OUT-OF-THE-ORDINARY ORTHOGRAPHY: THE USE OF TEXTISMS IN DUTCH YOUNGSTERS' WRITTEN COMPUTER-MEDIATED COMMUNICATION

LIEKE VERHEIJEN
Radboud University

Abstract

Recent decades have seen an explosive growth in computer-mediated communication (CMC). Since the language used in CMC can deviate from standard language conventions, concerns have been expressed that CMC may degrade youths' reading, writing, or spelling skills. However, before studying the possible impact of CMC on traditional literacy, the ways in which 'CMC language' differs from the standard language need to be established. This article discusses the first findings of an ongoing large-scale corpus study examining the register of written CMC of Dutch youngsters between the ages of twelve and twenty-three, revealing how their CMC language differs from Standard Dutch in various dimensions of writing. The focus here is on a salient orthographic feature, namely the use of textisms (unconventional spellings). A range of CMC modes was investigated, including instant messages, text messages, and microblogs. It is shown that the extent to which CMC users deviate orthographically from the standard language and the degree to which they use particular textism types depends both on CMC mode and on individual user characteristics such as age.

Keywords: computer-mediated communication (CMC), writing, register, orthography, spelling, corpus study

1. Introduction

Computer-mediated communication (CMC) has been defined as "the practice of using networked computers and alphabetic text to transmit messages between people or groups of people across space and time" (Jacobs 2008:470). Simply put, it is communication that takes place via modern communication tools, such as personal computers, mobile (smart) phones, and tablets. Since CMC is used more and more nowadays by even young children, groups such as parents, teachers, and the popular media are afraid that it may have a detrimental impact on youths' literacy skills. Such concerns have been documented for English-speaking countries (Thurlow 2006) and the Netherlands (Postma 2011). A *Daily Mail* article by Humphrys (2007), titled "I h8 txt msgs: How texting is wrecking our language," expresses this widespread fear about the language used in CMC. Humphrys describes texters as:

> vandals who are doing to our language what Genghis Khan did to his neighbours eight hundred years ago. They are destroying it: pillaging our punctuation; savaging our sentences; raping our vocabulary. And they must be stopped. (¶15)

Some linguists, on the other hand, argue for the positive effects of CMC. They mention the creative, innovative use of written language; increased motivation to read and write; more exposure to written text and extra opportunities to engage with writing; and greater phonological and metalinguistic awareness, that is, sensitivity to the underlying (sound) structure of language (since certain types of abbreviations used in CMC reflect an understanding of grapheme-phoneme patterns). In sum, opinions about CMC and its impact on literacy vary greatly. As Swartzlander (2010:9) put it, CMC has caused "a tsunami of anxiety, excitement, paranoia, enthusiasm, fear and fascination."

 ISSN 1758-0315

Before we can study if and if so, how, CMC affects youths' reading or writing skills, we need to establish the ways in which their 'CMC language' differs from the standard language norms. The present article addresses this question by discussing the first results of a large-scale corpus study into the register of CMC writing produced by Dutch youngsters.

2. *Background*

2,1 The landscape of computer-mediated communication

A prerequisite for studying the register of CMC is to establish what belongs to the domain of computer-mediated communication. CMC is an umbrella term covering an array of new/social media. Figure 1 presents a classification of CMC modes (based on Merchant 2007; Herring 2012; Van Dijck 2013), along with popular examples. A well-known CMC mode is text messaging, also known as texting and SMS (Short Message Service). There is also online chat, of which there are two kinds: chatting in web-based chat rooms and instant messaging (IM). IM can occur through four kinds of technologies: using a mobile app (e.g. *WhatsApp*, *Telegram*, *Google Hangouts*) or a desktop application (formerly *MSN Messenger*, now *Skype text chat*), through a social networking site (*Facebook chat*), and within an online gaming network or virtual world (*World of Warcraft*, *Second Life*). Another CMC mode, widely used for both personal and professional purposes, is emailing. CMC also occurs through social networking sites (*Facebook*, *Google+*, *MySpace*), microblogging platforms (*Twitter*, *Tumblr*), and visual media sharing platforms (*YouTube*, *Instagram*, *Pinterest*). Furthermore, CMC includes blogging and online forums or discussion boards.[1] All this shows that CMC covers a broad range of computer-mediated genres.

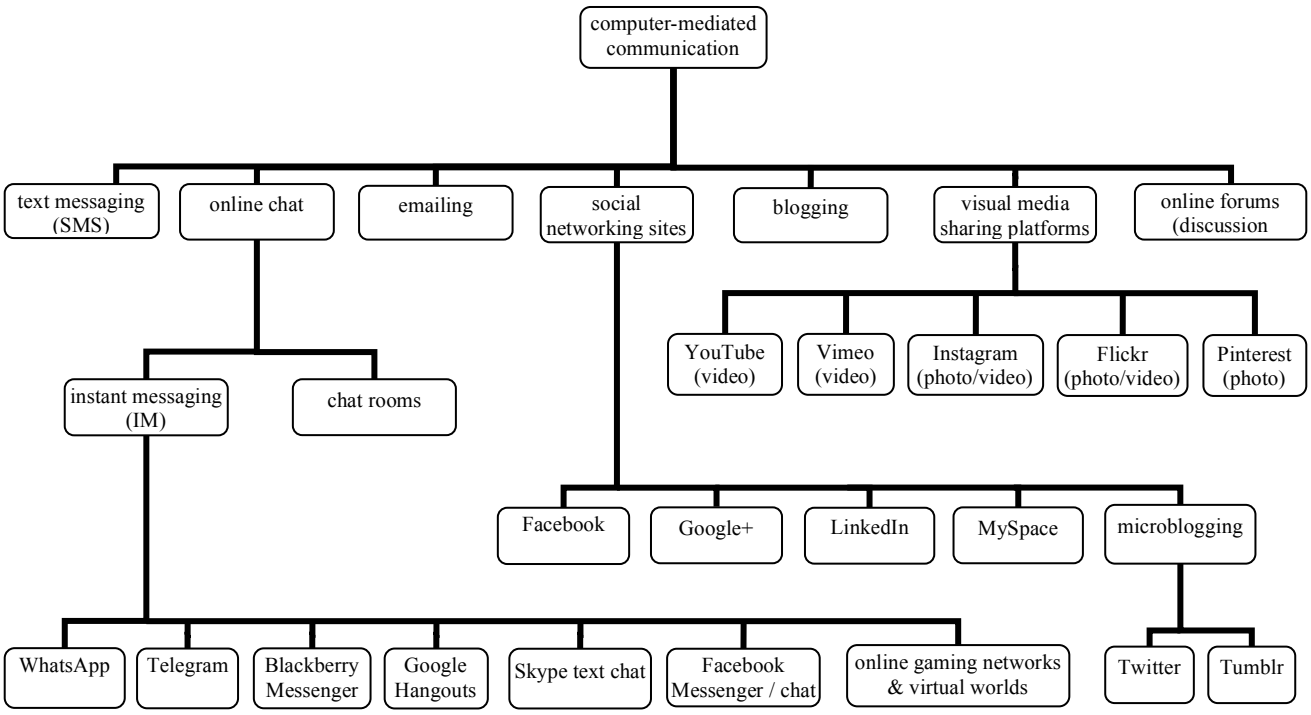

Figure 1: classification of computer-mediated communication

[1] One could argue that web pages and wikis are part of CMC as well, but I have excluded them from myclassification, because their main function is not transmitting interpersonal messages but conveying information.

