DE GRUYTER MOUTON Folia Linguistica 2017; 51(1): 253–281

Benny De Decker and Reinhild Vandekerckhove*

Global features of online communication in local Flemish: Social and medium-related determinants

DOI 10.1515/flin-2017-0007
Submitted July 2, 2015; Revision invited March 21 2016; Revision received April 19, 2016; Accepted May 31, 2016

Abstract: The present case study tests the frequency of four clusters of features of computer-mediated communication (CMC) in a written chat corpus of more than two million words, produced by Flemish adolescents living in northern Belgium. The main focus is on the correlation between the occurrence of these features and the social variables gender and age, on the one hand, and the impact of the CMC-medium, on the other. For the latter parameter, a distinction is made between asynchronous and synchronous real time conversation. While gender has received ample attention in recent research on chat and texting practices, the findings qualify the impact of this social variable, and point to age and medium as two significant determinants of the use of typical CMC-features. Finally, the results call for a systematic distinction between expressive and playful vs. more strictly functional or economical features, since precisely that distinction enhances our understanding of the attested gender and age patterns.

Keywords: corpus linguistics, sociolinguistics, CMC, adolescent language, Flemish Dutch, gender

1 Introduction

In the early years of computer-mediated communication (CMC), lay sources such as “press reports, lay wordlists and popular dictionaries” as well as academic linguistic research on online writing practices contributed to the so-called “netspeak myth” (Androutsopoulos 2006: 420; see also Dürscheid 2004). Especially informal online peer group discourse was claimed or supposed to

***Corresponding author: Reinhild Vandekerckhove,** Department of Linguistics, University of Antwerp, Grote Kauwenberg 18, 2000 Antwerpen, Belgium,
E-mail: reinhild.vandekerckhove@uantwerpen.be
Benny De Decker, Department of Linguistics, University of Antwerp, Grote Kauwenberg 18, 2000 Antwerpen, Belgium, E-mail: benny.dedecker@uantwerpen.be

be marked by an abundance of typical and unique netspeak features (Crystal 2001: 18), which gave the new genre(s) an aura of inaccessibility for non-practitioners. In the past decade, academic research has gradually shifted from merely listing and describing typical CMC-features towards contextualizing them (Androutsopoulos 2006: 421). While this has resulted in a growing awareness of social correlates and medium-related determinants (see Sections 2.1 and 2.2), it has also put into perspective the relative weight of these typical informal CMC-features (see below), as several studies warned against overestimating their frequency (e. g., Baron 2004; Tagliamonte and Denis 2008; and parallel findings by Thurlow and Poff 2013 for texting).

Furthermore, the alleged homogeneity of the genre has been questioned and local appropriation of global CMC-features has become an issue. Many pioneering studies (e. g., Crystal 2001; Baron 2004) were based on corpora produced by English speaking subjects, and much of the present-day research still is. Moreover, many CMC-features that are known and used worldwide are actually English-based. Yet, there is no such thing as one global online chat culture: global characteristics or strategies of informal online communication need some localization or local reshaping in order be suitable or acceptable for local peer groups (Androutsopoulos 2006; Androutsopoulos 2010), all the more so in non-English speaking environments.

The present case study fits in this CMC-research approach. It investigates the presence of some typical informal CMC-features in electronic data produced by teenagers whose native language is a Flemish variety of Dutch. Apart from analyzing their actual frequency and the ways in which global online communication strategies are shaped in local non-English discourse, the present paper investigates potential age, gender and medium-correlates. Section 2 presents the CMC-research context based on a review of related research, which leads to the research questions. The corpus and some methodological issues are dealt with in Section 3. Section 4 discusses the results and Section 5 presents the subsequent conclusions.

2 Informal online communication: Social and contextual factors

2.1 Genre and medium-related characteristics

The typical characteristics of informal CMC that are generally dealt with in the relevant literature relate to spelling or to the visual rendering of (parts of) words, utterances and paralinguistic and non-verbal cues (see, e. g., Varnhagen et al. 2010; Herring 2012). The use of these features can largely be explained in terms of three main triggers. First of all, two principles appear to serve as implicit "rules of conduct" for informal CMC: *write like you speak* and *write as fast (or as concisely) as possible* (see Vandekerckhove and Nobels 2010: 658–659; De Decker et al. 2016). To some extent, informal written CMC appears to be "conceptually oral" (Schlobinski 2005: 132): due to the (nearly) synchronous interaction in many chat media and the informal character of most chat conversations, online conversations share some features of spoken interaction. However, this does not imply that informal CMC offers a perfect reflection of spoken language. The chatters in the Flemish corpus actually integrate a lot of phonological, morphological and syntactic dialect or substandard speech features in their chat discourse. Yet an accurate and faithful speech transcription is generally not aimed at. Rather they create a kind of regional style by the selective use of particular substandard markers, the result being a kind of stylized substandard for CMC (see Vandekerckhove and Nobels 2010). Moreover, other maxims interfere with the speech principle (see De Decker et al. 2016): since speaking generally proceeds much faster than typing, chatters appear to use a diverse range of strategies in order to maximize the speed of typing and guarantee smooth turn-taking, e. g., all kinds of abbreviations, acronyms and other economical spelling choices, in that way adhering to the second principle referred to above. And finally, chatters look for strategies to compensate for the lack of some of the paralinguistic and non-verbal cues that are inherent to spoken interaction. This "semiotics of compensation" (Androutsopoulos 2011) results, for example, in the use of all kinds of emoticons (see, e. g., Dresner and Herring 2012), but other features like the (excessive) use of capital letters and the repetition or *flooding* of letters or certain punctuation marks have similar pragmatic functions.

More so than the speech-characteristics, it is the features that result from the speed- or the economy-principle and the compensatory pragmatic features that constitute the typical informal CMC-characteristics. Therefore, for the present case study, we selected features which belong to either of these two categories, i. e., flooding, leetspeak and several abbreviatory strategies. First of all, the repetition or flooding of letters (e. g., *suuuuppppeeeerrrrr* 'super') is a typical compensatory strategy for expressing strong social or emotional involvement. Second, leetspeak forms, i. e., word forms in which letters are combined with symbols or figures, can be functional abbreviations and/or compensatory creative transformations: while they often serve brevity (e. g., *w8* 'wait'), in some cases they are merely creative adaptations (e. g., *g33k* 'geek'; cf. Crystal 2001: 88–89). Finally, acronyms, single-word abbreviations and grapheme cluster reductions at first sight unambiguously meet the speed-principle.[1] More explanation on the linguistic features that function as the dependent variables in the present case study can be found in Section 4.

Whether or not specific CMC-media stimulate or inhibit the use of these features remains fairly unclear. The impact of the medium can be interpreted in terms of synchronicity or the absence of it. Synchronous conversations create more time pressure, which might lead to reduced linguistic complexity (Herring 2001: 617). However, we might wonder whether this still holds today: the young digital natives are extremely dexterous in dealing with (touchscreen) keyboards, autocorrect software, (adjusted internal) dictionaries (in their mobile phones), etc. Yet, according to Gheuens (2010: 17–18), adolescents pay more attention to spelling (norms) on forums with asynchronous communication than in chat rooms with real time synchronous interaction. Asynchronous written communication leaves more room for checking and (more careful) editing of a text before it is posted. From that perspective, more "standard writing" is to be expected in asynchronous CMC. However, asynchronicity also implies more time for experimenting, which might favor the production of non-standard spelling forms and all kinds of creative inventions. Apart from synchronicity, there is also the private–public dimension: people may feel more comfortable when having a private chat conversation with friends from a close peer group network than when posting messages for a wider audience, even if that audience consists of peers. But 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. In other words, the impact of the medium on the use of protypical informal CMC-features seems somewhat ambiguous and hard to predict.

