WhatsApp with social media slang? Youth language use in Dutch written computer-mediated communication

Lieke Verheijen, *Radboud University*

Abstract

Communication via new media or social media, i.e. computer-mediated communication (CMC), is now omnipresent. The 'CMC language' that youngsters use in such media often diverges from the 'official' spelling and grammar rules of the standard language. Many parents and teachers are thus critical of CMC language, because they view Standard Dutch as a strict norm. Yet among youths it enjoys a certain status, and is regarded as playful, informal, and cool. So an interesting power conflict exists between the overt prestige of the standard language and the covert prestige of CMC language among youngsters. To determine how Dutch youths' language use in computer-mediated messages differs from Standard Dutch, an extensive register analysis was conducted of about 400,000 tokens of digital texts, produced by youths of two age groups – adolescents (12-17 years old) and young adults (18-23 years old), in four social media – SMS text messages; instant messages, viz. MSN chats and WhatsApp messages; and microblogs, namely tweets. This corpus study focuses on various linguistic features of four writing dimensions: orthography (textisms, misspellings, typos), typography (emoticons, symbols), syntax (omissions), and lexis (borrowings, interjections). The results suggest that the variables of age and medium are of crucial importance for (Dutch) youths' online language use.

Keywords: social media, computer-mediated communication (CMC), youth language, writing, WhatsApp

1 INTRODUCTION[1]

The use of social media has increased massively in recent years, both worldwide and in the Netherlands. Communication via these new media is called 'computer-mediated communication,' abbreviated to CMC. This 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). A growing number of communication tools are now at our disposal on computers, mobile phones, and tablets, and their users appear to get younger by the day. In informal CMC, young people often use what can be called 'CMC language' (in Dutch: '*digi-taal*'). The definition of this, as used in this paper, is as follows:

CMC language is a digitally written language variant that is especially used by youths in informal communication via new media, and is characterized, to a greater or lesser extent, by deviations from the standard language norms at different levels of writing, such as spelling, grammar, and punctuation.

In fact, CMC language is an umbrella term which encompasses great variation in itself, depending on various characteristics such as the user who composed the text, the circumstances under which it was written, and the medium that was used to produce it (see section 1.2). So even though language use in CMC has several prominent linguistic peculiarities, computer-mediated texts do not always display the same features to the same extent. Yet because CMC language overall diverges markedly from the standard language, this has caused feelings of resistance among some people, particularly from older generations, as it is feared that these new media pave the way to 'language corruption' or 'language deterioration'. Such sentiments are based, however, on superficial observations, anecdotal evidence, and personal experiences with CMC – not on empirical research. To find out whether these fears are in any way justified, a large-scale systematic register analysis was conducted of digital texts composed in four new media, namely SMS text messaging, instant messaging via MSN Messenger, microblogging on Twitter, and instant messaging via WhatsApp Messenger, written by Dutch youths from two age groups, i.e. adolescents and young adults.

The research question that is central to this paper is as follows: how does the language used by Dutch youths in these social media differ from Standard Dutch? In addition, the following question is addressed: is this language dependent on age group and/or medium? In other words, is the linguistic variation within written CMC by youths from the Netherlands dependent on social and medium-related factors?

---

1   This chapter is a translated, extended, revised, and updated version of a Dutch conference paper by the author (Verheijen 2016).

1.1 New media

Research into new media requires clarity about what this term encompasses. In this day and age, numerous new media exist. Two relatively 'old' new media are text messaging and email, which first became popular two decades ago. Online chats are of a similar vintage, and two main kinds exist: chat rooms hosted on the Internet and instant messaging services, with the latter occurring via four kinds of technologies: mobile phone applications (e.g. *WhatsApp Messenger*, *Telegram*), Internet applications (*Google Hangouts*, *Skype*, formerly *MSN Messenger*), social networking sites (*Facebook chat*), and online gaming networks or virtual worlds (*World of Warcraft*, *Second Life*). Other new media include social networking sites (*Facebook*, *Google+*) and platforms for sharing visual media (*YouTube*, *Instagram*, *Pinterest*). Blogs and microblogs (*Twitter*, *Tumblr*) are also forms of new media. The concept further includes online forums or discussion boards (*4chan*, *FOK!forum*, *VIVA Forum*). This list indicates that new media are extremely varied, and thus the communication that takes place via these various platforms can also be surmised to be rather diverse. That is, each of these media differ in multiple characteristics that may affect the language used in CMC. Table 1 gives an overview of the various media analysed in this paper.

Table 1: Characteristics of four new media.

Medium characteristics	Instant messaging: MSN	Text messaging: SMS	Microblog: Twitter	Instant messaging: WhatsApp
message size limit	no	yes (max 160 characters)[2]	yes (max 140 characters)	no
synchronicity of communication	near-synchronous (real-time)	asynchronous (deferred time)	asynchronous (deferred time)	near-synchronous (real-time)
visibility	private	private	public, sometimes private (direct message)	private
interactivity	one-to-one or some-to-some (group chat)	one-to-one, sometimes one-to-many (broadcast message)	one-to-many, sometimes one-to-one (direct message)	one-to-one or some-to-some (group chat)
technology	computer	mobile phone (or computer)	mobile phone or computer	mobile phone (or computer)
communication channel	multimodal	textual or multimodal[3]	multimodal	multimodal

2   With the exception of concatenated text messages, in which messages are joined if the limit is exceeded.

3   The use of emoticons (see section 2.2) in SMS is textual, because they are composed of typographic characters. Smartphones, however, allow the use of emoji in SMS (but not in the present corpus): this leads to multimodality, because emoji are small images.

1.2 Computer-mediated communication

Certain attributes of CMC language, on various levels of writing, have cross-linguistically emerged from previous research. As for orthography, CMC language is prototypically known for the use of unconventional, non-standard spelling, 'textisms'; that is, transformations of conventionally spelt words.[4] As for typography, emoticons are a key novel feature of such communication (e.g. Silva 2011). Moreover, a frequently mentioned syntactic attribute is the omission of words, in particular function words (Ferrara et al. 1991, Werry 1996, Hård af Segerstad 2002, Crystal 2006, Frehner 2008, Bergs 2009, Winzker, Southwood and Huddlestone 2009, Herring 2012, Wood, Kemp and Plester 2013). A lexical attribute is the use of many English borrowings (Crystal 2008, Frehner 2008, De Decker and Vandekerckhove 2012). Graphical attributes are, for example, the use of hyperlinks and the incorporation of images, sound files, or videos; there can also be multimodality, a "blending of graphic with grapheme" (Carrington 2004: 218).[5]

CMC language thus tends to deviate from the standard language, a phenomenon that has roots in four main causes. Firstly, efficiency and speed are of great importance when communicating via new media, and tempo thus overrules 'correctness.' In addition, some media are limited in message size. For example, a single text message can only contain up to 160 characters, and a tweet no more than 140, so succinctness is crucial in these media. Secondly, words are often typed in computer-mediated messages as they are pronounced in informal spoken language (phonetic writing), to make the writing more like casual speech. Deviations from the standard language can, furthermore, increase expressivity: they can compensate for the lack of paralinguistic and prosodic elements in written (digital) language, such as stress, intonation, and volume, as well as the lack of body language, such as gestures and facial expressions. Androutsopoulos (2011: 149) summarizes these three principles as economy, orality, and compensation. Lastly, many youths like to be creative and original when communicating via new media, and such playing with language can contribute to their social identities. We can infer from this that many deviations in CMC language are functional: they are often resourceful, practical adaptations for which youths, in the context of the current study, make optimal use of the linguistic possibilities of written CMC in order to reach their communicative goals, despite the technological limitations of new media and the pragmatic limitations of written language.