CMC modes differ in their medium variables. Characteristics and constraints of the CMC modes analysed in this study are summarized in Table 1.

Characteristic	**MSN**	**SMS**	**Twitter**
Message size limit	no	yes (max. 160 characters)[i]	yes (max. 140 characters)
Synchronicity of communication	synchronous (real time)	asynchronous (deferred time)	asynchronous (deferred time)
Visibility	private	private	public, sometimes private[ii]
Level of interactivity	one-to-one, sometimes many-to-many (group conversation)	one-to-one, sometimes one-to-many (broadcast message)	mostly one-to-many, sometimes one-to-one (direct message)
Technology	computer	mobile phone	mobile phone or computer
Channel of communication	multimodal	textual	multimodal

Table 1: medium variables of three CMC modes

[i] Except for concatenated text messages: messages linked together when the limit is exceeded.

[ii] Only direct messages (DMs) are private on Twitter.

2.2 The language of computer-mediated communication

CMC language, which can differ markedly from the standard language, has been described with numerous terms, including 'text/SMS/IM/chat/Internet/cyber language/speak', 'text talk', 'textish', and 'textese'. There are various reasons for CMC users to disregard standard language conventions. First of all, because efficiency comes first: effective CMC requires speed rather than correctness (Silva 2011). Another reason is that some CMC modes are limited in message size, which makes brevity crucial: tweets are limited to 140 characters, text messages to 160. Furthermore, orthographic deviations are frequently used to increase expressiveness: they can compensate for paralinguistic features that are present in speech but lack in writing – prosody, such as stress and volume, and body language, so gestures and facial expressions (Thurlow & Brown 2003). Finally, deviating from the standard spelling and grammar is seen as playful and cool by many young people: they mark in-group belonging and are part of youth culture (Bergs 2009).

Previous research has identified several features characteristic of CMC language, where the writing is affected by the brevity, speed, and creativity of CMC. In terms of orthography, CMC language includes emoticons/smileys, such as :D indicating great joy and ;-) symbolising a wink, and abounds with 'textisms', i.e. spellings deviating from the standard language (see section 3.2 for examples). Textisms are not just used in CMC in English, but also, for example, in German (Bieswanger 2006; Bergs 2009), French (Rúa 2005; Anis 2007; Fairon & Klein 2010), Italian (Pietrini 2001; Herring & Zelenkauskaite 2009), Spanish (Rúa 2005; Alonso & Perea 2008), Portuguese (Silva 2011), Finnish (Plester et al. 2011), and Swedish (Hård af Segerstad 2002). A syntactic feature of CMC language is the omission of function words (Crystal 2006; Frehner 2008; Herring 2012). For example, the sentence "will leave hotel 3 Feb," in which a personal pronoun, article, and preposition have been omitted, is a perfectly acceptable CMC variant of "I will leave the hotel on 3 February." A lexical feature is the use of borrowings (Crystal 2008; Frehner 2008; De Decker & Vandekerckhove 2012): CMC in languages other than English often contains English loan words and loan

textisms, such as *lol* (for *laughing out loud*) and *btw* (*by the way*). Graphically speaking, CMC language stands out for the addition of images, videos, or other multimedia; for the use of colour; and for the inclusion of hyperlinks. Whether these features also occur in Dutch CMC has not been systematically analysed yet. The present paper, focusing on the orthographic feature of textisms, is the first step of such an analysis.

Deviations from the standard language are what catches the eye in computer-mediated messages. Still, this does not mean that CMC language is entirely riddled with opaque abbreviations and rebuses. Crystal (2006:128) rightly points out that the "graphological deviance" in CMC messages is not universal. In fact, the extent to which CMC users deviate from standard language depends on diverse factors. The first of these is individual user characteristics, such as age, gender, regional background, ethnic background, familiarity with CMC, and personal preferences. Secondly, it depends on situational factors, such as discourse topic, recipient of the message, and communicative intent. It is also subject to particulars of the medium: CMC modes differ in message size limits, synchronicity, level of interactivity, and technology (as shown in Table 1). All these factors make CMC language stylistically diverse. This paper looks into the effects of both medium and user age on Dutch CMC language, by investigating writings of three CMC modes (MSN, SMS, Twitter) and two age groups (adolescents, young adults).

3. Methodology

The method employed in this study is a register analysis, investigating how Dutch CMC writings deviate from Standard Dutch. The research is of a quantitative nature: the frequencies of a large set of linguistic features were charted. This paper focuses on the most striking orthographic feature of CMC, namely textisms. It was also investigated how the independent variables age group and CMC mode affect this linguistic feature.

3.1 Materials

The corpus of CMC writings used so far contains instant messaging conversations, text messages, and microblogs. These have been obtained from SoNaR ('STEVIN Nederlandstalig Referentiecorpus', see Sanders 2012; Treurniet et al. 2012; Treurniet & Sanders 2012; Oostdijk et al. 2013). SoNaR is a reference corpus of contemporary written Dutch, including a variety of text sources, both conventional text types and texts from new media. SoNaR includes texts from the Netherlands and Flanders; Flemish texts were excluded in this study. Since the texts in SoNaR are accompanied by metadata on e.g. age, gender, and residency, it was possible to select texts by adolescents (12–17) and young adults (18–23).[2] The selected texts were written under non-experimental conditions between 2009 and 2011. Specifics can be seen in Table 2.

[2] These groups were chosen so that they match the age groups of participants who will partake in later studies in my PhD project about the impact of computer-mediated communication on literacy.

CMC mode	Year(s) of collection	Age group	Mean age	No. of words	No. of conversations or contributors[i]
instant messaging: MSN	2009–2010	12–17	16.2	45,051	106
		18–23	19.5	4,056	21
		total		49,107	127
text messaging: SMS	2011	12–17	15.4	1,009	7
		18–23	20.4	23,790	42
		total		24,799	49
microblogging: Twitter	2011	12–17	15.9	22,968	25
		18–23	20.6	99,296	83
		total		122,264	108
		grand total		196,170	284

Table 2: overview of selected CMC writings

[i] Number of conversations: MSN, number of contributors: SMS, tweets.

