutors	Instant messaging with multiple interlocutors: group chats
Asynchronous	Email, private messages, texts, …	Public posts or reactions on social media or online fora

8 We note that other typologies are possible too, as the two selected parameters are not the only ones, nor are they necessarily the most influential ones for all phenomena or markers: e.g. the type of keyboard or electronic device – computer or mobile device such as smartphone or tablet – can have a large influence as well. However, our choice is determined by practical constraints, as these are the only parameters we have information on.

Synchronous CMC (instant messaging) consists of real-time chat sessions in which all interlocutors are online at the same time (Baron 2004: 298). In *asynchronous* CMC (or non-instant messaging), only the emitter is online and not the receiver, or at least not necessarily so (Herring 2001). Both types can contain one-to-one just as well as one-to-many messages.

Both the synchronicity of the medium and the public versus private nature of the communication can impact on language use. Different hypotheses can be found in related research, relying on different views on digital media platforms and different theories about the ease or automaticity with which people use standard language.

As for the impact of the synchronicity of the electronic medium, some linguists argue that people write in a more standard-oriented way on asynchronous platforms, as they experience less time pressure[9] than in synchronous communication and therefore have more time to check and edit their posts (Herring 2001: 617, Gheuens 2010: 17-18, Verheijen 2015: 134). Others, however, hypothesize that chatters might use the extra time in asynchronous posts for experimenting and linguistic innovation (De Decker 2014: 64, De Decker & Vandekerckhove 2017: 256).

As for the public versus private dimension, Verheijen (2015: 134) notes that the public (one-to-many) character of some asynchronous channels could encourage people to turn to more standard orthography, to avoid 'being chided for their spelling'. But De Decker & Vandekerckhove (2017) add that even though private conversations with close peers can be more comfortable, 'this need not imply that private interaction favors experimenting more than public interaction, since self-presentation on public networking sites might also be a trigger for creative language use' (256) and the use of chatspeak features 'might raise [youngsters'] personal attractiveness to outsiders' (277). The 'showing-off' function is also identified by Verheijen (2016: 289) who observes abundant use of English in public tweets of Dutch youngsters who enjoy demonstrating to a large audience how cool they are.

Verheijen (2015: 133-134) generally observed a strong impact of medium in Dutch online communication. Instant messages appeared to contain much more non-standard writing than text messages and tweets (microblogging). The latter had the lowest score for non-standard forms. De

9 Verheijen (2015: 129) notes that although the speed principle may not hold for asynchronous media, brevity can still be important, as some asynchronous genres have limited message size (e.g. tweets). In this paper, however, message size is no (sub)variable, as none of the medium variants represented in our corpus have limitations with respect to the length of the messages.

Decker & Vandekerckhove (2017: 277-278) call for a distinction between expressive or playful CMC features and highly functional economical spelling choices: abbreviations appeared to be more frequent in synchronous data, whereas the expressive marker of flooding scored higher in asynchronous interaction. With respect to intensifier use, Herring (2001: 617) observes that synchronous media trigger a higher frequency of intensifiers because communication there is less formal than on asynchronous media.

As may be deduced from the above, impact of the medium is hard to predict. Moreover, apart from the public versus private character of the medium and the degree of synchronicity, there are other determining factors, such as the above-mentioned formality of the interaction and the contents of the messages. There may be huge differences between several asynchronous media with respect to these parameters. The tweets analyzed by Verheijen (2015, 2016) are often quite neutral in terms of formality, but the asynchronous messages examined in the present study (and in De Decker 2014 and De Decker & Vandekerckhove 2017) have been extracted from a social media site which triggered quite personal and in most cases highly informal communication between youngsters (see section 4.1).

3.4 Hypotheses and research questions

The discussion of the related research in the previous sections leads to the following hypotheses: since girls are generally supposed to have a stronger focus on establishing social and emotional connections, we assume they will produce more expressive markers than boys. In view of the fact that the older adolescents (see 4.1) are beyond the adolescent peak period, whereas the younger ones are in the midst of it or heading towards it, we hypothesize that the younger adolescents will outperform the older group in the frequency of use of the expressive markers. Finally, we assume that both due to the importance of linguistic self-presentation in the selected public asynchronous media and due to the greater time pressure in the synchronous media, fewer expressive markers will be used in the synchronous chat conversations.

The strength of the present study lies in the fact that it combines several independent variables and includes a wide range of expressive markers. The former enables us to discover the relative strength of several factors: what variable displays the strongest correlation with the use of expressive markers? What about the relative impact of the others? What are the implications for future CMC research? Furthermore, the inclusion of several types of expressive markers allows for a more detailed analysis of the preferences for specific markers by particular groups or in particular media.

For instance, irrespective of the potential gender differences with respect to the overall frequency of the expressive markers, boys and girls might display distinct preferences with respect to choice or even realization of particular markers. Do these findings corroborate or nuance the overall age, gender or medium preferences?

4 Methodology

In this section, we will describe our corpus (section 4.1) and the data extraction and processing (section 4.2).