---

4   The term 'textism' is obviously derived from the phrase 'text messaging,' but these unconventional spellings also occur in CMC via other media.

5   Bergs (2009) rightly stresses that not all of these deviations from the standard language were first invented during communication via new media. Some features of CMC language were already present in earlier writing genres, such as telegrams, postcards, informal personal letters, and newspaper headlines.

Still, Crystal (2006: 128) is right when he remarks that "the graphological deviance noted in [new media] messages is ... not universal": digital texts diverge from the standard language to different extents. Such differences stem from a variety of factors (Herring 2001, Hård af Segerstad 2002, Crystal 2006, Crystal 2008, Drouin and Davis 2009, Proudfoot 2011):[6]

- user characteristics, such as age, gender, region, ethnic background, familiarity with textisms, personal preferences;

- situational characteristics, such as conversational topic, (social distance to) receiver of the message, communicative intent;

- medium characteristics, such as a possible message size limit, (a)synchronicity, interactivity, visibility.

All this makes CMC language stylistically diverse. That is why, as Hård af Segerstad (2002: 234) rightly argues, CMC should not be regarded as "one single mode of communication." Rather, each new media user determines their own unique way of communicating every time they compose a digital message, depending on their personal profile, the medium they use for communication, and various situational features.

1.3 Polarization and prestige

CMC language has evoked a range of sentiments. A so-called 'Gr8 Db8' (great debate) exists about CMC language and its impact on reading, writing, and spelling (Crystal 2008), and it has become quite polarized. On the one hand, the language used in new media is negatively described by critics, with terms such as 'language corruption', 'modern scourge', 'linguistic ruin', 'vandalism', 'foe of literacy' and 'bane', while on the other hand, positive terms are used by those who are optimistic about the linguistic potential of CMC, such as 'language enrichment', 'opportunity', 'resource', 'valuable', 'frNd of literacy' and 'blessing'.

Dutch youths' CMC language is thus, as it were, embroiled in a power conflict with Standard Dutch. The standard language has overt prestige, because it is openly esteemed by many as the norm (Labov 1966): 'official' Dutch is dominant within the Netherlands. Although what used to be known as 'Civilized Dutch' (in Dutch: '*Algemeen Beschaafd Nederlands*') is nowadays perhaps less used in spoken language, for one reason due to the rise of 'Polder Dutch' ('*Poldernederlands*': a speech variant that has increased in popularity in the last decades, especially among young highly-educated women, Stroop 2010), many people still regard

---

6 Many of these factors are not exclusive to new media texts: they also explain (in part) other forms of language variation.

Standard Dutch as a strict norm in its written form. They consider the 'incorrect' and inconsistent language use in social media as a detrimental influence on their beloved language. The following reactions by parents and teachers, prompted by an article about 'language errors' by youths, illustrate this:

> "Got the feeling that language deterioration has been going on for years …, particularly among youths, and is getting worse. Some seem to just enjoy communicating in a kind of semi-slang. Maybe also caused by modern communication tools WhatsApp, Facebook etc … in which it is not so important whether something is spelled correctly as long as it is understood by friends."

> ('*Heb het idee dat er al jaren … taalverloedering is, met name onder jongeren, en steeds erger wordt. Sommigen lijken het ook gewoon leuk te vinden om in een soort semi-straattaal te communiceren. Misschien ook veroorzaakt door huidige communicatiemiddelen Whatsapp, Facebook etc…waarin het niet zo van belang is of iets juist gespeld is als het maar door vrienden begrepen wordt.*') (TN 2014)

> "Social media such as Facebook and WhatsApp definitely affect language deterioration"

> ('*Sociale media zoals Facebook en Whatsapp hebben zeker invloed op taalverloedering*') (Robin F 2014)

The following example from a public Internet forum shows similar concerns. A contributor is convinced that social media "cause language corruption": they "sometimes get the impression that with the advent of Facebook & Co, the Netherlands spontaneously became dyslexic collectively" (social media '*leid[en] tot taalverloedering (krijg soms de induk dat met de komst van Facebook & Co Nederland spontaan collectief dyslectisch is geworden)*') (w00t00w 2015). Another forum participant shares this critical outlook and when comparing language use in old and new media, he observes, "With newspapers and publishers, contributors could hardly afford to make a spelling error back then. With social media, this does not matter anymore at all" ('*Bij kranten en uitgevers konden de inzenders zich toen nauwelijks een spelfoutje permitteren. Bij de sociale media maakt dat nu allemaal geen bal meer uit*') (EricMM 2015). In short, non-standard language use on social media is criticized openly and often, and in various contexts. The overt prestige of Standard Dutch is also clear from the success of non-academic publications about language 'errors,' such as the immensely popular books and online communities of *Taalvoutjes* (Bogle and Hollebeek 2013), in which Dutch 'language errors' are made fun of.

By contrast, unconventional CMC language enjoys covert prestige among many youths, who value this non-standard language variety. They consider it as playful,

informal, and cool. The use of CMC language is thus part of youth culture (Bergs 2009), may express humour, rebelliousness, and youthfulness (Shaw 2008), and is often used to mark one's social identity (Wood, Kemp and Plester 2013). In this way, CMC language bears resemblances to so-called street language (in Dutch: '*straattaal*'), an urban youth language which is spoken in the streets, particularly in multi-ethnic cities, and is characterized by influences from immigrant languages and American slang. That, too, is an informal youth language which deviates from Standard Dutch, and is therefore regarded with suspicion by many (older) people, whereas many youths consider it as fashionable and cool.[7] Street language and CMC language foster a sense of belonging to a group and help youths create their own social space (De Rooij, in Truijens 2009), and this covert prestige of CMC language also reveals itself through creativity with language in new media, such as novels and poetry written in the form of text messages or tweets. This paper examines the linguistic characteristics to which Dutch youths' CMC language owes its covert status. Put differently, this work investigates in which ways this language variant diverges from Standard Dutch, and whether these divergences are dependent on the variables of medium and age group.

2 MATERIALS AND METHODOLOGY

2.1 Data collection

For this register analysis of new media messages produced by Dutch youths, texts written in three media were selected from SoNaR ('*STEVIN Nederlandstalig Referentiecorpus*', Oostdijk et al. 2013), an existing reference corpus of written Dutch, while additional texts from one further medium, WhatsApp, were also collected. The WhatsApp messages were gathered especially for the present study: a website was created with instructions on how Dutch youths could voluntarily contribute their authentic (private) messages by sending them to a specific email address (Verheijen and Stoop 2016). Data collection was promoted via diverse national and regional media, and an added incentive for young people to donate their messages was a prize raffle among all contributors with the chance to win gift certificates. The final corpus used for this study contains 392,169 tokens of instant messages (MSN chats and WhatsApp messages), text messages, and tweets, composed by youths aged 12 to 23. These were divided into two age groups: adolescents (between the ages of 12 and 17) and young adults (18 up to 23 years old). The specifics of the corpus, and the distribution of tokens over the

---

7   Just like CMC language, street language is a heterogeneous phenomenon. CMC texts from different media and by different users are distinct; likewise, there are different kinds of street language, which cannot be simply lumped together in any formal analysis.

media and age groups, are shown in Table 2. To be clear, messages in the different media – not only those in the added WhatsApp component, but overall – came from different individuals, so the corpus was not longitudinal. Due to the distribution of new media texts in SoNaR, the corpus is unfortunately imbalanced for the independent variables of medium and age group, but this does not skew the tables and figures presented below, because the frequencies of the linguistic features have been normalised per 10,000 words.

Table 2: Corpus of new media texts for analysis.