3.2 Classification

A comprehensive taxonomy of the different textism types that occur in Dutch CMC was made on the basis of Thurlow and Brown's (2003) and Plester, Wood, and Joshi's (2009) classifications of textisms in English CMC. It is presented here with Dutch examples:[3]

- initialism (alphabetism, acronym): first letters of each word/element in a compound word, phrase, (elliptical) sentence, or exclamation, e.g. *sv* < *samenvatting* ('summary'), *hvj* < *hou van je* ('love you'), *omg* < *o mijn God* ('oh my God');
- contraction: omission of letters (mostly vowels) from middle of word, e.g. *vnv* < *vanavond* ('tonight'), *grtjs* < *groetjes* ('greets'), *idd* < *inderdaad* ('indeed');
- clipping: omission of final letter of word (mostly silent *-n* or *-t*), e.g. *lache* < *lachen* ('laugh'), *truste* < *trusten* ('good night'), *nie* < *niet* ('not');
- shortening (truncation): dropping of ending or occasionally beginning of word, e.g. *eig* < *eigenlijk* ('actually'), *wan* < *wanneer* ('when'), *knuf* < *knuffel* ('hug');
- phonetic respelling: substitution of letter(s) of word by (an)other letter(s), while applying accurate grapheme-phoneme patterns of the standard language:

a) abbreviation, e.g. *fyn* < *fijn* ('nice'), *ff* < *effen* ('for a sec'), *nix* < *niks* ('nothing');
b) replacement, e.g. *ofso* < *ofzo* ('or something'), *jonguh* < *jongen* ('boy'), *owk* < *ook* ('also');
c) extension, e.g. *heej* < *hé* ('hey'), *okee* < *oké* ('okay'), *errug* < *erg* ('very');

- single letter/number homophone: substitution of entire word by phonologically resembling or identical letter/number, e.g. *k* < *ik* ('I'), *n* < *een* ('a'/'an'), *t* < *het* ('it'), *4* < *for*;
- alphanumeric homophone (rebus): substitution of part of word by phonologically resembling or identical letter(s) and/or number(s), e.g. *suc6* < *success* ('good luck'), *w88* < *wachten* ('wait'), *btje* < *beetje* ('little');

[3] A less exhaustive version of this taxonomy can be found in Verheijen (2013).

- reduplication: repetition of letter(s), e.g. *zoooo* < *zo* ('so'), *neeeee* < *nee* ('no'), *superrr* < *super* ('super');
- visual respelling: substitution of letter(s) by graphically resembling non-alphabetic symbol(s) (special character or number), e.g. *Juli@n* < *Julian* ('Julian'), *m%i* < *mooi* ('pretty'), *c00l* < *cool* ('cool');
- accent stylisation: words from casual, colloquial, or accented speech spelled as they sound, e.g. *hoezut* < *hoe is het* ('how are you doing'), *kweenie* < *ik weet het niet* ('I don't know'), *lama* < *laat maar* ('never mind');
- inanity: "nonsensical transmogrification" of word (Craig 2003:120), e.g. *plezierhr* < *plezier* ('fun'), *helaasje* < *helaas* ('too bad'), *laterz* < *later* ('later');
- standard language abbreviation: abbreviation that is part of the standard language, e.g. *aug* < *augustus* ('August'), *bios* < *bioscoop* ('cinema'), *info* < *informatie* ('information');
- unconventional use of spacing, punctuation, diacritics, and capitalisation.

Example (1), an excerpt from an MSN dialogue, illustrates how textisms were classified (all textisms are underlined):[4]

1. *hooooooooooowj*
   keb net de film klein beetje gmonteerd, ziet er strak uit jonguh!:D
   keb uhm in zwartwit oude film style staan nu is eg fat
   mja ben wieder weg
   kom strx nog trug
   mzzzzzzzzl (MSN0000001099)

This fragment contains the following textisms:

- *hooooooooooowj* < *hoi*: phonetic respelling (extension) + reduplication of letter;
- *keb* < *ik heb*: accent stylisation;
- gmonteerd < gemonteerd: contraction;
- *jonguh* < *jongen*: phonetic respelling (replacement);
- *keb* < *ik heb*: accent stylisation;
- *uhm* < *hem*: phonetic respelling (replacement);
- *zwartwit oude film style* < *zwart-witoudefilmstyle*: omission of hyphen + overuse of spacing;
- *eg* < *echt*: phonetic respelling (abbreviation) + clipping;
- *fat* < *vet*: phonetic respelling (replacement);
- *strx* < *straks*: contraction + phonetic respelling (abbreviation);
- *trug* < *terug*: phonetic respelling (abbreviation);
- *mzzzzzzzzl* < *mazzel*: contraction + reduplication of letter.

As can be seen, a single textism can include multiple textism types, as is the case for *hooooooooooowj*, *zwartwit oude film style*, *eg*, *strx*, and *mzzzzzzzzl*; these have been coded for all types present. Multiple words can also represent one textism, as with *zwartwit oude film style*, where elements of a compound word are separated with spaces.

[4] No English translations have been provided for this example and some others, because what matters is the orthographic *form* of the words rather than their *meaning*. In addition, since textisms deviate from the standard language, they are practically untranslatable on many occasions.

3.3 Procedure

Textisms have been identified and classified manually. After initial data coding, all the data were checked to make sure no textisms were overlooked and to filter out any possible misclassifications. The results presented here concern the total number of textisms and the number of textism types.[5] Figures and tables show the frequencies standardised per 10,000 words, because the subcorpora differ in their total amount of words (as shown in Table 2 above). Statistical tests were conducted on the unstandardised, raw frequencies.

4. Results and discussion

Figure 2 shows the results for all textisms per CMC mode – instant messages (MSN), text messages (SMS), and microblogs (tweets).