4.1 Corpus and participants

The corpus contains CMC data produced between 2007 and 2013 by Flemish adolescents aged 13 to 20.[10] So some adolescents are in their late teens or even rather young adults. We take into account this discrepancy between young teenagers and adolescents nearing adulthood when dealing with the variable age. Furthermore, all of them are Dutch-speaking teenagers living in the north of Belgium. The entire corpus consists of 400 808 posts (i.e. utterances, delimited by carriage returns) or 2 066 521 tokens.[11] The utterances were produced on both synchronous and asynchronous electronic media. The synchronous or Instant Messaging (IM) media were MSN (i.e. Windows Live Messenger), which does not exist anymore, and Facebook Chat (Messenger). The rest of the corpus consists of posts produced on the – at that time – very popular Belgian social networking site Netlog. For

10 Apart from some additions, the corpus largely corresponds to the one used in De Decker & Vandekerckhove (2017). It was composed by the research group CLiPS of the University of Antwerp. Numerous students of the University of Antwerp contributed to the data collection (we note that these contributions were filtered, so that only utterances from teenagers aged 13-20 remained): they collected data in their own networks and donated these data, together with the information on the demographic profile of the chatters (age, gender, region, and in some cases also educational track). The data that were collected in this way were mainly produced on private synchronous media. The Netlog data (see below) were originally collected for the CLiPS project 'A safer internet: (Semi)automatically recognizing internet paedophilia in multilingual online social networks'. For more information on the project and the data collection, see Peersman, Daelemans & Van Vaerenbergh (2011). All of the data were anonymized: the information on the social profile of the chatter is no longer linked to the name of the chatter, nor can the names be traced back. More information on the entire corpus can be found in De Decker (2014: 23-28).

11 These tokens are the result of splitting the text on whitespace. They were counted automatically. A token can be a word, an emoticon or isolated punctuation marks.

some time, Netlog was considered the European equivalent of Facebook, but in recent years it could no longer compete with Facebook and the site closed in December 2014. Unlike the Facebook data in the present corpus, the Netlog data in our corpus do not only contain chat conversations, but also and predominantly data from asynchronous communication, such as blog posts, profile texts and comments on pictures. In other words, whereas the IM-corpora only cover data from synchronous conversations in real-time, data from mainly asynchronous and to a minor extent synchronous communication are mixed within the Netlog-corpus. Moreover, the Netlog-posts generally have a more public character: the posts and the reactions on the posts reach a wider audience (of peers) than the private IM-conversations. Therefore, we distinguish the private synchronous instant messages from the (largely) asynchronous public messages on the social media site Netlog (see Table 2).

For the age variable, we distinguish a younger group (aged 13-16) and an older group (aged 17-20) of adolescents. Table 2 shows the distribution of the tokens over the age and gender groups and the two media. Although there is an imbalance in the amount of data available for all three social variables (e.g. more male than female material), the smaller subcorpora are always sufficiently large and thus do not exclude valid testing for the three variables.

4.2 Data extraction and processing

The present section provides some explanation on the automatic extraction and quantitative processing of the tokens for the expressive markers.

4.2.1 Typographic and onomatopoeic expressive markers

All occurrences of the typographic and onomatopoeic expressive markers were detected and counted automatically by using Python scripts. The software's performance was evaluated and judged accurate on a test set of 1000 randomly chosen posts (5595 tokens) from the corpus by comparing a human annotator's decisions to the software's output. For the seven

Table 2 Distribution of variables in the corpus

	GIRLS		BOYS		
	YOUNGER	OLDER	YOUNGER	OLDER	Total
Private SYNC.	118 694	176 233	29 146	973 061	1 297 134
Public ASYNC.	463 277	67 257	162 077	76 776	769 387
Total	581 971	243 490	191 223	1 049 837	2 066 521

automatically detected expressive variables, the average precision – i.e. the (relative) number of detected occurrences of a marker that actually are legal occurrences of that marker – is very high: 98%. The average recall – i.e. the (relative) number of occurrences of a marker that were actually detected as occurrences of that marker – is high as well: 95%.

4.2.2 Intensifiers

The intensifiers were automatically extracted using a predefined list covering most of the lemmas (and their variants) present in our corpus. Yet, this method is not exhaustive, as less popular or less obvious intensifying modifiers are not retrieved, nor are intensifiers containing unexpected spelling mistakes or typographical errors. Because of the large size of the corpus, however, the impact of such errors can be assumed to be minimal. With respect to the final selection, we added a frequency cutoff: only lemmas (types) that occurred at least fifteen times in the entire corpus, of which at least five times as an intensifier, were preserved. This cutoff resulted in a list of 23 intensifiers.[12]

We did not select intensifiers that appeared in a negative or interrogative context (cf. Ito & Tagliamonte 2003: 264 and Palacios Martínez & Pertejo 2012: 779). In these contexts, the adjectives or adverbs that follow the intensifiers are often mitigated rather than intensified or emphasized, as illustrated by example (20).

(20) *ma je moet **ni super**veel prentjes ebbn*
'but you **don't** need **that** many images'

After automatic extraction, we manually screened and filtered the software's output, i.e. for each utterance, we checked if the intensifying words were truly used as an intensifier. This finally rendered 14 269 tokens for the selected set of intensifiers. A test set of 700 intensifiers in context was screened by two annotators, who obtained a disagreement of only 1.57% (i.e. the percentage of truly ambiguous utterances containing an intensifier).

12 In alphabetical order: (1) *bere*, (2) *echt*, (3) *echt wel*, (4) *erg*, (5) *fucking*, (6) *gans*, (7) *heel*, (8) *kei*, (9) *kweetniehoe*, (10) *loei*, (11) *mass(as)*, (12) *massiv*, (13) *mega*, (14) *muug*, (15) *over*, (16) *overdreven*, (17) *so*, (18) *super*, (19) *vies*, (20) *vree*, (21) *zeer*, (22) *zo*, (23) *zot.* Adding the frequency cutoff was needed in the original study (Vercammen 2014-2015), where the use of intensifiers (on its own) was correlated with several variables: age, gender, region. Therefore, we needed enough tokens in each of the cells. Moreover, in view of the size of the corpus, it seems sensible not to include intensifiers with an extremely low occurrence.