Medium	Year(s) of collection	Age group	Mean age	# tokens	# chats or contributors[8]
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
Instant messaging: WhatsApp	2015	12-17	14.0	55,865	11 / 84
		18-23	20.4	140,134	23 / 132
		total		195,999	34 / 216
		grand total		**392,169**	

2.2 Data coding

The new media texts were examined quantitatively for various linguistic features that have been found in prior research, carried out on languages other than Dutch, to be relevant for CMC: the orthographic features of textisms, misspellings, and typos; the typographic features of emoticons and symbols; the syntactic feature of omissions; and the lexical features of borrowings and interjections.

The following spelling deviations of Standard Dutch have been classified in the analysis as textisms (adapted from Plester, Wood and Joshi 2009; see also Verheijen 2013):

- **initialism**: first letters of each word/element in a compound word, phrase, (elliptical) sentence, or exclamation (cf. Daniëls' (2009) 'lettero'), e.g. *hw < huiswerk* ('homework'), *gmj < goed, met jou* ('fine, how are you'), *hjb < houd je bek* ('shut up'), *wtf < what the fuck*

---

8 Number of chats: MSN, WhatsApp; number of contributors: SMS, Twitter, WhatsApp.

- **contraction**: omission of letters (mostly vowels) from middle of word (cf. Daniëls' (2009) 'shortje'), e.g. *ltr < later* ('later'), *hzo < hoezo* ('why'), *sws < sowieso* ('in any case')

- **clipping**: omission of final letter of word (mostly silent *-n* or *-t*), e.g. *morge < morgen* ('tomorrow'), *bes < best* ('rather'), *naa < naar* ('to')

- **shortening**: dropping of ending or occasionally beginning of word, e.g. *miss < misschien* ('maybe'), *opdr < opdracht* ('assignment'), *ns < eens* ('some time')

- **phonetic respelling**: substitution of letter(s) of word by (an)other letter(s), while applying accurate grapheme-phoneme patterns of the standard language (resulting in abbreviation, replacement, or extension), e.g. *sgool < school* ('school'), *meel < mail, owkeej < oké* ('okay')

- **single letter/number homophone**: substitution of entire word by a phonologically resembling or identical letter/number, e.g. *k < ik* ('I'), *m < hem* ('him'), *2 < too/to*

- **alphanumeric homophone**: substitution of part of word by phonologically resembling or identical letter(s) and/or number(s), e.g. *opdr8 < opdracht* ('assignment'), *id < idee* ('idea'), *hh < haha*

- **reduplication**: repetition of letter(s) (cf. De Decker's (2015) 'flooding' and Darics' (2013) 'letter repetition'), e.g. *cooool < cool, doeii < doei* ('bye'), *jaaa < ja* ('yes')

- **visual respelling**: substitution of letter(s) by graphically resembling non-alphabetic symbol(s) (special characters or numbers), e.g. *w00t < woot, j@n < Jan*

- **accent stylisation**: words from casual, colloquial, or accented speech spelled as they sound, e.g. *hoessie < hoe is het* ('how are you'), *das < dat is* ('that's'), *eik < eigenlijk* ('actually')

- **inanity**: miscellaneous spelling deviations, e.g. *eeyz < ey, duz < dus* ('so'), *chilliej < chill*

- **standard language abbreviation**: abbreviation that is part of the standard language,[9] e.g. *jan < januari* ('January'), *uni < universiteit* ('university'), *min < minuut* ('minute')

- unconventional use of spacing, punctuation, diacritics and capitalisation (incl. 'all caps,' i.e. entire words or utterances typed in capital letters).

---

9   Of course, standard language abbreviations do not deviate from the 'official' spelling; after all, they are included in dictionaries that codify Standard Dutch. Yet these abbreviations were still included in the present analysis of textisms, since they are also typical of the succinctness and speed of CMC.

In the classification of 'misspellings,' only a number of distinct spellings deviating from Standard Dutch have been coded, and these are deviations that are strongly denounced by prescriptivist linguists or language users. These concern 'spelling errors' with *d/t*, *ei/ij*, *is/eens*, *jou/jouw*, *n* (the letter *n* used to connect two words in Dutch, or final *n*), obsolete spelling, and with borrowings. Only a select group of deviations has thus been classified as 'misspelling'; the rest has been interpreted as textism, despite the fact that these are also regarded as 'incorrect' by those who hold the standard language as the norm for all writing.

The analysis only contains manifest typos (typing errors), where the writer clearly intended to type another word, given the context. Such deviations often differed by only one letter, e.g. *hey boek* ('thy book') instead of *het boek* ('the book').

Emoticons – a portmanteau word of the words 'emotion' and 'icon' – are understood to mean symbols composed of typographic characters (punctuation marks, letters, and/or numbers) which represent facial expressions with emotions, such as :-) (a smiling face, or 'smiley') to indicate joy. These help to express the writer's feelings. Both Western variants, which should be understood by tilting one's head, and Asian/Japanese variants, which can be interpreted at face value (e.g. ^^ and -_-), have been included, although the latter (also called 'kaomoji') only occurred rarely in the present corpus. Emoticons do not include the nowadays popular emoji – which, in the new media analysed here, only occur in WhatsApp: these small, standardised images are not part of typography.

The symbols encountered in this new media corpus are as follows: *&* (and), + (and, plus), = (is, equals), <, >>, --> (arrow), € (euro, money), <3 (heart), *X* or *x* (kiss), *K* or *k* (kiss), *(K)* or *(k)* (kiss), *o* (hug, as in xoxo), *(L)* or *[L]* (love), *(H)* or *(h)* (heart or cool), *(A)* or *(a)* (angel), *(Y)* or *(y)* (yes, okay), and *** (correction, emphasis, or action).

All omissions have been coded and subsequently classified on the basis of the part of speech of the omitted elements: articles, subject pronouns (personal or demonstrative pronouns that function as the grammatical subject), other pronouns (personal/demonstrative pronouns with another grammatical function, such as object or possessive pronouns), auxiliary verbs, copula verbs, lexical verbs, combinations of subject pronoun and verb (plus possibly object pronouns), conjunctions, prepositions, and other elements (e.g. adverbs). Further analysis of these types of omissions was outside the scope of this paper.

The following lexical elements have been classified as borrowings: borrowed words, borrowed phrases, borrowed sentences, borrowed interjections, and borrowed textisms. Words that originate from other languages, but have now been officially acknowledged as part of Standard Dutch, have not been coded. The criterion used to objectively determine whether a word has been acknowledged as

part of Standard Dutch was inclusion in the *Dikke Van Dale Online* dictionary, an authority among Dutch lexicons.

Interjections are expressions or utterances that do not constitute a grammatical constituent of a sentence, but stand on their own. They are mainly used to express sentiment or to imitate sounds, for example onomatopoeias conveying laughter.

These features were identified and classified entirely manually. To increase the reliability of the results, all data were checked twice by the first coder (the author). Moreover, a subset of the data (*n* = 10,010 tokens, a random sample of at least 1,000 from each subcorpus) was also coded independently by a second coder, who before this process began took part in two training sessions with the first coder to get a full grasp of the codebook. The intercoder reliability for this subset of the data was measured with Cohen's *κ*. It was calculated per linguistic feature, to ensure acceptable levels of reliability (except for the omissions, which were only coded by a single coder). Values ranged from 0.68 to 0.92 (see Table 3); the average intercoder reliability was *κ* = 0.83.

Table 3: Reliability coefficients per linguistic feature.

Linguistic features	Kappa
textisms	.92
misspellings	.70
typos	.68
emoticons	.98
symbols	.85
omissions	-
borrowings	.82
interjections	.83

2.3 Data analysis

The results reported here have been separated for medium and age group and normalised per 10,000 words, because the total number of words analysed differs per medium and age group. The results have also been subjected to statistical testing with IBM SPSS Statistics, through seven loglinear analyses and one chi-square test.