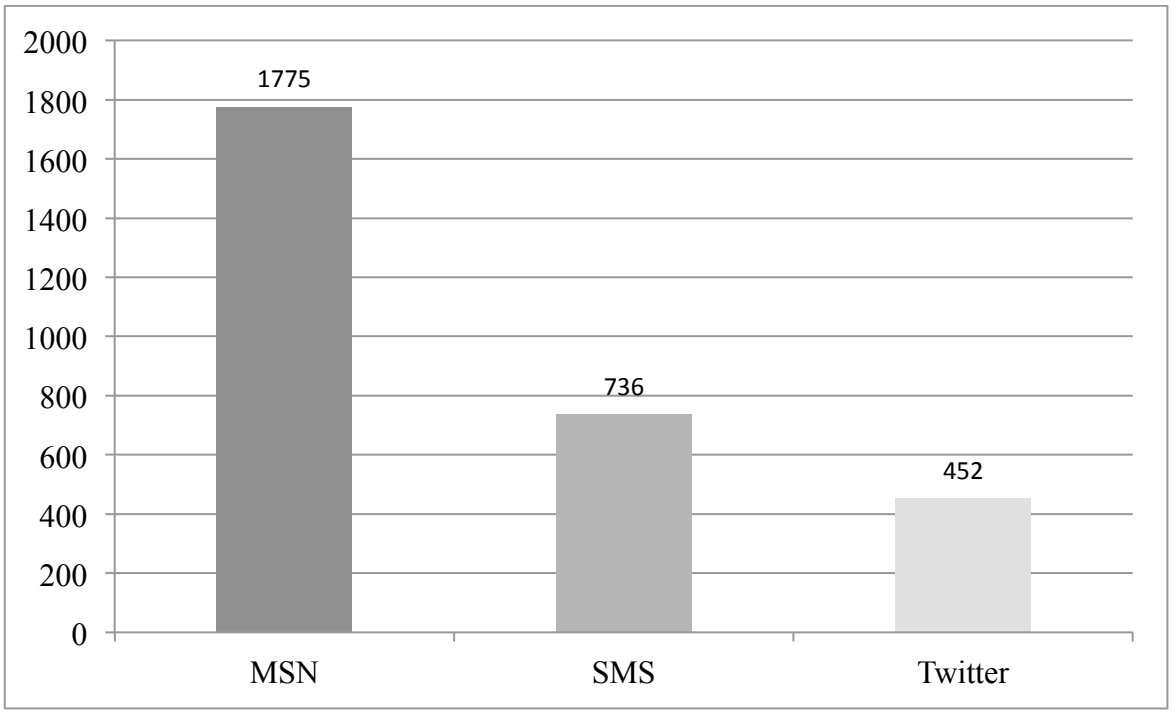

Figure 2: total no. of textisms, per CMC mode

The graph makes it clear that CMC modes diverge in frequency of textisms. A Pearson chi-square test confirms that there was a significant association between CMC mode and total number of textisms (χ^2 (2, n = 196,170) = 8183.20, p < .001). The genre of instant messages contains the most textisms and so deviates most from Standard Dutch in terms of orthography, whereas writings from microblog Twitter contain the least textisms. This result supports the view that we should not approach CMC as one way of communicating (Hård af Segerstad 2002). Rather, each CMC mode is a unique mode of communication with its own orthographic conventions, caused by a complex interplay of medium variables (see Table 1).

[5] The sum of textism types has also been computed, where textisms such as *strx*, which belong to two types, are counted twice. Since the sum is higher than the total number of textisms, it was chosen to report the latter instead, in order not to exaggerate textism use in the CMC writings.

Results show that Dutch youngsters' text messages and tweets contain fewer textisms than their instant messages. This might, in part, be attributed to the fact that these two CMC modes are asynchronous, with messages exchanged sequentially over time, which provides users with more time to edit their writing and to filter out textisms. Instant messaging, by contrast, is synchronous,[6] which makes it direct, immediate, and rushed: users have to respond rapidly to keep up with the conversational pace in order to maintain the floor, which gives them less time to revise their writing and check their spelling.

The lower frequency of textisms in tweets might be caused by their public character: most tweets are 'one-to-many', either visible for everyone to read or for one's entire list of followers. The presence of this larger audience may possibly constrain tweeters to adhere more to standard language norms, so as to avoid being chided for their spelling. Tweets where fierce criticism is passed on "spelling errors" in CMC language illustrate this point; examples (2)–(3) are from adolescents and (4)–(10) from young adults. Metalinguistic terms here are, for instance, *correct schrijven* ('correct writing') in (3), *typvouten/typfouten* ('typos') in (4) and (6), *taalfouten en incorrecte zinnen* ('language errors and incorrect sentences') in (5), and *spelfouten* ('misspellings') in (9). Instant messages and text messages, on the other hand, are sent only to selected recipients, usually friends and family, who may be less prone to point out their communication partner's deviations from the standard language, in order not to harm their relationship.

2. *@roderickmathieu Ellende schrijf je dus met dubbel L. Iets met #incapabele mensen en #opvoeden*
   '@roderickmathieu *Ellende* is written with double L. Something about #incompetent people and #education' (tweet0000000029)
3. *"@TOBIASGROOT: WAAROM KAN NOG STEEDS NIEMAND HET WOORD 'SOWIESO' CORRECT SCHRIJVEN?!" dat is zo'n groot irritatie-factor*
   '"@TOBIASGROOT: WHY IS EVERYONE STILL UNABLE TO WRITE THE WORD '*SOWIESO*' CORRECTLY?!" that is such a big factor of annoyance' (tweet0000000072)
4. a. *@brigadierREUSEL hij vertelt, stam + t. ;)*
      '@brigadierREUSEL *hij vertelt*, stem + t. ;)'
   b. *@orkestmagic typvouten zijn uit den boze.*
      '@orkestmagic typos are absolutely forbidden.' (tweet0000000075)
5. *ik snap dat ze blij zijn, maar zijn al die taalfouten en incorrecte zinnen nodig? 'KADDAFI GEDOOD' http://t.co/vJQeFiFF #nuandroid*
   'i get that they are happy, but are all those language errors and incorrect sentences necessary? 'KADDAFI KILLED' http://t.co/vJQeFiFF #nuandroid' (tweet0000000111)
6. *@mrlkrmr nog steeds typfouten? je hebt nu geen excuus meer hè :P*
   '@mrlkrmr still typos? you no longer have an excuse eh :P' (tweet0000000136)
7. *@IzJoen Ik wordt? Met dt? Ff serieus man, laat je testen op dyslexie ofzo...*
   '@IzJoen *Ik wordt*? With dt? Seriously man, have yourself tested for dyslexia or something... ' (tweet0000000178)
8. a. *TIS GVD FLUORESCEREND*
      'ITS *FLUORESCEREND* GODDAMNIT'
      Fluorizerend
      '*Fluorizerend*'
   b. *"Muggenbulden". Word gek hier*

[6] In reality, CMC cannot be as completely synchronous as spoken interaction, because of the time that inevitably passes between the typing and the receiving of a message. Therefore, it has also been called 'quasisynchronous' or 'near synchronous' (Hård af Segerstad 2002).

'"*Muggenbulden*". Going crazy here'
"Ik heb niks geruikt". Het wort steeds gekker
'"*Ik heb niks geruikt*". Its going too far'

c. *"dan maar een keer m'n rug versjouwen". Niemand kan er tegenwoordig noch Nederlands.*
'"*dan maar een keer m'n rug versjouwen*". No one can write Dutch anymore nowadays.' (tweet0000000194)

9. *@milouvangaans van jouw spelfouten krijg ik nog veel meer stress!*
'@milouvangaans your misspellings cause me even more stress!' (tweet0000000273)

10. *@silvertje Ik scheldt?! #foei ;-)*
'@silvertje *Ik scheldt*?! #shameonyou ;-)' (tweet0000000399)

The distributions of textism types within the CMC modes are shown in Figure 5 in appendix A. This reveals which types are principally responsible for the differences observed in Figure 2. Unfortunately, discussion of these distributions is beyond the scope of this article.