2.2 Social determinants

2.2.1 Gender

Gender is the only social variable that has been operationalized with some frequency in CMC-research in recent years. In some cases, the approach is more holistic and the focus transcends language use. Kapidzic and Herring (2011), for instance, analyze teenagers' visual and textual self-presentation in chat rooms and conclude that young girls show more emotional involvement

1 The results for cluster reduction will nuance this hypothesis for this particular feature (see Section 4.4 and the discussion in Section 5).

and empathy whereas adolescent boys appear more assertive, more dominant and generally also more distant. Similar observations are made by Schwartz et al. (2013: 9) who present a vocabulary study of a huge corpus of Facebook messages. They reveal that female participants use more emotion words and “mention more psychological and social processes”, whereas men use more object references and swear words. Finally, Ling et al. (2014: 425) refer to several studies on texting which prove that “[f]emales are more likely to use it for social purposes, and males are more likely to have an instrumental use”. All of these findings seem to support older sociolinguistic and pragmatic research on gender patterns in male and female discourse. In her 1990 best-seller, Deborah Tannen stressed, for example, that female interaction strongly focuses on establishing (emotional) connections and negotiating relationships whereas men’s talk is, amongst others, about negotiating status and giving information. Similarly, Holmes (1995) concluded that women are more concerned about the affective function of talk, while men generally focus on its referential function.

Huffaker and Calvert (2005), however, state that linguistic gender differences are less clearly profiled in CMC than in “real life”, because girls tend to take on less traditional gender roles in online interaction. They support Rodino’s warning (1997) against oversimplification when establishing parallels between “speech” and written CMC, because online communication can be assumed to create more room for flexible gender roles. Quite striking is that Huffaker and Calvert (2005) do not even attest gender differences for the use of emoticons, although many other CMC-studies do actually reveal a higher production of smileys in girls’ chat than in boys’ chat, both in synchronous and asynchronous media (e. g., Witmer and Katzman 1997; Wolf 2000; Lee 2003; Baron 2004; Tossell et al. 2012; Schwartz et al. 2013). A similar point is made by Parkins (2012), who combines several types of expressive markers in her analysis of social networking data from Australian men and women and concludes that women are “still” the emotionally expressive gender (Parkins 2012: 53). In other words, contrary to Huffaker and Calvert (2005), most CMC-studies seem to confirm classic gender patterns in this respect. Moreover, perception research by Ling et al. (2014: 429–430) shows that the American adolescents whom they interviewed on their texting practices are very much aware of these patterns: excessive use of certain expressive markers (e. g., emoticons, repetition of punctuation) is definitely not considered a “guy thing”. Boys “know” they should avoid using these features (except when flirting with girls), because they are incongruent with how they perceive (their) masculinity. In other words these gender related patterns and strategies seem to play an important role in adolescents’ identity construction.

Furthermore, Varnhagen et al. (2010: 729) find more “new language”-features in girls’ Instant Messaging than in boys’, but this is somewhat incongruent with Baron (2004: 415), who attests significantly more abbreviations in boys’ chat. Boys generally seem to be more concise than girls: their conversation units tend to be shorter and so are the individual turns of boys (see also Baron 2004). The same tendencies are attested in texting: more than their female peers, Finnish boys focus on speed (Kasesniemi 2003), while Norwegian girls tend to be more innovative in their lexical choices than their male colleagues (Ling 2005).

On the basis of these findings, we hypothesize that boys will use more acronyms and abbreviations, whereas girls might produce more letter flooding and playful leetspeak forms (see Section 2.3).

2.2.2 Age

Age differences have hardly been dealt with in CMC-research. Most of the research on chat language focuses on adolescents’ chat practices. A well-known assumption and often attested tendency with respect to adolescent speech is the so-called “adolescent peak”-principle (Holmes 1992), which implies that there is a gradual increase in non-conformist (linguistic) behavior during early adolescence, culminating around the mid-teens. Beyond that point, non-standard behavior (or speech production) gradually decreases. The translation of the adolescent peak-principle to digital writing is quite challenging: if the use of typical informal CMC-features falls under the label “non-conformist” or “non-standard”, we expect a higher number of expressive and experimental chat features in the youngest age group. On the other hand, the use of some deviant spelling forms and acronyms may have become highly conventionalized, to the extent that they have become the default in chat contexts, rather than a deviation of the/a “norm”. Moreover, since some CMC-features (e. g., acronyms) are extremely functional, there seems to be no reason why older teenagers should stop using them.

Yet, Tagliamonte and Denis (2008: 13) find that the use of a highly popular acronym such as *lol* gradually decreases as their Canadian adolescent informants (aged 15–20) grow older: “Perhaps as a result of habituation to the IM environment, it seems that adolescents quickly outgrow at least some of the IM forms”. However, Varnhagen et al. (2010: 729), whose corpus contains data of American teenagers aged 12–17, found no age grading effect on the production of “new language” forms (i. e., all kinds of Instant Messaging features). But Varnhagen’s age limit is 17 whereas Tagliamonte and Denis (2008) include older teenagers up to the age of 20. And so does the present case study. We divide the chatters in two age categories: 13–16 and 17–20 years old (see Section 4.1).

2.3 Research questions

The literature discussed in the previous sections points to substantial opportunities for future research: medium has hardly been operationalized systematically in CMC-research, and the age and gender findings are quite ambiguous. Some studies suggest there might be an adolescent peak effect, with young adolescents being more eager to integrate typical CMC-features in their online discourse than older ones (e. g., Tagliamonte and Denis 2008), while others (e. g., Varnhagen et al. 2010) attest no age grading patterns. With respect to gender, the same duality prevails: some studies minimize gender patterns in CMC (e. g., Huffaker and Calvert 2005), while others actually do reveal significant gender correlations, for instance a stronger tendency towards brevity in boys' CMC and more markers of emotional involvement in girls' CMC (for both, see, e. g., Baron 2004). Finally, the effect of the synchronous and private versus the asynchronous and public character of online interaction, which is related to the chat medium, is potentially twofold: synchronous chat may be a stronger trigger for economical linguistic choices, while asynchronous chat may leave more room for experimenting with playful or creative chat features. Furthermore, public self-presentation may be a trigger for creative language use, but at the same time people may feel more at ease when experimenting in private conversations.

In other words, much more research is needed if we want to determine which of the above tendencies hold or prevail (if any). The present case study hopes to contribute to this endeavor, by operationalizing the impact of gender, age and medium for an extensive chat corpus (see Section 3.1). It starts from the following research questions:

(1) Is there a significant impact of gender on the chat practice? Do girls use more expressive or emphatic CMC-features? Is there a tendency for boys to prefer shortening strategies?
(2) Do younger adolescents produce more CMC-features than older adolescents?
(3) What is the effect of the CMC-medium (private synchronous CMC versus public asynchronous CMC) on the production of typical features of informal online interaction?

3 Corpus and methodology

3.1 Corpus

The corpus contains chat conversations and online posts produced between 2007 and 2013 by Flemish adolescents aged 13–20. The entire corpus consists of 2 million words. For each of the informants, we have the following profile information: age, sex and place of residence. The teenagers chat and post information about all kinds of topics: going out, sports, gaming, school, friendship and love, birthdays, etc. All of the data have been anonymized: the names were replaced by codes that were linked to the sociodemographic profile of the informants. For the sake of privacy, the link between the names of the participants and the codes has been removed, so that the identity of the informants is no longer traceable. Table 1 presents the composition of the entire corpus:

Table 1: The chat corpus and its subcorpora.