5 Results and discussion

This section presents the results of the analyses. It starts with the general findings (section 5.1) and is followed by a more detailed discussion of some of the patterns on the level of the individual markers (section 5.2). To verify the statistical significance of our quantitative findings, we combined chi-square tests with a bootstrapping approach (Monte Carlo resampling).[13] With this approach, we can obtain more solid results than when performing one single chi-square test on the entire data set, because we can estimate the (sampling distribution of the) statistics: we first calculated the statistics of interest (chi-square value, p-value, etc.) for each sample and stored them, and finally, we computed the average values (as well as the corresponding standard deviations and confidence intervals). The statistical values reported in the next paragraphs are the mean values for all bootstrap samples.

5.1 General findings

An overview of all expressive markers in the corpus in terms of relative and absolute frequency is shown in Table 3. We note that the use of some markers is more limited than others, depending on their function (e.g. an emoticon can be inserted almost anywhere in an utterance, whereas several grammatical constraints limit intensifier use).

For the analyses, we quantified the degree of expressiveness by dividing the number of expressive markers in the (sub)corpus by the total number of tokens in the (sub)corpus. This approach led to relative expressiveness scores or ratios. The entire data set contained 295 127 expressive markers, which is a ratio of 14.28%. An overview of the ratios per independent variable is shown in Table 4. These percentages should be interpreted with caution. A score of 9.30% does not imply that 9.30% of all tokens in the relevant subcorpus contains an expressive marker. In fact, a smaller percentage of all tokens actually contains an expressive marker, since some tokens contain more than one expressive feature (e.g. combinations of letter flooding and allcaps in one word: *SUUUUUPER*). Yet these scores present

13 Bootstrapping is a statistical technique in which the original data set is resampled by picking *n* bootstrap samples randomly and with replacement, in order to estimate (the sampling distribution of) a statistic (Efron & Tibshirani 1998: 12, Field 2009: 782). By doing so, one is 'treating the data as a population from which smaller samples are taken' (Field 2009: 782). We resampled our corpus by picking 10 000 random samples, each containing 100 000 tokens, chosen with replacement (a same token could thus occur more than once in one sample).

Table 3 Absolute and relative frequencies for each expressive marker in the entire corpus

	Absolute number	Percentage of all markers	Percentage of total number of tokens (*) or question and exclamation marks (**)[14]
Laughter (*)	11 412	3.87 %	0.55 %
Emoticons (*)	150 895	51.13 %	7.30 %
Allcaps (*)	15 029	5.09 %	0.73 %
Kisses (*)	45 129	15.29 %	2.18 %
Flooding letters (*)	40 479	13.72 %	1.96 %
Flooding punctuation marks (**)	17 213	5.83 %	12.18 %
Combinations of question and exclamation marks (**)	701	0.24 %	0.50 %
Intensifiers (*)	14 269	4.83 %	0.69 %
Total	**295 127**	**100 %**	

a reliable indication of the relative representation of expressive markers in the adolescent groups and media. The asynchronous posts contain the highest ratio of expressive markers (28.35%), followed by the younger participants' texts (25.23%) and the girls' texts (21.77%).

Table 4 Overview of expressiveness ratios per subcorpus

Female	Male
21.77%	9.30%
Younger (13-16)	**Older (17-20)**
25.23%	7.74%
Public/Asynchronous posts	**Private/Synchronous posts**
28.35%	5.94%

General tendencies for the social variables are that girls use significantly more expressive markers than boys ($p < .001$, $\chi^2 = 3044.57$, $df = 1$) and that younger teenagers integrate significantly more of them than older ones ($p < .001$, $\chi^2 = 5850.01$, $df = 1$). Furthermore, expressive markers score much

14 As the use of some markers is more limited than others (e.g. because of grammatical constraints), they will naturally occur less frequently. We partially normalized these quantitative differences by counting features related to punctuation in a different way than the other markers. The relative frequency of punctuation flooding and of combinations of question and exclamation marks was obtained by dividing the absolute counts not by the number of tokens in the (sub)corpus (which was done for all other markers), but by the number of occurrences of question and exclamation marks in the (sub)corpus. This increased the otherwise very low relative frequency of these expressive markers.

higher on the public/asynchronous medium than on the private/synchronous media ($p < .001, \chi^2 = 9274.18$, df = 1). In view of the imbalance of several subgroups in relation to particular variables (e.g. older boys are dominant in the synchronous data), we also tested the impact of each independent variable while keeping the other variables constant (for every possible combination of subgroups). Apart from one exception, the observed tendencies were confirmed and turned out to be significant.[15] Moreover, these general tendencies also hold for *each* of the expressive markers: every single expressive marker occurs more frequently in female, younger and public / asynchronous texts than in male, older and private / synchronous texts respectively.

In order to assess the strength of the association between the linguistic and independent variables, we looked at the Cramer's V scores (here identical to Phi scores),[16] which rank from 0 to 1 (Field 2009: 699). The strongest association is found for medium (Cramer's V = 0.31), followed by age (Cramer's V = 0.24) and gender (Cramer's V = 0.17). Apart from that, we took into account the effect size – i.e. a 'measure of the magnitude of observed effect' (Field 2009: 56) – by calculating the odds ratio scores[17] per experiment. These ratios rank from 1 to infinite (or, in the inversed notation, from 0 to 1): 'an odds ratio of 1 would indicate that the *odds* of a particular outcome are equal in both groups' (Field 2009: 790). The odds ratios appear to display the same order as the Cramer's V or Phi scores: medium has the largest effect size (odds ratio = 6.27), followed by age (odds ratio = 4.02) and gender (odds ratio = 2.71). In other words, the odds that a token contains an expressive marker are 6.27 times higher if the token is produced within

15 We ran 12 subtests: 4 per social variable. We will illustrate our approach for gender. We compared the younger **girls'** synchronous data to the younger **boys'** synchronous data (test 1), the younger **girls'** asynchronous data to the younger **boys'** asynchronous data (test 2), the older **girls'** synchronous data to the older **boys'** synchronous data (test 3) and finally the older **girls'** asynchronous data to the older **boys'** asynchronous data (test 4). In these subtests, gender is always the only variable that changes; medium and age remain constant. The only subtest in which the observed tendency was not significant, was the final gender test: older girls used more expressive markers in asynchronous posts than their male peers, but not significantly so.