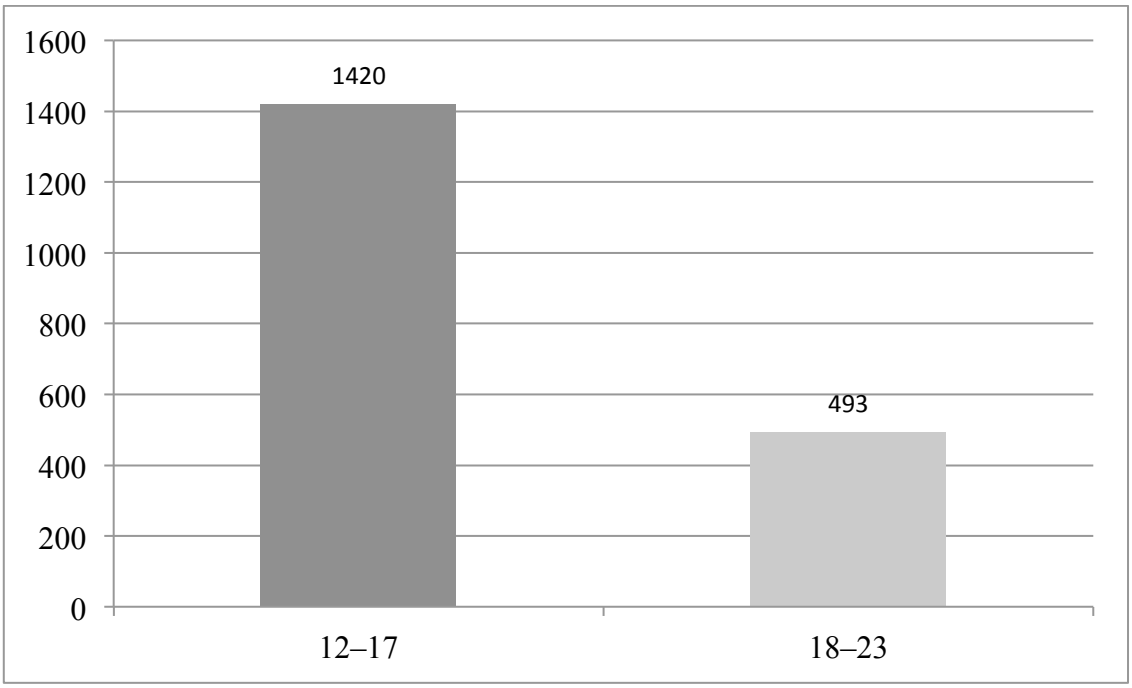

Figure 3: total no. of textisms, per age group

Figure 3 shows the results for all textisms per age group, separating adolescents between the ages of 12 and 17 from young adults between 18 and 23. It reveals that age affects the frequency with which textisms are used in CMC writings: taking all CMC modes together, adolescents use many more textisms than young adults. The relationship between age group and total number of textisms was significant (χ^2 (1, n = 196,170) = 5114.28, p < .001). At this point, we can only speculate about what causes this age effect, but it is possible that young adults regard textisms as somewhat childish. This is evident in the comments made by participants in Wood et al.'s (2011) study that using textisms was considered immature. Similarly, Grace et al. (2014) suggest that young adults' lower use of textisms may be occasioned by social pressure not to come across as immature. Figure 6 (appendix B) shows how different textism types are represented in the age groups. Again, the bulk of the differences in Figure 3 is brought about by only some of the textism types.

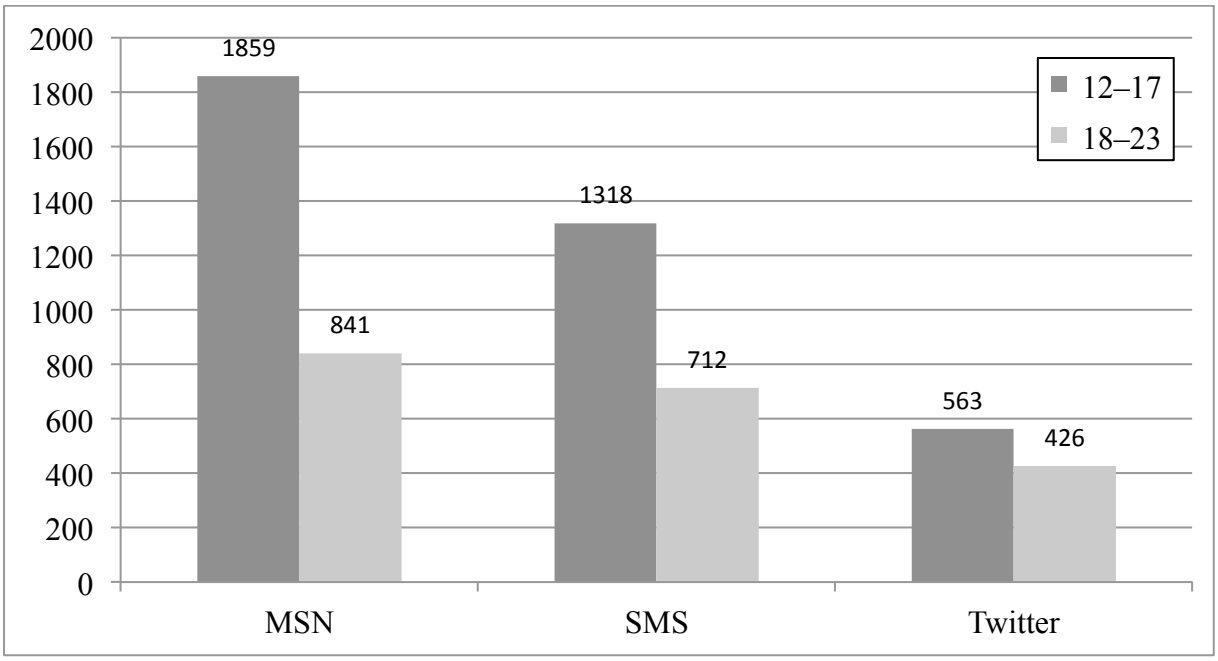

Figure 4: total no. of textisms, per CMC mode and age group

The results for all textisms, distinguishing between CMC modes as well as age groups, are presented in Figure 4. In all three CMC modes, textisms are used more by adolescents than young adults. A three-way loglinear analysis produced a final model that retained all effects, with a likelihood ratio of χ^2 (0) = 0, p = 1. This indicated that the highest-order interaction (CMC mode x age group x textisms) was significant (χ^2 (2, n = 196,170) = 100.81, p < .001). The difference is quite large for instant messages and text messages, but smaller for microblogs, because adolescents use fewer textisms in tweets.

Examples (11)–(16) present instant messages, text messages, and tweets by adolescents and young adults. These examples reflect the continuum shown in Figure 4: most textisms (again, all are underlined) are used by adolescents on MSN, while the lowest number of textisms is used by young adults on Twitter.

11. *mwa tvalt mee hoor, tis altijd IETSJE minder snapje, maar keb strxx ff wrongturn gdraaid, nouja tis gwoon goed omtkijke, je denkt niet van hmmm geen goede kwaliteit ofsow snapje, tis gwoon goed ma ligt natuurluk owk aan je film* (MSN0000001111)
12. *tineke wist het ook niet blijkens haar meel van vorige week* (MSN0000001370)
13. *Heeeeeeeeeee jarige!!! Happy birthday! Ik hoop dat je een leuke verjaardag krijgt ;^) Xx Love you skot* (SMS0000000176)
14. *Hey. Ik ben waarschijnlijk pas om 2 uur in Dordt.. De trein reed net voor m'n neus weg -_-* (SMS0000000153)
15. *@bernoutjee IK HEB EEN KOEKIEMONSTER SHIRT GEKOCHT :D* (tweet0000000147)
16. *Ben benieuwd naar bijeenkomst met schoolbesturen vlgde week. En uiteraard ook naar debat in #raad024 op 16 november. #schoolwijzer* (tweet0000000598)

In the instant message by an adolescent (11), the majority of words are textisms, representing various types. The instant message produced by a young adult (12) shows omission of capitalisation with a proper name (*tineke* < *Tineke*) and a phonetic replacement (*meel* < *mail*).