The loglinear analyses were performed on the raw frequencies, taking into account the total sample sizes. A hierarchical model was used for these analyses, containing all the lower-order interactions and main effects of the interactions examined. Seven of the eight linguistic features – textisms, misspellings, typos, emoticons, symbols, borrowings, and interjections – were treated as variables

in their own loglinear analyses. Textisms, for instance, were a variable in one analysis (NB: it was thus *not* the case that 'linguistic feature' was a variable in an overall analysis and the different features, such as textisms and misspellings etc., were its levels). For each of the seven linguistic features analysed with loglinear analyses, a separate analysis was conducted with the following variables: 'medium' (MSN, SMS, Twitter, or WhatsApp), 'age group' (adolescent or young adult), and 'linguistic feature' (feature present or absent), which were all weighted by the raw frequencies. The raw frequencies of 'feature absent' were computed as follows: the total number of words per medium and age group, minus the raw frequency of linguistic feature per medium and age group, e.g. for textisms in MSN by adolescents: 45,051 - 8,398 = 36,653. As an example, Table 4 shows what the SPSS data file for the statistical analysis of textisms looked like:

Table 4: Example data file for loglinear analysis: textisms.

MSN_SMS_Twitter_or_WhatsApp	Adolescent_or_young_adult	Textism_or_not	Raw_frequency
MSN	adolescent	textism	8398
MSN	adolescent	no textism	36653
MSN	young adult	textism	347
MSN	young adult	no textism	3709
SMS	adolescent	textism	133
SMS	adolescent	no textism	876
SMS	young adult	textism	1696
SMS	young adult	no textism	22094
Twitter	adolescent	textism	1298
Twitter	adolescent	no textism	21670
Twitter	young adult	textism	4255
Twitter	young adult	no textism	95041
WhatsApp	adolescent	textism	6317
WhatsApp	adolescent	no textism	49548
WhatsApp	young adult	textism	10206
WhatsApp	young adult	no textism	129928

Since the number of instances in the corpus that are *not* omissions cannot be computed (in theory, any number of omissions can exist; irrespective of the total number of words per subcorpus), instead of a loglinear analysis, a chi-square test was conducted on the standardised frequencies of the omissions.

3 RESULTS

The following tables show the findings of the corpus study: Table 5 presents the normalised frequencies and Table 6 the results of the statistical tests.

Table 5: Normalised frequencies of the linguistic features (per 10,000 words).

Linguistic features	Instant messaging: MSN		Text messaging: SMS		Microblogging: Twitter		Instant messaging: WhatsApp	
	12-17 yrs	18-23 yrs	12-17 yrs	18-23 yrs	12-17 yrs	18-23 yrs	12-17 yrs	18-23 yrs
	norm. freq.	norm. freq.	norm. freq.	norm. freq.	norm. freq.	norm. freq.	norm. freq.	norm. freq.
textisms	1864.11	855.52	1318.14	712.90	565.13	428.52	1130.76	728.30
misspellings	24.42	27.12	19.82	6.31	16.11	10.57	24.70	13.27
typos	39.73	22.19	79.29	40.77	29.17	16.42	137.65	57.59
emoticons	690.55	236.69	198.22	356.45	216.39	196.38	83.95	101.97
symbols	16.87	4.93	267.59	237.49	24.82	20.14	39.38	22.41
omissions	518.75	315.58	356.79	479.61	390.54	423.98	620.60	493.17
borrowings	131.41	71.50	148.66	76.92	149.77	114.81	194.76	144.72
interjections	559.81	332.84	317.15	253.05	179.38	114.41	485.10	304.14

Table 6: Results of the statistical tests of the linguistic features.

Linguistic features	Interaction medium × age group × linguistic feature (DF = 3)		Interaction medium × linguistic feature (DF = 3)		Interaction age group × linguistic feature (DF = 1)	
	χ^2	Sig	Partial χ^2	Sig	Partial χ^2	Sig
textisms	97.48	***	3574.71	***	1121.06	***
misspellings	5.61	n.s.	17.84	***	30.02	***
typos	5.68	n.s.	676.18	***	305.62	***
emoticons	174.43	***	3711.52	***	12.14	***
symbols	9.41	*	1461.77	***	36.74	***
omissions	75.14	***	-		-	
borrowings	5.67	n.s.	173.01	***	91.36	***
interjections	3.71	n.s.	1692.75	***	457.63	***

N.s.: non-significant, $p > .05$; significant * $p < .05$, *** $p < .001$. DF: degrees of freedom.
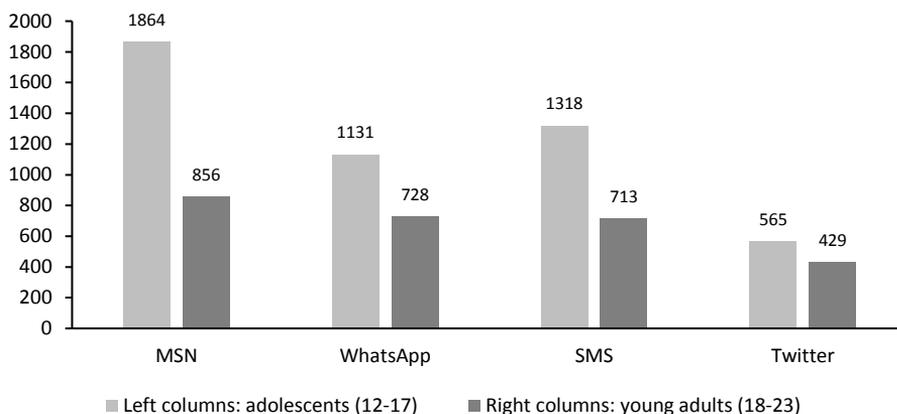
3.1 Orthography

3.1.1 Textisms

Figure 1: Normalised frequencies of textisms.

The statistical test reported in Table 6 shows that the three-way interaction medium × age group × textisms was significant (χ^2 (3) = 97.48, p < .001). Analysis of the normalised frequencies demonstrates that textisms were used more by adolescents than young adults in all media, but that this difference was dependent on medium: it was greatest in MSN chats, in which textisms occurred most, and smallest in tweets, in which they occurred least.