16 Cramer's V and Phi are 'measures of the strength of association between two categorical variables' (Field 2009: 695). In our experimental setup (with two categorical variables per experiment, each containing two subcategories), the two measures are identical (Field 2009: 698), and are 'calculated by taking the chi-square value and dividing it by the sample size and then taking the square root of this value' (Field 2009: 695).

17 Field (2009: 790) defines odds ratio as 'the ratio of the *odds* of an event occurring in one group compared to another'.

the asynchronous medium than when produced within the synchronous media in our corpus.[18] Medium definitely appears to be the strongest determinant of expressiveness. The correlation with the linguistic variables appears to be very strong and the effect size is much larger than for the other variables.

Some markers produce remarkably high odds ratios. This is the case for letter flooding (deliberate, expressive letter repetition) and the rendition of kisses (e.g. *xxx*), especially with regards to medium. The odds ratios are 51.85 (kisses – medium) and 16.33 (letter flooding – medium): the odds of a token containing a rendition of kisses (letter flooding, resp.) are 51.85 times higher (16.33, resp.) when that token is produced in a public/asynchronous utterance instead of in a private/synchronous post. Markers that were strongly associated with the two other independent variables were letter flooding (CV 0.11, OR 5.53) for gender, and letter flooding (CV 0.14, OR 8.99) and kisses (CV 0.13, OR 6.20) for age. In other words, girls and young adolescents show a strikingly stronger preference for letter flooding than boys and older adolescents, and *x*'s representing kisses are much more frequent in younger adolescents' CMC than in that of the older ones.

5.2 Patterns on the level of the individual markers

5.2.1 General tendencies

The data display some striking constants across all different subgroups with respect to certain patterns or preferences on the level of the individual markers. The present section presents a selection of the dominant tendencies. While the percentages reported in the next paragraphs are the relative counts for the *entire* corpus, the same tendencies were actually found in all six[19] subcorpora.

The most popular expressive markers in all groups are punctuation flooding and emoticons (with relative frequencies of 12.18% and 7.30% resp.). For punctuation flooding, the difference may be (partly) ascribed to the fact that the ratio was not calculated in the same way as for the other markers (see note 14). Since we relate the tokens of flooding of exclamation

18 Note that these numbers differ from the ratios reported in Table 3. Although both numbers express a similar concept, the calculation behind them is different, as sample sizes of both subcorpora are taken into account to calculate odds ratios and not to calculate the straightforward percentages.

19 The six subcorpora are: female texts, male texts, younger texts, older texts, and synchronous and asynchronous posts.

and question marks to all occurrences of these punctuation marks instead of to all tokens in the corpus, the ratio inevitably is higher than for the other markers. However, this does not apply to the emoticons. A possible explanation for their popularity is that these features are very explicit expressive markers: emoticons often literally represent a facial expression. They are very obvious and consequently favored expressive markers.

Another tendency concerns letter flooding: in all subgroups, mainly vowels are repeated (91% of all occurrences of this expressive marker) and hardly ever plosives (2%). Liquids, fricatives and nasals occupy an intermediate position in this respect. This supports the hypothesis that flooding is the (CMC-specific) orthographic representation of an oral phenomenon (Darics 2013: 144), i.e. the lengthening of sounds, which is most natural for vowels and impossible for plosives. Concerning the nature of the words that were emphasized through letter flooding, we found that many of the top lexemes are positively qualifying adjectives (30% of the top 100 types containing letter repetition), mainly variants of the Dutch adjective *mooi* ('beautiful') (22% of the top 100 types). While adolescent language generally has a strong focus on how people and things are valued and experienced (Taylor 2001: 299) with an abundance of evaluative vocabulary (Androutsopoulos 2005: 1497), the nature of the asynchronous data certainly contributes to the top position of the positive qualifiers in the flooding data: a large part of these social media posts are positive reactions to other users' profile pictures, which often involve some degree of pleasing or even flirting (see also section 5.2.2). A similar tendency could be found for the use of intensifiers: the adjective *mooi* represents 18% of the intensified adjectives and adverbs.

With respect to the use of allcaps as an expressive marker, we note the top position of the Dutch first person singular pronoun *ik* ('I') (1.44% of all capitalized lexemes in the entire corpus, and the type that was most often written in capitals by all subgroups). Function words are generally used more frequently than content words (Newman et al. 2008: 216, Pennebaker 2011: 27), but the top position of *ik* might be symptomatic of the intense personal self-expression of the teenagers. Quite often, the pronoun is integrated in an utterance that is consistently written in allcaps.[20]

20 Manual screening of the output revealed that the impact of typographic errors for this phenomenon is negligible: 'IK' was almost always capitalized deliberately (i.e. either integrated in an entirely capitalized sentence (58 out of the 61 cases), or emphasized in a lowercased sentence, in contrast with another pronoun, i.e. 'JIJ' (you) (1 occurrence)). Only in two cases it could not be excluded that the chatter capitalized the entire pronoun