The text message by an adolescent (13) contains reduplication of letters and omission of a diacritic (*Heeeeeeeeeee* < *Hé*) and reduplication of exclamation marks (*!!!*), as well as an accent stylisation (*skot* < *schat*). The text message (14) and tweet (16) written by young adults only contain contractions (*Dordt* < *Dordrecht*, *vlgde* < *volgende*). The tweet by an adolescent (15) obviously contains overuse of capitalisation. The variety in CMC language becomes apparent through these examples.

Table 3 below shows the top five textism types (again standardised per 10,000 words), separated per age group and CMC mode. It shows the age impact on youths' preferences for particular textisms types: adolescents and young adults prefer to use different types. Young adults mostly omit capitalisation, which is no great deviation from Standard Dutch, and they use many standard language abbreviations, which are typical of CMC language but are nevertheless part of Standard Dutch. This age group thus exhibits a more conventional attitude towards orthography. With adolescents, phonetic respellings are quite popular. These represent a kind of word play, since they involve creative experimentation with the existing grapheme-phoneme patterns of the language. This may be attributed to the so-called adolescent peak (Holmes 1992), which entails that youths are most non-conformist in their linguistic behaviour around the ages of fifteen and sixteen.

MSN	SMS	Twitter
	12–17	
1. phonetic respelling (541)	1. phonetic respelling (218)	1. omission of capitalisation (98)[i]
2. clipping (299)	2. omission of capitalisation (198)	2. reduplication of letter (90)
3. omission of spacing (245)[ii]	2. standard language abbreviation (178)	3. phonetic respelling (60)
4. omission of capitalisation (195)	4. contraction (119)	4. overuse of capitalisation (53)
5. contraction (162)	4. single letter homophone (119)	5. stand. lang. abbreviation (52)
	18–23	
1. omission of capitalisation (375)	1. omission of capitalisation (171)	1. omission of capitalisation (75)
2. stand. lang. abbreviation (254)	2. stand. lang. abbreviation (95)	2. stand. lang. abbreviation (54)
3. reduplication of letter (64)	3. phonetic respelling (77)	3. initialism (42)
3. omission of diacritic (64)	4. omission of diacritic (57)	4. phonetic respelling (37)
4. phonetic respelling (59)	5. single letter homophone (47)	5. overuse of spacing (31)

Table 3: top five textism types, per CMC mode and age group

i Omissions of capitalisation that occurred sentence-initially or in hashtags (in tweets) were excluded, so only capitalisation omitted from proper names and abbreviations was counted.
ii Omissions of spacing after punctuation marks and in hashtags were excluded; only spaces omitted between words were included in the counts.

5. Conclusions and future research

As part of an ongoing corpus study into the register of CMC language, the results observed thus far allow us to draw several conclusions. In terms of orthography (which is, according to Crystal 2006, most distinctive of CMC language), the written CMC of Dutch youngsters from twelve to twenty-three years old greatly deviates from Standard Dutch. The CMC modes selected from the SoNaR corpus deviate to different extents, instant messages containing the most textisms and microblogs the least, which confirms the impact of CMC mode. The impact of age group has also been detected: adolescents between twelve and seventeen use many more textisms in CMC than young adults between eighteen and twenty-three. This occurs across the board in CMC writings, but the difference is greater for instant messages

and text messages than for tweets. Moreover, adolescents and young adults prefer to use different textism types: while adolescents use many playful, self-invented spellings, young adults achieve the brevity and speed required in CMC by employing many standard language abbreviations. All this suggests that youths' written CMC clearly deviates from Standard Dutch, at least where orthography is concerned. This shows the potential of interference of youngsters' informal CMC register with their more formal school register – an issue which will be addressed in future studies part of this project into the impact of CMC on literacy.

Of course, linguistic features from other dimensions of writing, namely syntax and lexis, need to be added to this register analysis, to determine whether Dutch youngsters' CMC writings deviate from Standard Dutch in more than just orthography. In addition, data from other CMC modes which are not in SoNaR, but are at present very popular among Dutch youths need to be collected and analysed, such as WhatsApp and Facebook. Ultimately, this will yield linguistic writing profiles which characterise the language of various CMC modes. The CMC writings can then also be compared to samples of school writings produced by youngsters of similar ages, to explore the differences between these registers. It would also be interesting to interview youths to discover why they believe they use (specific types of) textisms; such self-reports could add valuable insights to this corpus study.

It is fascinating to observe that Dutch youngsters' orthography in their CMC writings can deviate from the standard language to a considerable extent and yet despite these deviations, the message being communicated largely remains understandable. Only on certain occasions communication breaks down, as is evident in the MSN conversation in (17), where the omission of spacing in and shortening of *erma* cause some textism confusion:

17.	*heb het erma is over;)*	'talk about it sumtime orso;)'	
	?	'?'	
	erma?	'orso?'	
	aaaaaah	'aaaaaah'	
	haha	'haha'	
	ja ik heb het er over:P	'yes I'll talk about it:P'	
	sorry	'sorry'	
	er maar	'or something'	
	:P	':P'	
	;)	';)'	(MSN0000001125)

This example shows that too much idiosyncrasy in orthography can impair the recognisability of a word, thereby causing problems of intelligibility and leading to unsuccessful communication. So although the use of textisms by youths is, on the whole, creative and cost-effective, their freedom to deviate from standard language norms is not unlimited.

Acknowledgements

My thanks go to Wilbert Spooren and Ans van Kemenade for their comments on previous drafts of this paper. Any remaining errors are my sole responsibility.

References

ALONSO, E., & M. PEREA. 2008. "SMS: Social and cognitive impact." *Escritos de Psicología* 2(1): 24–31. http://www.escritosdepsicologia.es/eng/previousnumbers/i_vol2num1/i_vol2num1_4.html.

ANIS, J. 2007. "Neography: Unconventional spelling in French SMS text messages." In: B. Danet & S.C. Herring (eds.) *The Multilingual Internet: Language, Culture, and Communication Online*. New York: Oxford UP. pp. 87–115.

BERGS, A.T. 2009. "Just the same old story? The linguistics of text messaging and its cultural repercussions." In: C. Rowe & E.L. Wyss (eds.) *Language and New Media: Linguistic, Cultural, and Technological Evolutions*. Cresskill, NJ: Hampton Press. pp. 55–73.

BIESWANGER, M. 2006. "2 abbrevi8 or not 2 abbrevi8: A contrastive analysis of different shortening strategies in English and German text messages." *Texas Linguistics Forum* 50. http://studentorgs.utexas.edu/salsa/proceedings/2006/Bieswanger.pdf.