3.1.2 Misspellings
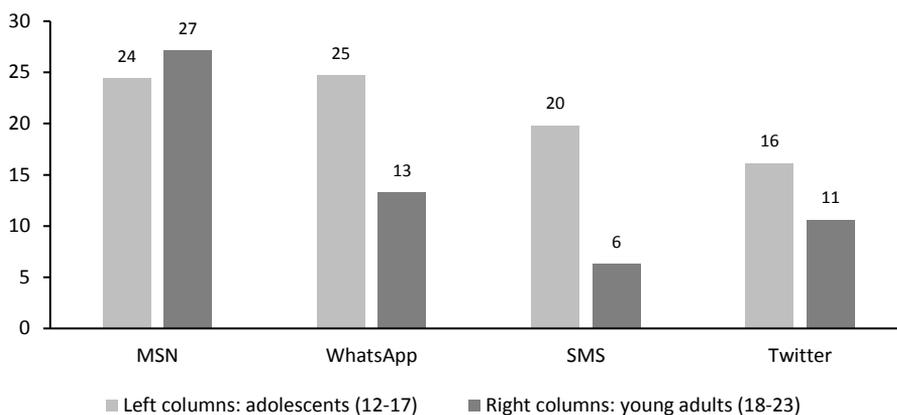
Figure 2: Normalised frequencies of misspellings.

It is apparent from the analysis that the two-way interactions medium × misspellings and age group × misspellings were significant (partial χ^2 (3) = 17.84, $p <$ .001, partial χ^2 (1) = 30.02, $p < .001$). Misspellings occurred more in MSN chats than in the other three media. They were produced more by adolescents than young adults, except in MSN.
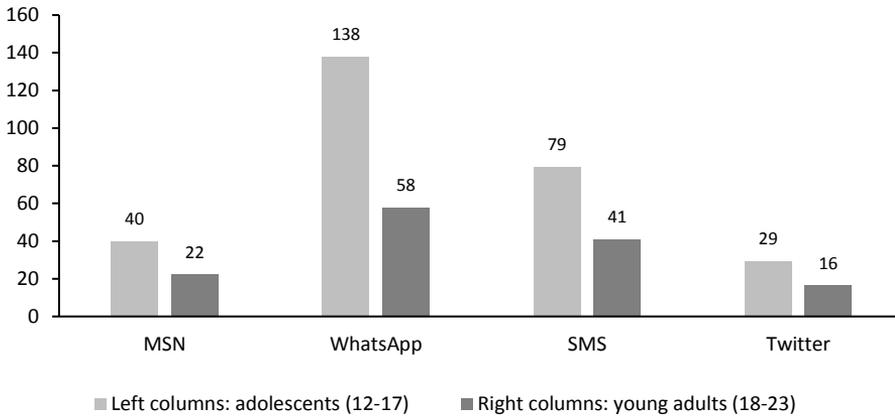
3.1.3 Typos

Normalised frequencies of typos, with values:
- MSN: adolescents 40, young adults 22
- WhatsApp: adolescents 138, young adults 58
- SMS: adolescents 79, young adults 41
- Twitter: adolescents 29, young adults 16

■ Left columns: adolescents (12-17)   ■ Right columns: young adults (18-23)

Figure 3: Normalised frequencies of typos.

The statistical tests show that both two-way interactions, namely medium × typos and age group × typos, were significant (partial χ^2 (3) = 676.18, $p < .001$, partial χ^2 (1) = 305.62, $p < .001$). More typos occurred in WhatsApp messages and then SMS text messages, than in the other two media. Adolescents made more typing errors than young adults in all four media.

3.2 Typography

3.2.1 Emoticons

Statistical tests reveal that the three-way interaction medium × age group × emoticons was significant (χ^2 (3) = 174.43, $p < .001$). In MSN chats, in which emoticons were most frequent, adolescents used many more of these than young adults. The situation was reversed for SMS text messages, in which it

was young adults who used more emoticons. The frequencies of emoticons in WhatsApp, in which emoticons were used least, and on Twitter were close together for the two age groups.
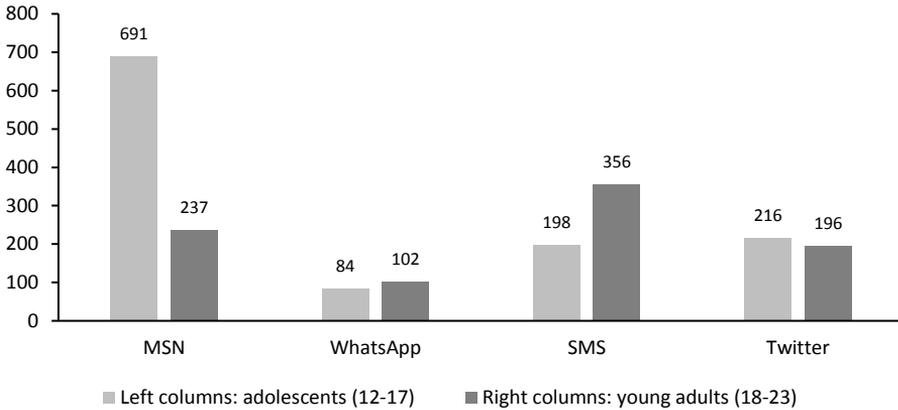

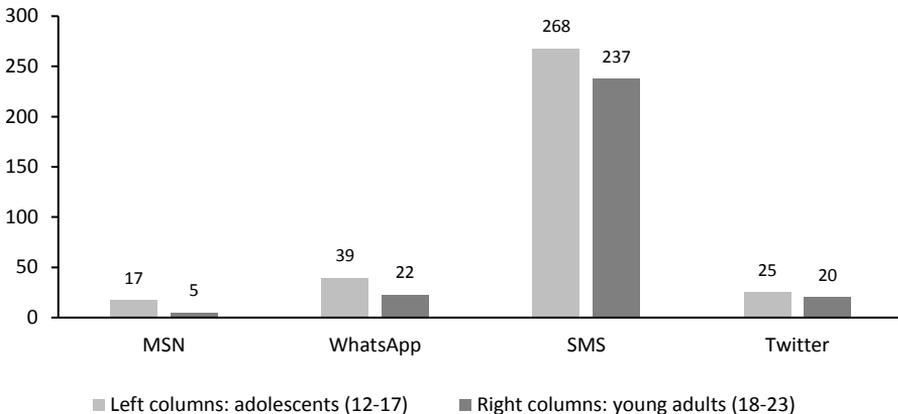**Figure 4: Normalised frequencies of emoticons.**

3.2.2 Symbols

Figure 5: Normalised frequencies of symbols.

Statistical testing shows that the three-way interaction medium × age group × symbols was significant (χ^2 (3) = 9.41, p < .05). Symbols were used much more in SMS text messages than in the other three media, and they were used somewhat more by adolescents than young adults across all media.

3.3 Syntax

3.3.1 Omissions

Figure values: MSN 519/316, WhatsApp 621/493, SMS 357/480, Twitter 391/424. Left columns: adolescents (12-17) Right columns: young adults (18-23)
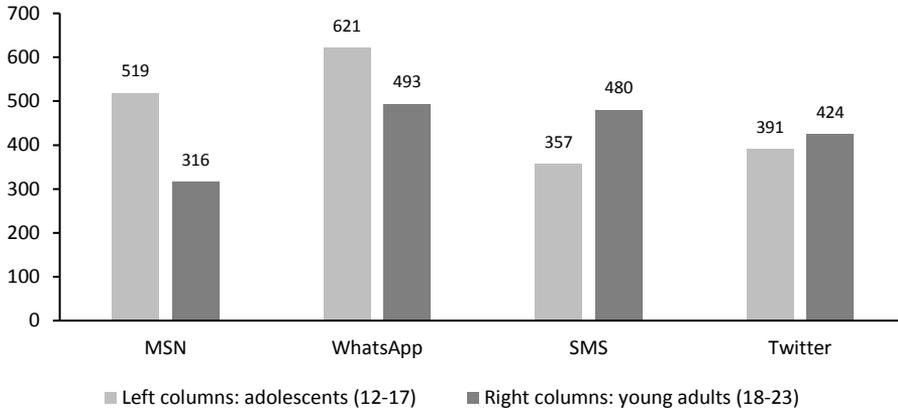
Figure 6: Normalised frequencies of omissions.