Furthermore, we note a preference for 'simple' variants for the rendition of laughter and kisses and the combination of question and exclamation marks. The three most popular ways of expressing laughter were *haha*, *hahaha* and *hihi*, which are the shortest variants (55% of all onomatopoeic renditions of laughter). The three most popular ways of expressing kisses were *x*, *xx*, *xxx*, also the shortest variants (95% of all renditions of kisses), and finally, the most popular combinations of question and exclamation marks were simply*!?* and*?!* (58% of all occurrences of this feature). These preferences could be interpreted in terms of the speed principle: typing the compact variants is more economical. Apart from that, the less elaborate variants simply seem to be highly conventional, even beyond CMC-contexts: *haha*, for instance, is a very international and common way to express laughter. An interesting (though not academic) tool to estimate the degree of conventionalization and 'internationalization' in informal 'speech' worldwide is the representation and interpretation of these features on Urban Dictionary.[21] The lemma *haha* for instance is identified as a 'short quick way of letting somebody know you are laughing, most likely at them' while its longer variant *hahaha* gets a deviant and more specific interpretation.[22] The same accounts for *x*, *xx* and *xxx*: they are all being identified as kisses on Urban Dictionary, but longer variants as *xxxx* or *xxxxx* are not defined as such. While this type of source has to be handled extremely carefully, it gives a clue with respect to the extent to which particular features are universal and mainstream in informal (online) communication.

Concerning emoticons, finally, we found that the Western variants are the most popular ones among all groups of participants (68% of all emoji in the corpus). They are among the oldest ones (together with the manually composed Asian variants) and are used worldwide, contrary to some of the emoticons that were typical of the Dutch-Flemish social medium of Netlog and of MSN.[23] Western emoticons are (at least for our Western participants)

unintentionally. So the potential 'mechanical' influence of capitalizing digitally (i.e. accidentally capitalizing not only the first letter at the beginning of a sentence, but the next one as well) for this particular token appeared to be very small.

21 urbandictionary.com

22 'To express on aim when something was funny, because just 'haha' isn't that dramatic and can be used as just aknowledging [sic] when someone has said something.'

23 We note that on these two platforms, the traditional facial expressions can be typed manually (e.g. smiling face as :)) and are then converted to a pictogram. For subtler or more elaborate expressions, the platform-specific images need to be selected from the interface. This explains why, for instance, the most popular Netlog variants do not contain smiling faces (as these are not Netlog-specific), but include hugging figures and a blushing face.

also quite easy to interpret and to create: they are simple visual representations of facial expressions. The more recent and highly popular Unicode emoji are not yet present in our corpus, which dates back to 2007-2013. The most popular Western emoticons in the corpus are:

:P or :p (sticking out tongue)
:D (laughing)
:) (smiling)

These variants figure in the top five emoticons for the entire corpus as well as in the top ten for each of the subcorpora, and thus appear to be very popular among all gender and age groups and on all platforms.

5.2.2 Correlations between gender, age and medium

Finally, the in-depth analyses for each of the expressive markers also lay bare correlations between the independent variables. Strikingly, parallel tendencies could be noted for texts written by female participants, by younger teenagers, and on the public/asynchronous medium. What these have in common is, for instance, that they contain many more expressive markers related to love and friendship than those of their male, older adolescent and private/synchronous counterparts. The most popular emoticons (top 3 or 2 for each of these three groups) were all related to love (e.g. the heart-emoticon *<3* or love-related emoticons bound to the specific chatroom and social media site used in this study). Heart-variants specifically figured quite frequently in these posts (9 to 10% of the emoticons used by each of the three groups). Furthermore, many of the top lexemes that were written in allcaps concerned love or friendship (at least 10% of the top 100 lexemes written in allcaps for each group) (e.g. *LOVEYOU*, *BFF*: 'best friend forever'). The same holds for the lexemes that contained letter flooding: 8 to 11% of the top 100 lexemes containing letter flooding for each group were love- or friendship-related (e.g. *iloveyouuu*).

These results manifest a strong discrepancy with boys' and older adolescents' CMC and with practices on the private/synchronous media. In these subcorpora, the top emoticons were not related to love or friendship, nor were heart-variants popular emoticons. On the contrary, they were even the least favored variants (0.40% to 2% of the emoticons used by each of these groups). Only few of the top lexemes containing letter flooding concerned love or friendship (0 to 5% of the top 100 flooded lexemes for each group), and even fewer of the lexemes written in capitals (1 to 3% of the top 100 allcaps lexemes for each group). These three groups' top emoticons contained more representations of negative emotions (e.g. :(, -_- and :/ , respectively a sad,

frustrated and confused face). Many of the lexemes written in allcaps were exclamations (*YEAH*, *WOW*, *BAM*) and 'tougher' words, such as curse words, insults and taboo words (*FUCK*, *ASS*, *GAY*, *GVD* – short for *godverdomme* 'goddamn it'). Finally, most of the boys' and older teenagers' flooded words were positively evaluating adjectives concerning appearance (e.g. *mooooi* for *mooi* 'beautiful'). These lexemes' relatively low frequency in the synchronic posts suggests that the positive evaluations primarily concern (profile) pictures, typical of the asynchronous medium. For boys and older teenagers, and on the synchronous medium specifically, interjections are often flooded (*pfffff*, *ooooh*, *aaaah*) as well as exclamations and greetings (*heeey*).