# words	Girls 13–16 years	Girls 17–20 years	Boys 13–16 years	Boys 17–20 years	Total
IM *(synchronous + private)*	122,569	185,082	29,342	979,405	1,316,398
Netlog *(asynchronous + public)*	430,864	64,443	152,884	73,519	721,710
Total	553,433	249,525	182,226	1,052,924	2,038,108

Though some cells contain far more data than others, each of the social variables is well-represented: the total number of words is 802,958 for the girls and 1,235,150 for the boys. The young adolescents produced 735,659 words, the older ones 1,302,449. For medium, the Instant Messaging data outnumber the Netlog data, but with 721,710 words Netlog is firmly represented too.

The Instant Messaging data result from synchronous and private online conversations between two (or sometimes more) people. In other words, they contain one-to-one and to a lesser extent one-to-many communication between peers from close adolescent networks. The IM-media were MSN, which does not exist anymore, and Facebook Chat. The data were collected by the authors of the present paper and numerous students of Dutch linguistics at the University of Antwerp. The rest of the corpus consists of posts produced on the – at that time – very popular Belgian social networking site Netlog.[2] For some time Netlog was considered the European equivalent of

2 These data were provided by the CLiPS research group (Computational Linguistics & Psycholinguistics, University of Antwerp) and are embedded in the Daphne-project.

Facebook. In 2008, it won an award in San Francisco in the category "Mainstream and Large Social Networks". 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 not only contain chat conversations, but also data from asynchronous communication, such as blog posts, profile texts and comments on pictures. In other words, whereas the IM-corpora only comprise 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. In other words, the Netlog-communication is essentially one-to-many communication. From now on, we refer to the IM-corpus as the private SYNC-corpus and to the Netlog-corpus as the public ASYNC-corpus.

This difference between a more public and a more private medium may also have consequences for the strength of the network ties between the participants: within the IM-media MSN and Facebook Chat (SYNC-corpus), users know their addressees, but this need not be the case for the Netlog data (ASYNC-corpus) since they include interaction with or self-presentation to strangers as well. In other words, generally speaking, the participants of the public ASYNC-corpus can be assumed to have weaker network ties than the chatters of the private SYNC-corpus. "Being one of them" may be a major driving force within synchronous media, but the adolescents with a Netlog profile also acquire social capital from presenting themselves as "cool" and "trendy", even if (part of) the addressees are unknown, since these Netlog addressees are potential new contacts. So once again, the effect of these parameters on the use of CMC-features is hard to predict: they might function as markers of in-group solidarity or they might be attractive tools in identity construction towards outsiders.

A final remark with respect to the corpus concerns the time span covered by the data: seven years is a long period in CMC-history. Symptomatic in this respect is the abovementioned fact that two of the online media represented in our data no longer exist: Facebook has replaced both MSN for Instant Messaging and Netlog for social networking. Moreover, a new competitor for Instant Messaging by mobile phone is steadily gaining ground: WhatsApp. While part of our SYNC-data are Facebook data, the 2007–2013 corpus does not contain any WhatsApp-conversations. Furthermore, CMC generally proceeded via laptop or pc in 2007, while today most adolescents' online peer group communication proceeds via mobile phone.[3] This transition was stimulated by the development of an (immediately popular) Messenger app for mobile phone developed by Facebook in 2011. An additional aspect, partly linked to the changing media, is that all kinds of CMC-features may abruptly gain popularity or gradually get outdated. Or some features may be quite popular in particular networks at a particular moment and much less or only much later in others.[4] Consequently, we had to integrate all of these parameters into our study or we had to "neutralize" them. We opted for the latter by carefully selecting features that generally mark Flemish online communication throughout these seven years. In this way, we hope to avoid distortions of the conclusions caused by, for example, short-lived hypes or features that are media-sensitive.

3.2 Methodology: Data extraction and processing

Four types of CMC-features (i. e., flooding, leetspeak, abbreviations and cluster reductions) were selected in order to test to what extent they were actually used by the Flemish adolescents and which of the parameters discussed above seemed to affect their use. In other words, we tested the impact of the independent variables age, gender and medium on the occurrence of the CMC-features. The initial searches for all of the features were performed automatically, by means of Python-scripts, but in the second phase, manual selection was needed to filter out the noise, i. e., items which are mistakenly identified as tokens of the selected feature. Python is a programming language with many functionalities, including the development of scripts for automatic data selection (see www.python.org). One of Python's assets with respect to the chat data is that its scripts can include a wide range of spelling variants For example, a search for the tokens of the acronym *omg* ('oh my god') has to yield all of the (potential) variants of this token (e. g., *zomg, omgg, omggg, OMG*, etc.). The Python-scripts can be designed in such a way that all of these variants can actually be traced.

For each of the features, the main analyses are based on the ratio of words or visual units that contain the relevant feature to the total number of words or visual units in the corpus (for more details, see Sections 4.1–4.4). With respect to

3 We are currently collecting new data (in Flanders) on a large scale. Time and again, we are struck by the fact that some adolescents even do not appear to know how to access Facebook via computer because they systematically use their mobile phones devices for doing so.
4 See the example of the acronym *yolo* in Section 4.3.

the quantitative and statistical data processing, the unequal contribution of individual chatters presents a major challenge. In spontaneous language data, some participants inevitably produce (much) more "speech" than others, and this might skew the quantitative analyses. In order to neutralize these individual discrepancies as much as possible, a *linear mixed effects model with random stratum for the chatters* (*lmer*)[5] was used; this statistical model allows us to test to what extent the effects of the independent variables (age, gender and medium) can be attested systematically in the chatters' discourse. The main benefit of the model is that the analyses can be performed on the individual data points of all of the informants and not on their averages. Apart from the effects of the fixed factors in the research design, the lmer analyses also quantify the random effect of the individual chatters, who represent an arbitrary selection of the total population of "chatters" and may cause error variance in the data that is not related to the (supposedly) systematic effects of the design factors (i. e., the independent variables of age, gender and medium).

4 Results

Table 2 presents an overview of the results of the lmer-analyses. These results serve as a starting point for the discussions below. The reference group (the intercept) contains the IM-conversations of the young teenage boys. Negative estimate and z-scores indicate that the reference group is dominant for a feature; the reverse holds for positive estimate and z-scores. For instance, the estimate scores and z-scores for the feature flooding and the independent variable gender are positive, which indicates that it is not the reference group (i. e., the boys) but the girls that have higher frequency scores for this feature. Estimate scores and z-scores are negative for the independent variable age for the same feature (flooding), because in this case the reference group (the young teenagers) scores higher than the other group (the older adolescents). The coefficient of determination r^2 remains fairly low in each of the analyses, since the data are in many respects highly variable, as stated above, and thus cause the data points to fall further from the regression line. Nevertheless, *p*-values lower than 0.05 (rendered in bold in the table) still signal significant correlations between the independent variables and the binary response variable (CMC variant x = 0; non-CMC variant x (any other word/visual unit in the corpus) = 1; e. g., all words/visual units with flooding vs.

5 See Bates et al. (2012); all analyses were conducted in *R* (R Development Core Team 2011).

Table 2: Results of the lmer-analyses.