The three-way interaction medium × age group × omissions turned out to be significant (χ^2 (3) = 75.14, p < .001). Adolescents used more omissions than young adults in WhatsApp messages and MSN chats, while young adults used more in SMS text messages and tweets.
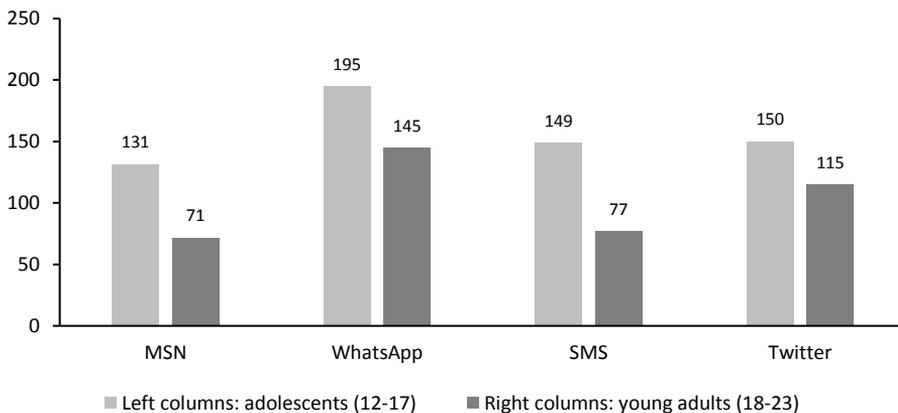
3.4 Lexis

3.4.1 Borrowings

Figure values: MSN 131/71, WhatsApp 195/145, SMS 149/77, Twitter 150/115. Left columns: adolescents (12-17) Right columns: young adults (18-23)

Figure 7: Normalised frequencies of borrowings.

Statistical testing reveals that the two-way interactions medium × borrowings and age group × borrowings were significant (partial χ^2 (3) = 173.01, p < .001, partial χ^2 (1) = 91.36, p < .001). Adolescents used more borrowed words, phrases, sentences, or textisms than young adults in the four media. Borrowings occurred most in WhatsApp, then on Twitter, and less frequently in SMS and MSN.
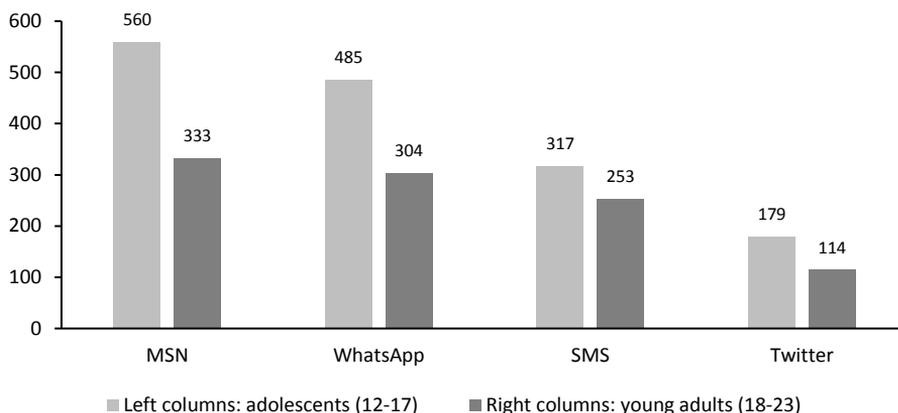
3.4.2 Interjections

■ Left columns: adolescents (12-17)   ■ Right columns: young adults (18-23)

Figure 8: Normalised frequencies of interjections.

Both two-way interactions medium × interjections and age group × interjections proved to be significant (partial χ^2 (3) = 1692.75, p < .001, partial χ^2 (1) = 457.63, p < .001). Interjections were used more by adolescents than young adults in all four media. They occurred most in MSN chats and least in tweets.

4 DISCUSSION

The results for the linguistic features that were analysed in this corpus study to-gether form the linguistic profiles of four new media and two age groups. These profiles ensue from the user characteristic age, and the various characteristics of the media examined.

4.1 Age

The results show that age plays a distinct role in the use of CMC language. This is consistent with findings by Hilte et al. (2016), who studied a corpus of Flemish computer-mediated messages and concluded that, in comparison to older youths (between 17 and 20 years old), adolescents (aged 13-16) more frequently used linguistic features of expressiveness deviating from the standard language. This was found, among other things, for reduplication of letters and punctuation, excessive use of capitalisation, emoticons, certain symbols (typographic kisses and hugs), and certain interjections (the onomatopoeic rendering of laughter) – each of these have been confirmed by the present study, with the exception of kisses, which in the present corpus were used more by young adults. Likewise, De Decker (2015), who also conducted a corpus study of Flemish CMC, observed that features such as 'flooding' (reduplication of letters), 'grapheme reductions' (phonetic abbreviations), and 'leetspeak' (incl. alphanumeric homophones and visual respellings) were used more by 13-to-16-year-olds than by 17-to-20-year-olds, as was the case in the present study. Adolescents were also found to diverge more from the standard language spelling in the Flemish written CMC studied by Peersman et al. (2016). The overall greater linguistic deviance of adolescents in CMC may be explained as follows. Teenagers, especially in puberty, are generally more non-conformist and innovative in their linguistic behaviour than adults (Eckert 1997, Androutsopoulos 2005). The most rebellious language behaviour is said to occur around the ages of 15-16, when youths feel the greatest pressure to rebel against the norms set by society, a period known as the adolescent peak (Holmes 1992). Young adults, on the other hand, feel a greater need to comply with the rules of the standard language, which has overt prestige in society. They start to feel social pressure not to appear immature, and so use Standard Dutch to conform to societal norms.

This explains why adolescents made significantly more use of textisms, typos, and symbols in all four media, and of misspellings in three media (all except MSN). In contrast, the young adults made a greater effort not to diverge from the standard language with regard to orthography and typography. The adolescents also used significantly more emoticons in MSN chats, whereas the young adults used more in SMS text messages. There appears to be no straightforward explanation for the lower frequency of emoticons in text messages by adolescents; it is possible that one or some of the contributors of text messages used very few emoticons, so an analysis of individual differences between the contributors could perhaps clarify this, especially given the rather low number of contributors of SMS text messages in the younger age group. Adolescents also diverged more from the standard language in terms of lexis: in all four media, they used relatively more borrowings, which are not (yet) part of Standard Dutch, and

interjections, which are characteristic of informal spoken language, but not for written standard language.

The results for the omissions were more complicated. The frequency of omissions was much higher with adolescents in MSN and WhatsApp, while it was higher with young adults in SMS and on Twitter. This is likely to be the result of a complex interaction between this linguistic feature with the variables age group and medium, as discussed below.

4.2 Medium

The medium used is found to have a large impact on CMC language use. In fact, it appears to have a greater effect than age group for all aspects except for misspellings (partial chi-squares of 17.84 vs. 30.02), for which age group had a greater impact. The partial chi-squares were higher for medium than age group for all other linguistic features – textisms (3574.71 vs. 1121.06), typos (676.18 vs. 305.62), emoticons (3711.52 vs. 12.14), symbols (1461.7 vs. 36.74), borrowings (173.01 vs. 91.36), and interjections (1692.75 vs. 457.63). This is in line with results reported by De Decker (2015) and Hilte et al. (2016), which show that medium was a significant determinant of the frequency of 'chatspeak' features and expressive markers in Flemish youths' CMC, even more so than age. Multiple medium characteristics play a part here (see Table 1), namely limitations in message size, (a)synchronicity, visibility, interactivity, and technology. These characteristics can either encourage or discourage deviations from the standard language.

The first characteristic concerns limitations in message size. SMS text messages and tweets are limited in number of characters, as opposed to MSN chats and WhatsApp messages. The message size limit in SMS (up to 160 characters) and on Twitter (a maximum of 140) requires considerable succinctness in communication. This explains the higher frequency of omissions in SMS text messages and tweets by young adults. Young adults apparently attempt to fill their text messages and tweets with as much information as possible without exceeding the message size limit, which they can achieve by means of omissions: leaving out nonessential elements, often function words. The lower frequency of omissions in adolescents' SMS text messages and tweets, in comparison with those sent by young adults, suggests that the latter more carefully formulate their utterances to be as concise as possible. The absence of a message size limit in MSN Messenger and WhatsApp provides young adults with the space needed to conform more to the norms of the (written) standard language with regard to syntactic completeness. This characteristic also partly explains the lower frequency of interjections in SMS and on Twitter, as the character limitations in

these media mean that nonessential words, such as interjections, are elided. The lack of such a limit in MSN chats and WhatsApp, by contrast, offers plenty of space for the use of interjections.