However, some caution might be needed when interpreting these correlations, as there is an imbalance in our dataset which could (partially) influence our results: young female participants in public asynchronous CMC are overrepresented in our corpus, and so are older male participants in private synchronous CMC (see Table 2). Still, similar correlations between gender and age were reported on before (Argamon, Koppel, Pennebaker, & Schler 2007, Pennebaker 2011, Schwartz et al. 2013). Stylistic correlations concern the use of function words: men and older people use them in similar ways (using more articles and prepositions), as do younger people and women (using more pronouns, conjunctions and auxiliary verbs) – a tendency which seems to hold across cultures, languages and time (Argamon et al. 2007: n.p., Pennebaker 2011: 66, Schwartz et al. 2013: 8-9). On a content-related note, correlations between the same two age and gender groups can be distinguished. Argamon et al. report that men and older people prefer topics like politics, religion and business, whereas women and younger people prefer discussing home, romance and fun (2007: n.p.). These findings correspond to the younger and female teenagers' preferences for expressive markers related to love and friendship.

As for medium, however, no correlations have been reported between the way people write on certain platforms and their gender or age. This could thus be an artefact of the imbalance in our dataset. Another possible explanation lies in the nature of our asynchronous texts. Although many posts on the asynchronous medium are public, the interaction often has a largely personal character. Many comments on this social medium involve pleasing and even flirting (e.g. in positive reactions to other users' pictures). In this respect, our asynchronous medium differs from other public social media, like Twitter, where the communication is less personal and more targeted at informing a wider audience, rather than at bonding or pleasing. The latter focus prevails in our public-asynchronous data: the medium is not only used for intensifying existing bonds, but

also for establishing new network connections, friendship ties and even for dating. By using the love and friendship-related expressive markers (and the other ones), young adolescents acquire social capital. This might explain the higher rate of these markers in the public medium than in the private media, in which people interact with friends from their existing peer group network.

6 Conclusion

This paper discussed linguistic expressiveness in a corpus of Flemish adolescents' computer-mediated messages. We included typographic CMC features (e.g. emoticons), an onomatopoeic variable (the rendition of laughter) and a lexical feature (the use of intensifiers) and looked for possible correlations between these linguistic variables and the author's profile (gender, age) versus the synchronicity and the public versus private character of the CMC medium.

Girls used more expressive markers than boys, and so did the younger adolescents compared to the older ones. The results were extremely consistent in this respect: the same tendencies could be observed for each of the expressive markers. Quite strikingly, however, medium appeared to have the largest impact (more expressive writing in asynchronous and largely public social media posts than in synchronous and mainly private instant messages). Furthermore, the qualitative analyses show that girls and younger teenagers produce more love-related expressive markers than boys and older adolescents. And again, remarkably, these types of correlations were found for medium too (with more love-related markers used in the public/asynchronous than in the private/synchronous posts).

The present research differs from previous research into expressive markers in CMC in that it includes a wider range of expressive markers (both lexical and typographic) combined with three independent variables (age, gender and medium). While gender and, to a minor extent, age have received ample attention in related research, the present findings highlight the importance of the variable medium. They call for refinement of this variable, since apart from (a)synchronicity and the public versus private character of the medium, the character and goal of the interaction seem to be determinant factors too and consequently need to be operationalized in future research. The behavior of the expressive markers is quite revealing in that respect. De Decker & Vandekerckhove (2017) included only one expressive marker (i.e. flooding)[24] and were struck by the

consistent age, gender and medium correlations for this variable. They suggested follow-up research with a wider inclusion of expressive markers in order to enhance insights in the operationalization of (emotional and social) expressiveness in CMC and the way it functions as an identity marker or as an identifying factor for specific subgroups. The present research does not only reveal that expressive markers are particularly age and gender-sensitive but suggests that they serve specific goals: bonding, pleasing, building up social capital, etc. Consequently, their use culminates on media in which these are the main driving forces of the interaction, as was the case with the Netlog medium in the present study, and in social groups that tend to invest heavily in these activities or goals. Young adolescents are intensively engaged in identity construction and extremely sensitive to peer group evaluation. Much of their interaction is driven by a 'need of acceptance and fear of rejection' (Taylor 2001: 298). Expressive markers (that often accompany positively qualifying adjectives – see above) seem to be favored tools in that process of both identity and social network construction, in which angling for approval may be a major determining factor. With respect to gender, female discourse is supposed to have a stronger focus on the establishment of social and emotional connections. The consistent gender findings suggest that girls stress their involvement through the use of expressive markers and especially through the use of features that express friendship and love. In view of young adolescents' dependence on peer group approval, it is hardly surprising that they share the latter preference with women/girls.

Our final conclusion, which concerns suggestions for future research, is that the impact of CMC media deserves more attention. We pointed to the importance of the goals and nature of the interaction, or the general communicative function of the medium. Apart from these aspects, other medium-related properties might be incorporated in the research design as well, like the technology or device that is used (e.g. smartphone or pc), the potential impact of spelling checkers or autocorrection, and limitations in message size. Furthermore, other features or devices for emotional expression could be included in future research, like lexical and syntactical expressions. Finally, it might be challenging to disentangle explicit expressions of positive or negative emoticons from subtler implicit ironic or sarcastic connotations.

24 Without making a distinction between letter flooding and punctuation flooding.

References

Androutsopoulos, Jannis (2005). Research on youth language. In: Ulrich Ammon, Norbert Dittmar, Klaus J. Mattheier & Peter Trudgill (eds.), *Sociolinguistics: An international handbook of the science of language and society* (*vol. 2*). Berlin: Mouton de Gruyter, 1496-1505.

Androutsopoulos, Jannis (2011). Language change and digital media: A review of conceptions and evidence. In: Tore Kristiansen & Nikolas Coupland (eds.), *Standard languages and language standards in a changing Europe*. Oslo: Novus Press, 145-161.