Flooding		**Variance**	**Std. dev.**		
(r^2 = 0.10)	Random effect	1.2929	1.1371		
		Estimate	**St. error**	***z***	***p***
	Intercept	−5.208	0.058	−89.060	<0.001
	Medium	1.397	0.055	25.530	**<0.001**
	Gender	0.614	0.025	24.680	**<0.001**
	Age	−0.603	0.030	−20.270	**<0.001**
Leetspeak		**Variance**	**Std. dev.**		
(r^2 = 0.05)	Random effect	42.019	6.4822		
		Estimate	**St. error**	***z***	***p***
	Intercept	−13.352	0.882	−15.144	<0.001
	Medium	1.718	0.838	2.051	**0.040**
	Gender	0.159	0.462	0.345	0.730
	Age	−0.390	0.573	−0.680	0.496
English		**Variance**	**Std. dev.**		
acronyms	Random effect	25.659	5.0655		
(r^2 = 0.02)					
		Estimate	**St. error**	***z***	***p***
	Intercept	−10.633	0.479	−22.220	<0.001
	Medium	0.143	0.421	0.339	0.735
	Gender	0.127	0.296	0.431	0.667
	Age	−0.594	0.385	−1.543	0.123
Dutch		**Variance**	**Std. dev.**		
acronyms	Random effect	9.4275	3.0704		
(r^2 = 0.06)					
		Estimate	**St. error**	***Z***	***p***
	Intercept	−7.846	0.218	−35.960	<0.001
	Medium	−1.283	0.179	−7.150	**<0.001**
	Gender	0.088	0.154	0.570	0.570
	Age	0.362	0.164	2.210	**0.027**
cluster		**Variance**	**Std. dev.**		
reductions	Random effect	3.1102	1.7636		
(r^2 = 0.03)					
		Estimate	**St. error**	***z***	***p***
	Intercept	−7.298	0.127	−57.610	<0.001
	Medium	0.909	0.117	7.760	**<0.001**
	Gender	−0.040	0.064	−0.630	0.527
	Age	−0.472	0.081	−5.840	**<0.001**

all words/visual units without flooding). The quantitative data for each of the features (i. e., exact number of tokens and their relative representation) and the discussion of the results can be found in the Sections 4.1–4.4.

4.1 Flooding

Flooding is a cover term for the repetition of letters, punctuation marks or symbols, which can be considered as a kind of “pragmatic lengthening” (Varnhagen et al. 2010: 725), symbolizing some form of “emphasis”. Flooding can be used for a variety of reasons, for instance to stress part of an utterance, to express enthusiasm, disbelief and other emotions. Since letters may be doubled by mistake (because of fast typing), we only selected words or utterances containing sequences of at least three identical graphemes or symbols: e. g., *superrr* ‘super’, *meeeega* ‘mega’, *!!!!* or:-))). Yet, even then, all kinds of unmistakably accidental grapheme clusters had to be filtered out manually.

Table 3 presents for each cell the ratio of words or visual units with flooding to the total number of words/visual units in that part of the corpus. Words in which several letters are flooded are counted as one flooding token. So *suuuuppppeeeerrrrr* ‘super’ represents one flooding token in the table below, not four.

Table 3: Flooding ratio: gender and age differences.

	Young adolescents	Older adolescents	Total
Girls	5.16 %	1.39 %	3.99 %
	28,537/553,433	3,466/249,525	32,003/802,958
Boys	3.08 %	0.32 %	0.73 %
	5,609/182,226	3,355/1,052,924	8,964/1,235,150
Total	4.64 %	0.52 %	2.01 %
	34,146/735,659	6,821/1,302,449	40,967/2,038,108

Some more examples from the corpus

(1) *prachtig mooiii haar* (prachtig mooi haar) ‘wonderfully beautiful hair’

(2) *ooohhh gent tofff* (oh, Gent, tof) ‘oh, [the city of] Ghent, cool’

Younger adolescents produce significantly more flooding than older teenagers. Gender appears to have a significant impact too, since girls show a higher preference for flooding than boys. In this respect, the present data confirm previous CMC research results which suggested that girls are more expressive or emphatic in their chat practices, thus generally manifesting greater emotional involvement (see Section 2.2.1).

The data also reveal a significant correlation with medium (see Table 4): participants in the ASYNC-corpus appear to apply significantly more flooding than chatters in the SYNC-corpus. In other words, the dependent variable flooding manifests a significant correlation with all of the independent variables and, quite strikingly, it is the only variable to do so (see Table 2).

Table 4: Flooding ratio: medium-related differences.

Private SYNC-corpus	Public ASYNC-corpus	Total
0.34 %	5.06 %	2.01 %
4,461/1,316,398	36,506/721,710	40,967/2,038,108

When looking at the flooding data, some more tendencies can be discerned. First of all, exclamatory interjections (e. g., *ooooh, heeey, pffff*) are popular flooding items, which confirms the pragmatic-expressive function of this feature. On the grapheme-level, graphemes which render vowels tend to be flooded more frequently than graphemes which render consonants: 77.85 % of the floodings affect the Dutch vowel-graphemes *a, e, i, o*, and *u*. With respect to the consonants, graphemes rendering plosives are seldom affected by flooding: plosives represent only 1,44 % of the flooding tokens (e. g., *hottt* 'hot'). This establishes a link with spoken language, since the articulation of plosives cannot be lengthened either. Finally, smileys can actually be subjected to flooding but examples are rare (e. g., *:-ooo*).

4.2 Leetspeak

The term *leetspeak* refers to the use of all kinds of characters and symbols in order to replace letters within a word unit.[6] In this case study, we only focus on the use of figures instead of letters, e. g., Dutch/English *w8* (wacht, 8 = acht) 'wait'. Some of these letter/number-homophones are assumed to be highly conventionalized and widely used in chatting and texting.

6 See: http://en.wikipedia.org/wiki/Leet.

Our initial automatic search selected all words that contained at least one letter combined by at least one figure. Furthermore, we included the use 2 for 'to(o)' and *4* for 'for', thus incorporating some highly conventionalized and more or less idiomatic English constructions that appear to be part of Flemish teenage CMC-jargon: e. g., *love u 2* 'love you too' and *4ever* 'forever'. In a second phase, manual selection was again needed to eliminate the large amount of noise caused by, for instance, hyperlinks, mathematical formulas, and accidental combinations of letters and figures in which the figures actually do represent a number and not a combination of letters (e. g., *2de* (tweede, twee = 2 'second')).

Table 5 presents for each subgroup the ratio of words or visual units with leetspeak to the total number of words/visual units in that part of the corpus. While leetspeak may be a well-known informal CMC-feature, judging from its frequency, there seems to be no reason to qualify it as a typical marker of informal online communication, at least not in the Flemish data. Hardly one in two thousand words appears to be affected by leetspeak transformations. Our findings present a striking parallel with the observations of Bieswanger (2006) who describes a remarkable discrepancy between English and German text messages with respect to the use of letter/number-homophones: in the German data, he found no letter/number-homophones at all, while the English corpus contained quite a lot of tokens of this feature. Clearly, the Flemish data takes an in-between position in this respect. Still, it would appear that this type of leetspeak is much more firmly rooted in CMC where English is the base language. In his conclusion, Bieswanger (2006) submits that "the German language does not provide the same potential in this category as English". This generalization does not seem to hold for Dutch, as is seen in examples (3) to (5) below. Yet in many cases the Flemish chatters copy English letter/number-homophones. In other words, they do not seem to maximally exploit the potentials in Dutch for this feature.

Table 5: Leetspeak ratio: gender and age differences.

	Young adolescents	Older adolescents	Total
Girls	0.13 %	0.04 %	0.10 %
	700/553,433	98/249,525	798/802,958
Boys	0.13 %	0.01 %	0.03 %
	234/182,226	142/1,052,924	376/1,235,150
Total	0.13 %	0.02 %	0.06 %
	934/735,659	240/1,302,449	1,174/2,038,108

Although the data in Table 5 suggest that the production of leetspeak decreases as the adolescents grow older, the lmer-analyses (Table 2) show no significant age correlation. The same holds for gender: girls do not produce significantly more leetspeak than boys. This could be related to a potential double function of these features: shortening (presumably preferred by boys) or expressivity and playfulness (preferred by girls, see Section 2.2.1). For medium, however, the pattern displayed in Table 6 is actually corroborated by the lmer-analyes: the data reveal a stronger preference for leetspeak in the public ASYNC-corpus than in the private SYNC-corpus.