Another difference between the new media lies in synchronicity, i.e. the simultaneity of communication. Instant messaging is a (near-)synchronous medium: the communication takes place in practically real-time, which puts users under more pressure to respond quickly. The speed inherent in instant messaging is conducive to deviations from the standard language, because the high pace of communication provides little time for spelling or grammar checks. SMS and Twitter are asynchronous, so more time passes between the exchange of messages. These media offer time to edit messages and reflect upon one's words. This explains the high frequency of misspellings in MSN chats and WhatsApp messages, and of textisms in MSN chats, in comparison to the other media. It also helps to explain the high frequency of interjections in MSN and WhatsApp: the near-synchronous communication in instant messaging makes these written media resemble a spoken conversation, in which interjections are common (although, of course, the conditions for verbalisation and mutual awareness in written CMC are not the same as those in spoken language). The asynchronous communication in SMS and on Twitter endows these media with more of the characteristics of written language. Synchronicity is also related to omissions. The higher frequency of omissions in adolescents' MSN chats and WhatsApp messages, in comparison to their SMS text messages and tweets, is inconsistent with the aforementioned limit on message size in SMS and on Twitter. This finding can be attributed to the synchronicity of instant messaging, which causes users to communicate in ways similar to informal speech – with many sentence fragments and omitted words. Young adults use this synchronicity slightly less eagerly: in MSN and WhatsApp, they also imitate an informal conversation, but take somewhat more time than adolescents to write syntactically more complete sentences; they are not pressed for time, because there are no limits on the message size.

New media also differ in terms of visibility and interactivity, two characteristics that are strongly linked. Communication in MSN chats, SMS text messages, and WhatsApp messages is private and typically one-to-one (interaction between two people), and so visible for a small number of selected interlocutors, whereas communication on Twitter is usually public and one-to-many, so it can be read by a greater number of people. Tweets are often more aimed at informing a wider audience rather than sending personal messages. The public character of tweets discourages users to diverge from the standard language norms, in contrast with the privacy of the other three media. This explains the low frequency of textisms, misspellings, and typos in tweets. The high frequency of symbols in SMS, notably of hearts (*<3*) and kisses (esp. *X* and *x*) to conclude SMS text messages, reflects the personal character of this medium. In addition, this characteristic explains the high frequency of emoticons in

especially the MSN chats written by adolescents and SMS text messages by young adults. This results from the one-to-one (or sometimes some-to-some, in MSN) private communication taking place via these media, in which emoticons are regularly used to convey the writer's feelings and to avoid misunderstandings about the sentiment behind an utterance, as opposed to the generally one-to-many public communication of tweets, which require fewer emoticons because their content is often more neutral and less focused on emotions. The lowest frequency of emoticons in WhatsApp has a completely different cause: in this medium, the pragmatic functions of emoticons are also fulfilled by emoji.[10] Furthermore, the characteristic of visibility explains the high frequency of English borrowings in tweets in particular. The English language currently enjoys prestige among Dutch youths, and using English words is thus seen as 'hip' and 'cool' among this group. That is why they are frequently used in tweets, whose public nature allows a large audience to witness how 'cool' the writer is. Yet this does not explain the high frequency of borrowings in WhatsApp messages, which may, in fact, be caused by a temporal development: perhaps the use of English words has become even more popular between the times of collecting the SoNaR data and the WhatsApp data.

Finally, new media are used on different technological devices. MSN Messenger was a chat program for computers; text messages and WhatsApp messages are usually sent via mobile phones; while tweets are sent from either computers or mobile phones. These devices differ as to their keyboards and possibilities of using a predictive dictionary. The frequency of textisms in SMS text messages, tweets, and WhatsApp messages, and of misspellings in the former two media, may be lower because mobile phones, from which these messages are usually sent, often contain a predictive dictionary (which users can choose to utilize or not, to their own liking): when typing the first letter(s), the software 'guesses' the rest of the word. The words in the digital dictionaries that are used for this are spelt according to the standard language orthographic rules, which decreases the chance of textisms. However, such a predictive dictionary was not used with MSN chats. Moreover, the frequency of typos in SMS text messages may be higher than otherwise because of the small keypads on mobile phones, which increase the risk of typos.[11] A computer keyboard, as was used with MSN chats, has larger keys and thus presents a lower risk of typos. Typos also seem to be more affected by technology than synchronicity, seeing that the asynchronous communication of SMS does offer sufficient time for checking and correcting typos. Finally, the frequency of omissions in WhatsApp as compared to MSN – both

---

10 Emoji could not be coded in the present study due to the file format in which WhatsApp messages were contributed to the corpus.

11 Mobile phones can have an alphanumeric keyboard, with which three or four letters and a number are assigned to a single key, or a (possibly touchscreen) QWERTY keyboard, which is comparable to a computer keyboard, but much smaller. This is likely to affect the risk of typos, but unfortunately there was no information available about the devices with which the new media texts in the corpus were produced.

near-synchronous media which encourage omitting some elements to achieve a conversational writing style – can be explained by technological differences. The frequency of omissions is even higher in WhatsApp, because the small keyboards of mobile phones provide users with an extra incentive to omit parts of speech, whereas the large computer keyboards used for MSN did not.

5 CONCLUSION

It can be concluded from the results of this corpus study that, as expected, the language Dutch youths use when they communicate via social media indeed diverges from Standard Dutch on several writing dimensions, namely orthography, typography, syntax, and lexis. As for orthographic peculiarities, this CMC language is overall characterized by textisms (which include deviations in letters as well as in spacing, diacritics, punctuation, and capitalisation), misspellings, and typing errors. Typographic features are symbols and emoticons – as well as emoji in WhatsApp, but those concern visuals rather than typography. Regarding syntax, CMC language deviates from the written standard by its many omissions. Characteristic of the vocabulary of CMC language are borrowings, especially English ones, and interjections.

More importantly, this register analysis clearly shows the effects of medium and age group on the frequency with which certain linguistic features occur in computer-mediated messages. All interactions between medium (MSN, SMS, Twitter, and WhatsApp) and each of the linguistic features were highly statistically significant, due to an interplay of different medium characteristics. This was also the case for all interactions between age group (adolescents, young adults) and the linguistic features. Factors such as age and especially medium, whose impact was even greater, thus make sure that 'CMC language' is not a homogeneous language variant – rather, it encompasses various registers. The present study thus emphasizes the crucial importance of the variables age and medium for online language use, as attested in (Dutch) written computer-mediated communication, and once more confirms that youths' online writings offer a wealth of linguistic diversity.

6 LIMITATIONS AND SUGGESTIONS FOR FURTHER RESEARCH

A drawback of this study is that the collection periods for different parts of the corpus were not the same. The SoNaR texts were collected between 2009 and 2011, thus quite some years ago. The WhatsApp messages are more recent,

collected in 2015. It is not inconceivable that Dutch youths' CMC language has changed somewhat between these collection periods; after all, language is subject to change, and this is particularly true for youth languages, which are dynamic and constantly evolving. This means that some of the differences found between the WhatsApp data and the data from the other three media could possibly be attributed not just to the characteristics of the various media, but also (partly) to temporal developments. Analysis of more recent data would, therefore, be a welcome addition to the current study.