CRAIG, D. 2003. "Instant messaging: The language of youth literacy." *The Boothe Prize Essays 2003*: 116–133.

CRYSTAL, D. 2006. *Language and the Internet*, second edition. Cambridge: Cambridge UP.

CRYSTAL, D. 2008. *Txtng: The Gr8 Db8*. Oxford: Oxford UP.

DE DECKER, B., & R. VANDEKERCKHOVE. 2012. "English in Flemish adolescents' computer-mediated discourse: A corpus-based study." *English World-Wide* 33(3): 321–352.

DIJCK, J. VAN. 2013. *The Culture of Connectivity: A Critical History of Social Media*. New York: Oxford UP.

FAIRON, C., & J.R. KLEIN. 2010. "Les écritures et graphies inventives des SMS face aux graphies normées" ["The inventive writings and scripts of the SMS vis-à-vis normative orthography"]. *Le Francais aujourd'hui* 170(3): 113–122.

FREHNER, C. 2008. *Email - SMS - MMS: The Linguistic Creativity of Asynchronous Discourse in the New Media Age*. Bern: Peter Lang.

GRACE, A., N. KEMP, F.H. MARTIN, & R. PARRILA. 2014. "Undergraduates' text messaging language and literacy skills." *Reading and Writing* 27(5): 855–873.

HÅRD AF SEGERSTAD, Y. 2002. *Use and Adaptation of Written Language to the Conditions of Computer-Mediated Communication*. PhD thesis, University of Gothenburg.

HERRING, S.C. 2012. "Grammar and electronic communication." In: C.A. Chapelle (ed.) *Encyclopedia of Applied Linguistics*. Oxford: Wiley-Blackwell. pp. 1–11.

HERRING, S.C., & A. ZELENKAUSKAITE. 2009. "Symbolic capital in a virtual heterosexual market: Abbreviation and insertion in Italian iTV SMS." *Written Communication* 26(1): 5–31.

HOLMES, J. 1992. *An Introduction to Sociolinguistics*. London: Longman.

HUMPHRYS, J. 2007, September 24. "I h8 txt msgs: How texting is wrecking our language." *Daily Mail*. http://www.dailymail.co.uk/news/article-483511/I-h8-txt-msgs-How-texting-wrecking-language.html.

JACOBS, G.E. 2008. "People, purposes, and practices: Insights from cross-disciplinary research into instant messaging." In: J. Coiro, M. Knobel, C. Lankshear, & D.J. Leu (eds.) *Handbook of Research on New Literacies*. New York: Routledge. pp. 469–490.

MERCHANT, G. 2007. "Writing the future in the digital age." *Literacy* 41(3): 118–128.

OOSTDIJK, N., M. REYNAERT, V. HOSTE, & I. SCHUURMAN. 2013. "The construction of a 500-million-word reference corpus of contemporary written Dutch." In: P. Spyns & J. Odijk (eds.) *Essential Speech and Language Technology for Dutch: Results by the STEVIN Programme*. Heidelberg: Springer. pp. 219–247.

PIETRINI, D. 2001. "X' 6 :-(?": Gli sms e il trionfo dell'informalità e della scrittura ludica ("X' 6 :-(?": The SMS and the triumph of informality and ludic writing). *Italienisch* 46(2): 92–101.

PLESTER, B., M.-K. LERKKANEN, L.J. LINJAMA, H. RASKU-PUTTONEN, & K. LITTLETON. 2011. "Finnish and UK English pre-teen children's text message language and its relationship with their literacy skills." *Journal of Computer Assisted Learning* 27(1): 37–48.

PLESTER, B., C. WOOD, & P. JOSHI. 2009. "Exploring the relationship between children's knowledge of text message abbreviations and school literacy outcomes." *British Journal of Developmental Psychology* 27(1): 145–161.

POSTMA, K. 2011. *Geen paniek! Een analyse van de beeldvorming van sms-taal in Nederland*. Master thesis, VU University Amsterdam.

RÚA, P.L. 2005. "Shortening devices in text messaging: a multilingual approach." *Neuphilologische Mitteilungen* 106(2): 139–155.

SANDERS, E. 2012. "Collecting and analysing chats and tweets in SoNaR." *Proceedings LREC (Language Resources and Evaluation) 2012*: 2253–2256.

SILVA, C. 2011. "Writing in Portuguese chats :). A new wrtng systm?" *Written Language & Literacy* 14(1): 143–156.

SWARTZLANDER, C.L. 2010. *The Language of Texting: The New Language of a Digital Nation*. Master thesis, Annenberg School for Communication.

THURLOW, C. 2006. "From statistical panic to moral panic: The metadiscursive construction and popular exaggeration of new media language in the print media." *Journal of Computer-Mediated Communication* 11(3): 667–701.

THURLOW, C., & A. BROWN. 2003. "Generation txt? The sociolinguistics of young people's text-messaging." *Discourse Analysis Online* 1. http://www.shu.ac.uk/daol/articles/v1/n1/a3/thurlow2002003-paper.html.

TREURNIET, M., O. DE CLERCQ, H. VAN DEN HEUVEL, & N. OOSTDIJK. 2012. "Collecting a corpus of Dutch SMS." *Proceedings LREC 2012*: 2268–2273.

TREURNIET, M., & E. SANDERS. 2012. "Chats, tweets and SMS in the SoNaR corpus: Social media collection." In: D. Newman (ed.) *Proceedings of the 1st Annual International Conference Language, Literature & Linguistics*. Singapore: Global Science and Technology Forum. pp. 268–271.

VERHEIJEN, L. 2013. "The effects of text messaging and instant messaging on literacy." *English Studies* 94(5): 582–602.

WOOD, C., S. MEACHEM, S. BOWYER, E. JACKSON, M.L. TARCZYNSKI-BOWLES, & B. PLESTER. 2011. "A longitudinal study of children's text messaging and literacy development." *The British Journal of Psychology* 102(3): 431–442.

A.J.P. (Lieke) Verheijen
Department of Dutch Language and Culture
Radboud University
postal address: PO box 9103, NL-6500 HD Nijmegen
visiting address: Erasmusplein 1, 06.23, NL-6525 HT Nijmegen, the Netherlands
email :lieke.verheijen@let.ru.nl