Table 6: Leetspeak ratio: medium-related differences.

Private SYNC-corpus	Public ASYNC-corpus	Total
0.013 %	0.14 %	0.06 %
175/1,316,398	999/721,710	1,174/2,038,108

When analyzing the leetspeak data, two types of leetspeak can be discerned. The first type is the "functional" shortening type: one figure replaces several letters on the basis of phonological similarity, which implies that the leetspeak transformation saves a number of keystrokes. English appears to be an important source of inspiration for the Flemish adolescents (e. g., *4ever* 'forever'), but the process is applied to the Dutch lexicon too:

(3) *leerk8te* (leerkrachten, 8 = acht, the <r> is missing) 'teachers'

(4) *suc6* (succes, 6 = zes) 'success'

(5) *9r* (neger, 9 = negen) 'black person'

The second type is creative rather than functional. Chatters play with the visual rendering of the words and replace one letter by one figure on the basis of typographic resemblances. The origin of leetspeak can be related to this type of visual transformation, since these alternative spelling forms enabled chatters to escape the automatic censorship of search machines: the English form *pr0n* 'porn' (with the figure 0 instead of the letter O, and a deliberate metathesis of the letters <r> and <o>) has become a typical example of this type of leetspeak. The following examples from our corpus illustrate the creative aspect of these visual leetspeak forms:

(6) *g00gl3-sk1llz* 'google-skills'

(7) *s4nd3rr* 'Sander' (name)

(8) *gy zijt l3kkerder* (gij zijt lekkerder) 'you are more tasty'

The creative leetspeak transformations are much more frequent in the ASYNC-corpus than in the SYNC-corpus: in the public asynchronous data they represent 310 of the 999 leetspeak tokens (or 31.03 %), compared to 29 of the 175 tokens (or 16.57 %) in the private chat data. An explanatory factor is most probably that these experimental leetspeak forms demand some cognitive effort and time and that asynchronous CMC is more convenient in this respect than the fast turn-taking in private chat.

The overall low scores for leetspeak may be accounted for in the same terms: even the time-saving effect of the functional leetspeak forms may not outweigh the loss of time because of the extra cognitive and even physical effort (since figures are somewhat remote from letters on the keyboard or demand the use of the shift-key). We cannot but conclude that leetspeak is not as omnipresent in informal CMC as some descriptions of the genre seem to suggest (see, e. g., Axtman 2002; Roos 2006; or the examples in Den Ouden and Van Wijk 2007).

4.3 Acronyms and abbreviations

These shortenings may be the most typical informal CMC-features; their presence can be directly related to the speed principle (see Section 2.1): using acronyms and word abbreviations may save a lot of keystrokes. But the question is whether they are actually so prominent and whether speed is really the determining factor. For this dependent variable, we distinguish between English versus Dutch acronyms and abbreviations.

For the English chat acronyms and abbreviations, we selected 10 well-known and widely used items (see De Decker and Vandekerckhove 2012): *btw* 'by the way', *aight* 'all right', *wtf* 'what the fuck', *brb* 'be right back', *imo* 'in my opinion', *omg* 'oh my god', *bk* 'back', *ofc* 'of course', *atm* 'at the moment' and *np* 'no problem'. The selection process was not entirely unproblematic. First of all, it is hard to find out which acronyms or abbreviations are highly representative for the present-day chat practices and which are not (anymore), since the use of these forms appears to be extremely trend-sensitive. The following example may illustrate this: when the compilation of the database started, in 2007, the acronym *yolo* 'you only live once' was not used at all in Flanders, since it was launched only at the end of 2011 by the American rapper Drake. But by the end of 2012, it was attributed the status of youth slang word of the year both in Belgium and in Germany.[7] Yet, *yolo* is completely absent in our corpus data, even in the recent data of the years 2012–2013. Apparently, it was not popular at all within the social networks our informants belonged to. Furthermore, we decided to exclude one English acronym, *lol* 'laughing out loud', which is extremely popular both in Flemish (De Decker and Vandekerckhove 2012: 331) and in English chat (Baron 2004: 411; Tagliamonte and Denis 2008: 12). The reason is that its acronym-status is ambiguous for Dutch chatters since the word *lol* 'fun' actually exists in Dutch (e. g., *lol hebben* 'having fun'). In other words, the English acronym coincides with a Dutch lexeme. Yet English influence has certainly contributed to its popularity: the lexeme *lol* originates in Netherlandic Dutch, but has become extremely popular in Flemish Dutch too due to its use in social media.

Since the full variants of the English acronyms are rarely used,[8] it seemed to make no sense to systematically quantify the ratio of these acronyms versus their full equivalents. The low frequency of the full variants is hardly surprising, because many chatters learn the English acronyms via social media and therefore may simply not have the full equivalents in their mental lexicon. The acronyms have become highly conventionalized features of chatting practice. Therefore, we calculated the ratio of English acronyms and abbreviations to the whole of the tokens/visual units in the entire corpus rather than to their full equivalents. This decision has the advantage that the data processing for this variable is in line with the other variables (flooding and leetspeak), which simply cannot be presented in a dichotomous way. The results for the variables age and gender are presented in Table 7.

Since we selected only ten English acronyms and abbreviations and excluded the most popular one (*lol*), we cannot draw conclusions from the low relative scores as such. Quite strikingly, no consistent age- and gender-related patterns can be discerned: young girls use more acronyms and abbreviations than older girls, but the reverse pattern holds for the boys. This feature is the only one for which the older boys get the highest score. Yet overall, age and gender differences are not significant. Moreover, medium does not appear to be

7 http://woordvanhetjaar.vandale.be/ and http://www.jugendwort.de/#worum_geht_s.

8 For the 10 selected items, the corpus contains 3693 tokens for the acronyms versus only 123 tokens for their full equivalents (i. e., 3.22 % for the latter category). *Back, of course* and *no problem* are sometimes rendered as full forms, but this is seldom the case for more "classic" acronyms, such as *wtf* and *btw* and even never for *brb, imo* and *atm*.

Table 7: The ratio for English acronyms and abbreviations: gender and age differences.

	Young adolescents	Older adolescents	Total
Girls	0.16 %	0.05 %	0.13 %
	899/553,433	120/249,525	1,019/802,958
Boys	0.13 %	0.23 %	0.22 %
	246/182,226	2 428/1,052,924	2,674/1,235,150
Total	0.16 %	0.20 %	0.18 %
	1,145/735,659	2,548/1,302,449	3,693/2,038,108

Table 8: English acronyms and abbreviations: medium-related differences.

Private SYNC-corpus	Public ASYNC-corpus	Total
0.20 %	0.15 %	0.18 %
2,585/1,316,398	1,108/721,710	3,693/2,038,108

a determining factor either (Table 8). In sum, for none of the independent variables significant correlations can be found.

Nevertheless, the fact that the asynchronous material contains less real time conversations than the synchronous chat data has consequences for the type of acronyms used by the chatters. Acronyms which structure a conversation, such as *brb* 'be right back' or *bk* 'back', are less relevant and consequently less frequent in the asynchronous corpus than in the synchronous chat data, see (9). In the public ASYNC-corpus, the dominant acronyms appear to be *omg* 'oh my god' and *wtf* 'what the fuck', see (10).

(9) *kga ff ne wafel hale, brb* (ik ga even een wafel halen, be right back)
'I'm going to get a waffle, brb (be right back, or: back in a moment).'

(10) *idd wtf ben jij zielig eerst kopieren en dan wat bewerken en zeggen dat jou skin is omg* (inderdaad, what the fuck ben jij zielig, eerst kopiëren en dan wat bewerken en zeggen dat het jouw skin is, oh my god)
'Indeed, what the fuck, you are pathetic, first you copy and edit and then you claim this is your skin, oh my god.'