Argamon, Shlomo, Moshe Koppel, Jonathan Fine & Anat Rachel Shimoni (2003). Gender, genre, and writing style in formal written texts. *Text* 23, 321-346.

Argamon, Shlomo, Moshe Koppel, James W. Pennebaker & Jonathan Schler (2007). Mining the blogosphere: Age, gender and the varieties of self-expression. *First monday* 12, n.p.

Argamon, Shlomo, Moshe Koppel, James W. Pennebaker & Jonathan Schler (2009). Automatically profiling the author of an anonymous text. *Communications of the ACM. Inspiring women in computing* 52, 119-123.

Baron, Naomi S. (1984). Computer mediated communication as a force in language change. *Visible language* 18, 118-141.

Baron, Naomi S. (2004). See you online: Gender issues in college student use of instant messaging. *Journal of language and social psychology* 23, 397-423.

Baron, Naomi S. (2008). Are instant messages speech? The world of IM. In: Naomi S. Baron (ed.), *Always on: Language in an online mobile world*. Oxford: Oxford University Press, 45-70.

Bing, Janet M. & Victoria L. Bergvall (1996). The question of questions: Beyond binary thinking. In: Jennifer Coates (ed.) *Language and gender: A reader*. Oxford: Blackwell, 495-510.

Coates, Jennifer (1993). *Women, men and language: A sociolinguistic account of gender differences in language*. London/New York: Longman.

Crystal, David (2001). *Language and the internet*. Cambridge: Cambridge University Press.

Daelemans, Walter (2013). Explanation in computational stylometry. In: *International conference on intelligent text processing and computational linguistics*. Berlin: Springer, 451-462.

De Decker, Benny & Reinhild Vandekerckhove (2017). Global features of online communication in local Flemish: Social and medium-related determinants. *Folia linguistica* 51, 253-281.

De Decker, Benny (2014). *De chattaal van Vlaamse tieners: Een taalgeografische analyse van Vlaamse (sub)standaardiseringsprocessen tegen de achtergrond van de internationale chatcultuur*. University of Antwerp, unpublished doctoral dissertation.

Eckert, Penelope (1997). Age as a sociolinguistic variable. In: Florian Coulmas (ed.), *The handbook of sociolinguistics*. Oxford: Blackwell, 151-167.

Eckert, Penelope (1998). Gender and sociolinguistic variation. In: Jennifer Coates (ed.), *Language and gender: A reader*. Oxford: Blackwell, 64-75.

Eckert, Penelope (2003). Language and adolescent peer groups. *Journal of language and social psychology* 22, 112-118.

Efron, Bradley & Robert J. Tibshirani (1998). *An introduction to the bootstrap*. Boca Raton/London/New York/Washington D.C.: Chapman & Hall/CRC.

Eisikovits, Edina (2006). Girl-talk/boy-talk: Sex differences in adolescent speech. In: Jennifer Coates (ed.), *Language and gender: A reader*. Oxford: Blackwell, 42-54.

Field, Andy (2009). *Discovering statistics using SPSS*. Los Angeles/London/New Delhi/Singapore/Washington DC: SAGE.

Herring, Susan C. (1996). Two variants of an electronic message system. In: Susan C. Herring (ed.), *Computer-mediated communication: Linguistic, social and cross-cultural perspectives*. Amsterdam: John Benjamins, 81-106.

Herring, Susan C. (2001). Computer-mediated discourse. In: Deborah Schiffrin, Deborah Tannen & Heidi E. Hamilton (eds.), *The handbook of discourse analysis*. Malden/Oxford: Blackwell, 612-634.

Herring, Susan C. & Anna Martinson (2004). Assessing gender authenticity in computer-mediated language use: Evidence from an identity game. *Journal of language and social psychology* 23, 424-446.

Herring, Susan C. (2012). Grammar and electronic communication. In: Carol A. Chapelle (ed.), *Encyclopedia of applied linguistics. S.l.*: Wiley.

Holmes, J. (1992). *An introduction to sociolinguistics*. London/New York: Longman.

Huffaker, David A. & Sandra L. Calvert (2005). Gender, identity, and language use in teenage blogs. *Journal of computer-mediated communication* 10, n.p.

Ito, Rika & Sali A. Tagliamonte (2003). Well weird, right dodgy, very strange, really cool: Layering and recycling in English intensifiers. *Language in society* 32, 257-279.

Jespersen, Otto (1922). *Language: Its nature, development and origin*. London: George Allen & Unwin.

Koch, Peter & Wulf Oesterreicher (2001). Gesprochene Sprache und geschriebene Sprache / Langage parlé et langage écrit. In: Günter Holtus, Michael Metzeltin & Christian Schmitt (eds.), *Lexikon der Romanistischen Linguistik* (*vol. 1:2*). Tübingen: Max Niemeyer Verlag, 584-627.

Koch, Peter & Wulf Oesterreicher (2011). *Gesprochene Sprache in der Romania: Französisch, Italienisch, Spanisch*. Berlin/New York: De Gruyter.

Kucukyilmaz, Tayfun, B. Barla Cambazogly, Cevdet Aykanat & Fazli Can (2006). Chat mining for gender prediction. In: *International conference on advances in information systems*. Berlin: Springer, 274-283.

Lorenz, Gunter R. (1999). *Adjective intensification. Learners versus native speakers: A corpus study of argumentative writing*. Amsterdam/Atlanta: Rodopi.