Appendix A

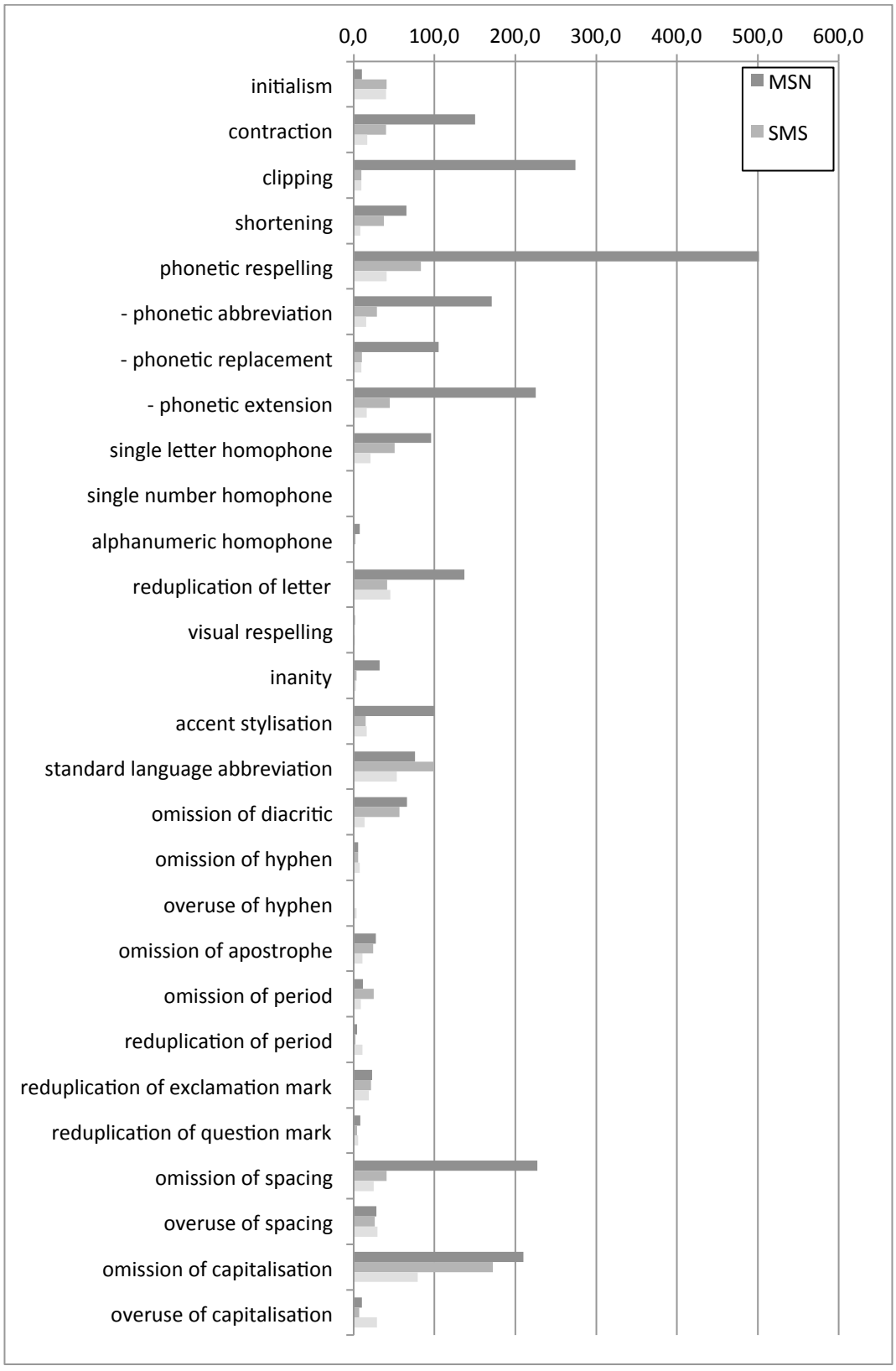

Figure 5: textism types, per CMC mode

Appendix B

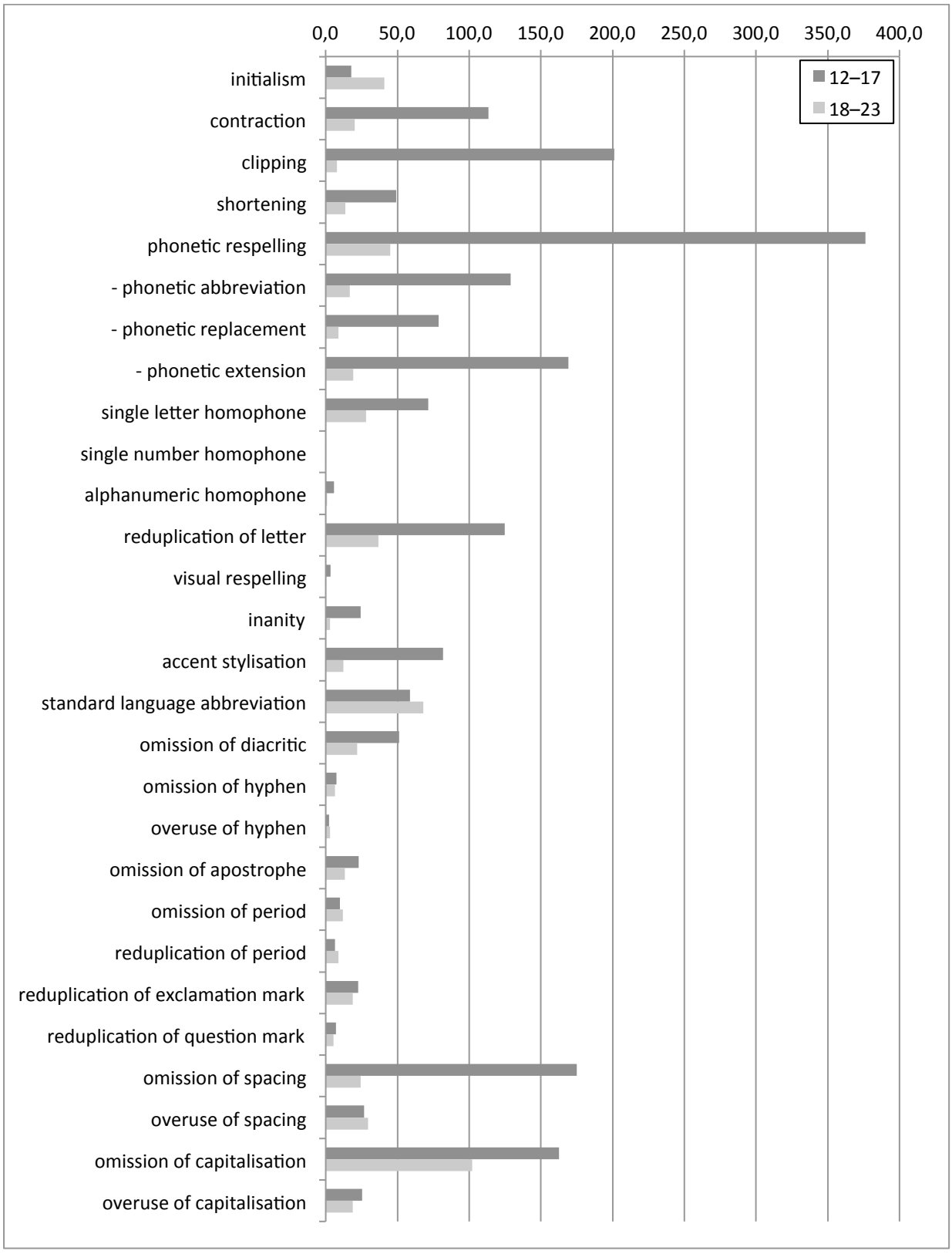

Figure 6: textism types, per age group