For the Dutch abbreviations and acronyms, we also quantified their relative frequency with respect to the entire corpus (Table 9), but we added a table with the relative frequency of the reduced variants compared to their full equivalents

(Table 10), because – contrary to what is the case for the English acronyms – we assume that these full forms are part of the productive lexicon of all Flemish chatters. Abbreviations and acronyms that have been part of the Dutch written standard language for a long time (e. g., *m.a.w.* (*met andere woorden*) 'in other words'), have been excluded from the analysis. We selected four typical Dutch CMC-abbreviations and two Dutch CMC-acronyms. For the latter category it was hard to find representative tokens, since the Dutch chatters heavily rely on English when using acronyms. The selected abbreviations are: *idd* (inderdaad) 'indeed', *mss* (misschien) 'maybe', *wss*/*wrs*/*wsl*/*wsch*/*wrsch* (waarschijnlijk) 'probably' and *ff* (even, "effe" in Flemish slang) 'for a short moment'. Furthermore, we searched the corpus for the following acronyms: *iig* (in ieder geval) 'in any case' and *(k)hvj* ((ik) hou van je) '(I) love you'. Table 9 presents the share of these 6 types relative to the entire corpus:

Table 9: Dutch acronyms and abbreviations: gender and age differences.

	Young adolescents	Older adolescents	Total
Girls	0.17 %	0.36 %	0.23 %
	965/553,433	888/249,525	1,853/802,958
Boys	0.17 %	0.53 %	0.48 %
	315/182,226	5,584/1,052,924	5,899/1,235,150
Total	0.17 %	0.50 %	0.38 %
	1,280/735,659	6,472/1,302,449	7,752/2,038,108

Contrary to the data for the English abbreviations and acronyms, the Dutch data does reveal a significant correlation with the age of the informants. Strikingly, the older adolescents use significantly more abbreviated forms than the younger ones. Gender clearly is not a determining factor.

More distinct gender patterns appear in Table 10, where the use of the compressed forms is opposed to the occurrence of the full variants. Surprisingly though, unlike what Table 10 seems to suggest, additional lmer analyses (not included in Table 2) for the data in this table reveal that neither gender ($z = -1.404$; $p = 0.160$) nor age ($z = 0.757$; $p = 0.449$) have a significant impact on the production of the full versus the abbreviated variants.

Generally speaking, the frequency score for the selected features is high: in 86 % of the cases, the Flemish chatters prefer the abbreviated forms. The observation that the production of these features does not decrease as the chatters grow older is analogous to the findings for the English acronyms, and so is the absence of significant gender differences. This might imply that these

Table 10: The ratio for Dutch acronyms and abbreviations to their full equivalents: gender and age differences.

	Young adolescents	Older adolescents	Total
Girls	70.34 %	77.15 %	73.44 %
	965/407	888/263	1,853/670
Boys	74.64 %	92.76 %	91.57 %
	315/107	5,584/436	5,899/543
Total	71.35 %	90.25 %	86.47 %
	1,280/514	6,472/699	7,752/1,213

acronyms and abbreviations are stable and functional, rather than purely innovative or trendy aspects of informal CMC, which prove useful both for younger and older boys and girls.

With respect to medium, Table 11 shows a higher preference for the Dutch abbreviations in synchronous chat than in the asynchronous posts, which is supported by the lmer-analyses (Table 2). Contrary to the medium-related data for the English acronyms, this pattern seems to confirm the importance of economical and speed-enhancing strategies in synchronous interaction.

Table 11: Dutch acronyms and abbreviations: medium-related differences.

Private SYNC-corpus	Public ASYNC-corpus	Total
0.51 %	0.15 %	0.38 %
6,672/1,316,398	1,080/721,710	7,752/2,038,108

An example from the corpus with two of these abbreviations in one and the same sentence is (11):

(11) *dan ist* <u>*mss*</u> <u>*idd*</u> *verstandig om zijn geld op iet anders te zette* (dan is het misschien inderdaad verstandig om zijn geld op iets anders te zetten)
'Then it is maybe indeed wise to invest his money in something else.' (Then it might indeed be wise…)

Finally, it may not be a coincidence that, for the four selected Dutch abbreviations, the longest word, i. e., *waarschijnlijk* 'probably', is shortened most frequently (96.14 %), whereas the shortest one, i. e., *even/effe* 'just for a moment', has the lowest "reduction" rate (78.95 %). The scores for the two Dutch acronyms are symptomatic of the inherent variability of online discourse (Thurlow and Poff 2013: 10): *hvj* (hou van je) 'love you' gets a frequency score of 50.56 % whereas *iig* (in ieder geval) 'in any case' is only used in 20.45 % of the cases instead of its full equivalent.

4.4 Cluster reduction

The final dependent variable (or CMC-feature) concerns the reduction of clusters of two graphemes to one grapheme. So the result is also a shortened form, but unlike the abbreviations dealt with in the previous section, the affected lexemes are not completely compressed. The chatters use an alternative (and non-standard) spelling for a grapheme cluster, but the rest of the word remains unaffected. In Dutch CMC, this type of spelling adaptation generally affects only the following three clusters: *ch* which is reduced to *g* ([x], e. g., *egt* instead of *echt* 'real'); *ij* > *y* ([ɛi], e. g., *wy* instead of *wij* 'we'); *ks* > *x* ([ks], e. g., *nix* instead of *niks* 'nothing').

We selected a number of words that have a high lexeme frequency in the corpus and are subjected to this type of conversion. The latter may seem a strange a priori selection criterion, but the process appears to be lexeme-bound. It therefore makes no sense to select lexemes that are never affected, since we want to detect gender-, age- or medium-related patterns. In theory, the cluster reduction could be applied to any word that contains the grapheme clusters, but in practice only a limited number of (mostly high-frequency) words are affected. For example, the words *links* 'left' (or '(hyper)links' to websites) and *reeks* 'series' could be transformed into *linx* and *reex*, but these forms are completely absent from the corpus. For the *ks* > *x* conversion, in fact only *niks* > *nix* 'nothing' appeared to be a useful lexeme. For the other clusters, there were more candidates. Eventually, the following lexemes were selected:

– *ch* > *g: schoon* 'beautiful', *echt* 'real', *toch* 'yet', *school* 'school', *schat(je)* 'darling';
– *ij* > *y: wij* 'we', *zijn* 'to be', *altijd* 'always', *blij* 'glad', *tijd* 'time';
– *ks* > *x: niks* 'nothing'.

Both Table 12, which presents the proportion of cluster reduction relative to the entire corpus, and Table 13, which compares frequency of reduced to non-reduced variants, reveal distinct age differences. These differences appear to be significant: for the data presented in Table 12, we can refer to the general survey in Table 2. The additional lmer-analysis for Table 13 confirms that young teenagers apply the grapheme cluster conversions significantly more often than their older peers ($z = -11.878$; $p < 0.001$). For gender, the analyses are contradictory: Table 2 reveals

Table 12: Cluster reductions: gender and age differences.

	Young adolescents	Older adolescents	Total
Girls	0.49 %	0.15 %	0.38 %
	2,706/553,433	385/249,525	3,091/802,958
Boys	0.50 %	0.08 %	0.14 %
	915/182,226	833/1,052,924	1,748/1,235,150
Total	0.49 %	0.09 %	0.24 %
	3,621/735,659	1,218/1,302,449	4,839/2,038,108

Table 13: cluster reduction versus no cluster reduction: gender and age differences.