Lorenz, Gunter R. (2002). Really worthwhile or not really significant? A corpus-based approach to the delexicalisation and grammaticalisation of intensifiers in Modern English. In: Diewald Wischer (ed.), *Typological studies in language* (*vol. 49*)*: New reflections on grammaticalization*. Amsterdam/Philadelphia: John Benjamins, 143-161.

Méndez-Naya, Belén (2003). On intensifiers and grammaticalization: The case of swithe. *English studies* 84, 372-391.

Newman, Matthew L., Carla J. Groom, Lori D. Handelman & James W. Pennebaker (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse processes* 45, 211-236.

Paradis, Carita (2000). It's well weird. Degree modifiers of adjectives revisited: The nineties. In: John M. Kirk (eds.), *Corpora galore. Analyses and techniques in describing English*. Amsterdam: Rodopi, 147-160.

Parkins, Róisín (2012). Gender and emotional expressiveness: An analysis of prosodic features in emotional expression. *Griffith working papers in pragmatics and intercultural communication* 5, 46-54.

Partington, Alan (1993). Corpus evidence of language change: The case of intensifiers. In: Mona Baker, Gill Francis & Elena Tognini-Bonelli (eds.), *Text and technology: In honour of John Sinclair*. Amsterdam/Philadelphia: John Benjamins, 177-192.

Peersman, Claudia, Walter Daelemans & Leona Van Vaerenbergh (2011). Predicting age and gender in online social networks. In: *Proceedings of the 3rd international workshop on search and mining user-generated contents*, 37-44.

Peersman, Claudia, Walter Daelemans, Reinhild Vandekerckhove, Bram Vandekerckhove & Leona Van Vaerenbergh (2016). The effects of age, gender and region on non-standard linguistic variation in online social networks. <http://arxiv.org/abs/1601.02431>

Pennebaker, James W. (2011). *The secret life of pronouns: What our words say about us*. New York: Bloomsbury Press.

Peters, Hans (1994). Degree adverbs in early modern English. In: Dieter Kastovsky (ed.), *Studies in early modern English*. Berlin/New York: De Gruyter, 269-288.

Pyles, Thomas & John Algeo (1993). *The origins and development of the English language*. Boston: Heinle.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, Jan Svartvik & David Crystal (1985). *A comprehensive grammar of the English language*. New York: Longman.

Schlobinski, Peter (2005). Mündlichkeit/Schriftlichkeit in den Neuen Medien. In: Ludwig Eichninger & Werner Kallmeyer (eds.), *Standardvariation: Wieviel Variation verträgt die Deutsche Sprache?* Berlin: De Gruyter, 126-142.

Schwartz, H. Andrew, Johannes C. Eichstaedt, Margaret L. Kern, Lukasz Dziurzynski, Stephanie M. Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin E.P. Seligman & Lyle H. Ungar (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS ONE* 8, n.p.

Stenström, Anna-Brita, Gisle Andersen & Ingrid Kristine Hasund (2002). Non-standard grammar and the trendy use of intensifiers. In: Anna-Brita Stenström, Gisle Andersen & Ingrid Kristine Hasund (eds.), *Trends in teenage talk: Corpus compilation, analysis and findings*. Amsterdam: John Benjamins, 131-163.

Stoffel, Cornelis (1901). *Intensives and down-toners*. Heidelberg: Carl Winter.

Tagliamonte, Sali A. & Chris Roberts (2005). So weird; so cool; so innovative: The use of intensifiers in the television series Friends. *American speech* 80, 280-300.

Tagliamonte, Sali A. (2008). So different and pretty cool! Recycling intensifiers in Toronto, Canada. *English language and linguistics* 12, 361- 394.

Tagliamonte, Sali A. & Derek Denis (2008). Linguistic ruin? Lol! Instant messaging and teen language. *American speech* 83, 3-34.

Thurlow, Crispin & Michele Poff (2013). Text messaging. In: Susan Herring, Dieter Stein & Tuija Virtanen (eds.), *Pragmatics of computer-mediated communication*. Berlin/New York: Mouton de Gruyter, 163-190.

Varnhagen, Connie K., G. Peggy McFall, Nicole Pugh, Lisa Routledge, Heather Sumida-MacDonald & Trudy E. Kwong (2010). Lol: New language and spelling in instant messaging. *Reading and writing* 23, 719-733.

Vercammen, Jens (2014-2015). *"Echt massiv sketchy". Versterkers in de chattaal van Vlaamse jongeren: Een corpusonderzoek naar gender-, leeftijds- en regionale verschillen. Unpublished master thesis*, University of Antwerp.

Verheijen, Lieke (2015). Out-of-the-ordinary orthography: The use of textisms in Dutch youngsters' written computer-mediated communication. In: *Proceedings of the second postgraduate and academic researchers in linguistics at York* (*PARLAY 2014*), 127-142.

Verheijen, Lieke (2016). De macht van nieuwe media: Hoe Nederlandse jongeren communiceren in sms'jes, chats en tweets. In: Dorien Van De Mieroop, Lieven Buysse, Roel Coesemans & Paul Gillaerts (eds.), *De macht van de taal: Taalbeheersingsonderzoek in Nederland en Vlaanderen*. Leuven/Den Haag: Acco, 275-293.

Wolf, Alecia (2000). Emotional expression online: Gender differences in emoticon use. *Cyberpsychology & behavior* 3, 827-833.

Over de auteurs

Lisa Hilte, CLiPS research center, University of Antwerp
E-mail: lisa.hilte@uantwerpen.be

Reinhild Vandekerckhove, CLiPS research center, University of Antwerp
E-mail: reinhild.vandekerckhove@uantwerpen.be

Walter Daelemans, CLiPS research center, University of Antwerp
E-mail: walter.daelemans@uantwerpen.be