	Young adolescents	Older adolescents	Total
Girls	29.53 %	8.62 %	22.68 %
	2,706/6,458	385/4,079	3,091/10,537
Boys	33.32 %	4.47 %	8.18 %
	915/1,831	833/17,803	1,748/19,634
Total	30.40 %	5.27 %	13.82 %
	3,621/8,289	1,218/21,882	4,839/30,171

no significant gender patterns when the reductions are related to the entire corpus, but additional lmer-analyses for Table 13 suggest that boys tend to produce more of these reduced variants than girls ($z = -4.456$; $p < 0.001$). This finding is somewhat surprising, since we expected a stronger orientation towards creative and innovative CMC-forms in girls' chat than in boys' chat. However, the overall impact of the factor age appears to be much bigger and more consistent than that of gender.

The lmer-test for medium-related differences (see Table 2) confirms the patterns revealed by the figures in Table 14. Just like flooding and leetspeak, cluster reduction occurs more frequently in the asynchronous data. This suggests that brevity cannot be the main determinant for the production of these cluster reductions: speed is more of an issue in synchronous chat conversations and yet they contain less of these grapheme-reductions. So playfulness and expressivity may be more decisive (cf. leetspeak and flooding). Example (12) contains two of the cluster reductions in question:

(12) *tis nix sgatje ('t is niks, schatje)* 'it is nothing, darling' (nothing to worry about, darling)

Table 14: Cluster reduction: medium-related differences.

Private SYNC-corpus	Public ASYNC-corpus	Total
0.09 %	0.50 %	0.24 %
1,197/1,316,398	3,642/721,710	4,839/2,038,108

5 Discussion

While gender does not appear to be a major determinant of variation patterns in the informal online interaction of the Flemish teenagers for the features included in the present study, the results for this social variable are most interesting in two respects. First of all, the only CMC-feature which shows a (highly) significant gender correlation is the single feature which unambiguously belongs to the category of the compensatory expressive CMC-features (see Section 2.1), i. e., the flooding of letters. Flooding implies strong emphasis, it makes exclamations even more emphatic (e. g., *ooohhh* versus *oh*) and puts extra stress on, for instance, adjectives that express all kinds of evaluations and emotions (e. g., *mooiii* versus *mooi* 'beautiful' /*suuuuppppeeeerrrrr* versus *super*). The observation that girls outperform boys in flooding corroborates previous findings on girls expressing emotional involvement in CMC much more than boys. The latter are supposed to be more concerned with information transfer (see Section 2.2.1; Kasesniemi 2003; Baron 2008; Kapidzic and Herring 2011). At the same time, the overall findings warn against overestimating the impact of gender. The hypothesis that boys would show a greater preference for shortening strategies is not confirmed (see Sections 2.2.1 and 2.3). Huffaker and Calvert (2005) conclude that gender differences are less marked in CMC, because girls tend to take on less traditional gender roles in online interaction. This claim does not seem to hold with respect to emotional expressiveness, yet the absence of gender correlations for the other four CMC-features corroborates their findings to some extent (see also below).

The age patterns also point to a dichotomy in the class of CMC-features. The English and Dutch acronyms and abbreviations are economical spelling choices, the use of which does not decrease as the chatters grow older. The overall presence of Dutch abbreviations is even higher in the data of the older adolescents than in those of their younger peers, while the production of English acronyms shows no significant age differences. Apparently, these features are so useful and functional, especially for high frequency words (compare *mss* to *misschien* 'maybe' and *wss* to *waarschijnlijk* 'probably') and high frequency expressions (compare *omg* to *oh my god*), that they have acquired a fairly stable position in (Flemish) CMC. The age patterns for flooding and grapheme cluster reduction, however, confirm the adolescent peak principle. Younger teenagers appear to be much more attracted to these non-standard writing patterns than older ones. The cluster reductions save only one keystroke (e. g., *egt* instead of *echt* 'really'), which most probably makes the alternative spelling forms playful rather than strictly functional, and therefore less appealing to older adolescents. The same holds for flooding. The category of leetspeak, however, comprises forms with a purely playful and innovative character (e. g., *g00gl3* 'google') as well as more functional word compression (e. g., *w8* for Dutch *wacht* and English *wait*). Apart from the extremely low frequency of this feature, which is interesting in itself (since popular media tend to overrate its use), the heterogeneity of the leetspeak category might explain why no age patterns emerge.

The data for medium corroborate this dichotomy once again, though this time leetspeak is clearly positioned on the non-economical side: flooding, leetspeak and cluster reductions score significantly higher in the (largely) asynchronous and public corpus than in the synchronous chat data. Asynchronous communication is not driven by fast turn-taking and leaves more time for experimenting than synchronous interaction. From that perspective, it is hardly surprising that flooding and leetspeak are more frequent in the asynchronous data. Flooding lengthens words and utterances and therefore runs counter to time and space saving strategies. The same applies to leetspeak, since the integration of numbers and other symbols in words demands extra cognitive and physical effort. Although cluster reduction is much less complex than leetspeak and moreover saves one keystroke, it does not behave like the much more efficient and economical acronyms and abbreviations. Cluster reductions are favored in the asynchronous corpus much more than in the synchronous chat conversations, which seems to corroborate their playful rather than strictly functional status once again. Furthermore, the expressive versus playful dimension of respectively flooding versus leetspeak and cluster reduction may be much more of an issue in a public medium like (asynchronous) Netlog than in the private (synchronous) chat conversations, since the participants in the public medium not only communicate with people from their personal networks but also with a wider audience that might include potential new contacts. By using leetspeak and cluster reductions, adolescents demonstrate CMC-linguistic dexterity (see Deumert and Lexander 2013: 535) and by applying flooding, they signal emotional and social involvement. Both of these create social capital or prestige. In other words, the use of these CMC-features might raise their personal attractiveness to outsiders.

Moreover, the latter factor might also explain why young adolescents, who intensively engage in identity construction, score higher for some of these features than older adolescents nearing their twenties.

For the Dutch and English acronyms and abbreviations, different patterns emerge: medium is a significant factor for the Dutch acronyms and abbreviations. This time, however, the SYNC-scores are higher than the ASYNC-scores, which is in line with expectations, given the time constraints of synchronous communication. The same pattern might be expected for the English acronyms and abbreviations, but once again they produce no significant correlations. In other words, these English acronyms and abbreviations are not sensitive to medium, age or gender variation. They seem to be the most stable markers of the genre: contrary to leetspeak, cluster reductions and flooding, they are not features to show off with, but useful and efficient CMC-tools.

Finally, the analyses of the Flemish chat data reveal that expressive and playful CMC-features produce other age, gender and medium correlates than the often highly conventionalized and functional economical spelling choices. In other words, there appears to be an interaction between age, gender and medium patterns, on the one hand, and the status or function of these features on the other. The systematic distinction between these two categories of CMC-features (i. e., expressive/playful vs. functional CMC-features) offers a promising line for future research, especially if we want to enhance our understanding of gender patterns in online communication. Girls do not lag behind in the use of chat features that make CMC more efficient. Generally, they master and apply many CMC-features to the same extent and in much the same way as boys, except when it comes to the use of the expressive marker that we called *flooding*. With respect to (emotional) expressiveness, it would appear that old gender patterns are reproduced within new media (see also the related research discussed in Section 2.2.1). A prominent gender scholar like Deborah Tannen (1990), for instance, stressed the importance of the connective dimension in women's discourse (see also Holmes 1995). Precisely this dimension might be at issue here: the girls seem to outperform the boys in establishing social and emotional connections by stressing their involvement through the use of expressive markers. These findings call for further refinement of the operationalization of emotional expressiveness in CMC and a broader selection of expressive markers (cf. Parkins 2012). At the same time, this endeavor could further our insights into, first of all, age grading patterns during adolescence as they surface in CMC and, secondly, the way these markers function as tools for identity construction (see also Ling et al. 2014) in (early) adolescence, since creative and expressive markers appear to be particularly age-sensitive too.

Acknowledgment: We sincerely thank the anonymous reviewers and the editor for their pertinent comments and helpful remarks on the previous version of this article.

References

Androutsopoulos, Jannis. 2006. Introduction: Sociolinguistics and computer-mediated communication. *Journal of Sociolinguistics* 10(4). 419–438.

Androutsopoulos, Jannis. 2010. Localizing the global on the participatory web. In Nikolas Coupland (ed.), *The handbook of language and globalization*, 203–231. Oxford: Blackwell.

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*, 145–161. Oslo: Novus.

Axtman, Kris. 2002. 'r u online?': The evolving lexicon of wired teens. *The Christian Science Monitor*. http://www.csmonitor.com/2002/1212/p01s01-ussc.html (accessed 25 January 2017).

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. *Always on: Language in an online and mobile world*. Oxford: Oxford University Press.

Bates, Douglas, Martin Mächler & Ben Bolker. 2012. *lme4: Linear mixed-effects models using S4 classes (R package version 0.999375-42)*. http://cran.r-project.org/web/packages/lme4/index.html (accessed 7 April 2016).

Bieswanger, Martin. 2006. 2 abbrevi8 or not 2 abbrevi8: A contrastive analysis of different shortening strategies in English and German text messages. *Texas Linguistics Forum* 50 – *SALSA XIV*. http://salsa-tx.org/proceedings/2006/Bieswanger.pdf (accessed 7 April 2016).

Crystal, David. 2001. *Language and the Internet*. Cambridge: Cambridge University Press.

De Decker, Benny & Reinhild Vandekerckhove. 2012. English in Flemish adolescents' computer-mediated discourse. A corpus-based study. *English world-wide: A journal of varieties of English* 33(3). 321–351.

De Decker, Benny, Reinhild Vandekerckhove & Dominiek Sandra. 2016. When two basic principles clash: About the validity of written chat language as a research tool for spoken language variation. Flemish chatspeak as a test case. *Journal of Language Contact* 9. 101–129.

Den Ouden, Harry & Carel Van Wijk. 2007. 'Om vet gaaf op te kicken': Over jongerentaal en het gebruik ervan in productadvertenties ['Om vet gaaf op te kicken': About youth language and its use in product advertising]. *Tijdschrift voor Communicatiewetenschap* 35. 232–248.

Deumert, Ana & Kirstin V. Lexander. 2013. Texting Africa: Writing as performance. *Journal of Sociolinguistics* 17(4). 522–546.

Dresner, Eli & Susan C. Herring. 2012. Emoticons and illocutionary force. In Dana Riesenfeld & Giovanni Scarafile (eds.), *Philosophical dialogue: Writing in honor of Marcelo Dascal*, 59–70. London: College Publication.

Dürscheid, Christa. 2004. Netzsprache – ein neuer Mythos. In Michael Beißwenger, Ludger Hoffmann & Angelika Storrer (eds.), *Internetbasierte Kommunikation*. Special Issue, *Osnabrücker Beiträge zur Sprachtheorie* 68. 141–157.

Gheuens, Koen. 2010. *De spelling voorbij: Een onderzoek naar het hedendaagse gebruik van spelling bij Vlaamse en Nederlandse jongeren in formele en informele contexten* [Beyond the spelling: A study of current spelling practices by Flemish en Dutch teens in formal and informal contexts]. Leuven: University of Leuven MA thesis.

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

Herring, Susan C. 2012. Grammar and electronic communication. In Carol Chapelle (ed.), *Encyclopedia of applied linguistics*. Hoboken, NJ: Wiley-Blackwell. http://ella.slis.indiana.edu/~herring/e-grammar.pdf (accessed 7 April 2016).

Holmes, Janet. 1992. *An Introduction to sociolinguistics*. London: Longman.

Holmes, Janet. 1995. *Women, men and politeness*. London: Longman.

Huffaker, David A. & Sandra L. Calvert. 2005. Gender, identity, and language use in teenage blogs. *Journal of Computer-Mediated Communication* 10(2). http://jcmc.indiana.edu/vol10/issue2/huffaker.html (accessed 7 April 2016).

Kapidzic, Sana & Susan C. Herring. 2011. Gender, communication, and self-presentation in teen chatrooms revisited: Have patterns changed?. *Journal of Computer-Mediated Communication* 17(1). 39–59.

Kasesniemi, Eija-Liisa. 2003. *Mobile messages: Young people and a new communication culture*. Tampere: Tampere University Press.

Lee, Christine. 2003. *How does instant messaging affect interaction between the genders?* Stanford: The Mercury Project for Instant Messaging Studies. http://www.stanford.edu/class/pwr3-25/group2/pdfs/IM_Genders.pdf (accessed 7 April 2016).

Ling, Richard. 2005. The socio-linguistics of SMS: An analysis of SMS use by a random sample of Norwegians. In Richard Ling & Per E. Pedersen (eds.), *Mobile communication: Renegotiation of the social sphere*, 335–349. London: Springer.

Ling, Richard, Namoi S. Baron, Amanda Lenhart & Scott W. Campbell. 2014. 'Girls text really weird': Gender, texting and identity among teens. *Journal of Children and Media* 8(4). 423–439.

Parkins, Róisín. 2012. Gender and emotional expressiveness: An analysis of prosodic features in emotional expression. *Griffith Working Papers in Pragmatics and Intercultural Communication* 5(1). 46–54.

R Development Core Team. 2011. *R: A language and environment for statistical computing*. http://www.R-project.org

Rodino, Michelle. 1997. Breaking out of binaries: Reconceptualizing gender and its relationship to language in computer-mediated communication. *Journal of Computer-Mediated Communication* 3(3). http://jcmc.indiana.edu/vol3/issue3/rodino.html (accessed 7 April 2016).

Roos, Lilian. 2006. D Alm8ige sms-tl: De invloed van sms en msn op het Nederlands [The almighty sms-language: The influence of sms and msn on Dutch]. *Onze Taal* 7(8). 192–194.

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

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(9). http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0073791 (accessed 7 April 2016).

Tagliamonte, Sali A. & Derek Denis. 2008. Linguistic ruin? Lol! Instant Messaging and teen language. *American Speech* 83(1). 3–34.
Tannen, Deborah. 1990. *You just don't understand. Women and men in conversation*. New York: Ballantine Books.
Thurlow, Crispin & Michele Poff. 2013. Text messaging. In Susan. C. Herring, Dieter Stein & Tuija Virtanen (eds.), *Pragmatics of Computer-Mediated Communication*, 163–190. Berlin & New York: Mouton De Gruyter.
Tossell, Chad C., Philip Kortum, Clayton Shepard, Laura H. Barg-Walkow, Ahmad Rahmati & Lin Zhong. 2012. A longitudinal study of emoticon use in text messaging from smartphones. *Computers in Human Behavior* 28(2). 659–663.
Vandekerckhove, Reinhild & Judith Nobels. 2010. Code Eclecticism: Linguistic Variation and Code Alternation in the Chat Language of Flemish Teenagers. *Journal of Sociolinguistics* 14. 645–665.
Varnhagen, Connie K., G. P. McFall, Nicole Pugh, Lisa Routledge, Heather Sumida-MacDonald & Trudy E. Kwong. 2010. Lol: New language and spelling in instant messaging. *Reading and Writing: An Interdisciplinary Journal* 23(6). 719–733.
Witmer, Diane F. & Sandra L. Katzman. 1997. On-line smiles: Does gender make a difference in the use of graphic accents? *Journal of Computer-Mediated Communication* 2(4). http://jcmc.indiana.edu/vol2/issue4/witmer1.html (accessed 7 April 2016).
Wolf, Alecia. 2000. Emotional expression online: Gender differences in emoticon use. *CyberPsychology & Behavior* 3(5). 827–833.