It would also be interesting to expand the analysis in terms of age groups, with the addition of digital texts written by children (for instance, aged 6-11 years). Yet due to practical and ethical considerations, collecting such private texts from young children could pose a real challenge. Besides expanding the corpus in age, it could also be enlarged in terms of medium. The study reported here has examined four well-known new media, while of course there are many more, and those that are popular among young people change very rapidly. Future research could thus analyse other media. It would be valuable to complement this register analysis with, for example, Facebook posts. In fact, these were already collected by the author between December 2015 and May 2016, so such an analysis would be a viable option for a future study.

Online language variability among new media could also be studied more in depth by including even more media characteristics into the research design, e.g. focusing on the software used to compose the messages, such as whether or not it includes predictive dictionaries, autocorrection, or spelling checkers. Additional user or situational characteristics, such as (the users' relationship with / profile of) the conversational partner and the communicative purpose of the interaction, would also be exciting ways to expand the analysis.

As a concluding suggestion, one more possibility for future corpus-lingusitics studies into CMC would be to include an extra independent variable, besides age and medium, with an obvious choice being gender. Other research suggests that there are differences between girls and boys in the use of several linguistic features of digital writing (e.g. Wolf 2000, Baron 2004, Parkins 2012, Hilte et al. 2016). This could be further explored for Dutch computer-mediated messages, to gain an even more nuanced picture of the registers that exist within CMC language.

Given that the language with which Dutch youths communicate via social media clearly diverges from Standard Dutch, chances are that this informal CMC language interferes with their more formal 'school language.' However, prior research does not provide a conclusive answer as to whether this is indeed the case. Therefore, this open issue will be investigated in future studies of the author's ongoing (doctoral) research project into the impact of CMC on literacy. As such, the present

corpus study is only a first step in studying Dutch youngsters' written CMC. The next steps will dig deeper into the possible relation between Dutch youths' social media use and their writing skills. This will be examined in both a correlational study and an experimental study. The former to see if any evidence for a relationship can be found, the latter to explore the causality of this relationship (if it exists at all), and thus whether it is indeed CMC that affects literacy, and not vice versa. In this extended outlook, let me briefly outline the design of these two studies.

Youths who will participate in the correlational study will be tested at school, so in an educational setting. They will first write an essay – with the text genre of expository discussion – to measure their formal writing skills. Subsequently, they will fill in questionnaires about their social media use. The essays will be analysed for several measures of writing quality, namely lexical richness, syntactic complexity, formality, and writing productivity. It will then be examined whether participants' CMC use (in terms of frequency, variety, intensity, use of textisms, etc.), as self-reported in the surveys, correlates with the writing quality of their essays. This work will thus study whether participants' private online writing habits are related to the quality of the 'offline' texts they write at school.

The experimental study will use social media as the experimental prime. All school classes that participate will be divided into two groups: an experimental group, who will communicate via WhatsApp together during the priming phase, and a control group, who will spend that time on a control task, namely colouring mandalas. All participants will then write stories – with the genre of narrative storytelling – to test their productive writing skills, which will again be analysed for several measures of writing quality. Next, they will complete a grammaticality judgement task (GJT), to test their receptive grammar and spelling skills: they will be presented with sentences in which they have to spot and correct 'language errors,' i.e. deviations from Standard Dutch. It will then be measured whether the immediately preceding use of WhatsApp has a direct impact on the writing quality of the experimental groups' stories or on their performance on the GJTs.

Both studies will involve youths from different educational levels and age groups, to find out if these are mediating factors in the potential impact of Dutch youths' informal written CMC on their more formal writing skills. We hypothesize that writers of a younger age group or lower educational level could experience a greater extent of interference of social media on their school writings. Irrespective of what these future studies will find, it is nevertheless important to point out to all youngsters, no matter their age or education, that the informal digital language they use in computer-mediated messages and the standard language are different variants (registers) of Dutch – variants they ought to keep separate and employ effectively depending on the context.

Acknowledgments

References

Androutsopoulos, Jannis, 2005. Research on youth language. Ammon, Ulrich, Norbert Dittmar, Klaus J. Mattheier and Peter Trudgill (eds.): *Sociolinguistics: An International Handbook of the Science of Language and Society* 2. Berlin: Mouton de Gruyter. 1496–1505.

Androutsopoulos, Jannis, 2011. Language change and digital media: A review of conceptions and evidence. Kristiansen, Tore and Nikolas Coupland (eds.): *Standard Languages and Language Standards in a Changing Europe*. Oslo: Novus Press. 145–161.

Baron, Naomi, 2004. See you online: Gender issues in college student use of instant messaging. *Journal of Language and Social Psychology* 23/4. 397–423.

Bergs, Alexander, 2009. Just the same old story? The linguistics of text messaging and its cultural repercussions. Rowe, Charley and Eva L. Wyss (eds.): *Language and New Media: Linguistic, Cultural, and Technological Evolutions*. Cresskill, NJ: Hampton Press. 55–73.

Bogle, Vellah and Inger Hollebeek, 2013. *Taalvoutjes: het boek*. Utrecht: Van Dale.

Broeren, Karin, 2012. De 10 irritantste taalfouten op social media. *Ze.nl.* http://www.ze.nl/p/141435/de_10_irritantste_taalfouten_op_social_media. (Last accessed 29 June 2017.)

Carrington, Victoria, 2004. Texts and literacies of the Shi Jinrui. *British Journal of Sociology in Education* 25/2. 215–228.

Crystal, David, 2006. *Language and the Internet* (2nd ed.). Cambridge: Cambridge University Press.

Crystal, David, 2008. *Txtng: The Gr8 Db8*. Oxford: Oxford University Press.

Daniëls, Wim, 2009. *Sms & msn: hoest begonnuh?, hoe sgrijf jut?, ist errug?* Houten: Prisma.

Darics, Erika, 2013. Non-verbal signalling in digital discourse: The case of letter repetition. *Discourse, Context & Media* 2/3. 141–148.

De Decker, Benny and Reinhild Vandekerckhove, 2012. English in Flemish adolescents' computer-mediated discourse: A corpus-based study. *English World-Wide* 33/3. 321–352.

De Decker, Benny, 2015. Prototypische chatspeakkenmerken in Vlaamse tienerchattaal: de invloed van gender, leeftijd en medium. *Taal en Tongval* 67/1. 1–41.

Drouin, Michelle and Claire Davis, 2009. R u txting? Is the use of text speak hurting your literacy? *Journal of Literacy Research* 41/1. 46–67.

Eckert, Penelope, 1997. Age as a sociolinguistic variable. Coulmas, Florian (ed.): *The Handbook of Sociolinguistics*. Oxford: Blackwell. 151–167.

EricMM, 2015. Taalfouten: geen verloedering maar verandering. *Joop*. http://www.joop.nl/opinies/taalfouten-geen-verloedering-maar-verandering. (Last accessed 29 June 2017.)

Ferrara, Kathleen, Hans Brunner and Greg Whittemore, 1991. Interactive written discourse as an emergent register. *Written Communication* 8/1. 8–34.

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

Hård af Segerstad, Ylva, 2002. *Use and Adaptation of Written Language to the Conditions of Computer-Mediated Communication*. Dissertation University of Gothenburg.

Herring, Susan, 2001. Computer-mediated discourse. Schiffrin, Deborah, Deborah Tannen and Heidi E. Hamilton (eds.): *Handbook of Discourse Analysis*. Oxford: Blackwell. 612–634.

Herring, Susan, 2012. Grammar and electronic communication. Chapelle, Carol A. (ed.): *Encyclopedia of Applied Linguistics*. Oxford: Wiley-Blackwell. 1–11.

Hilte, Lisa, Reinhild Vandekerckhove and Walter Daelemans, 2016. Expressiveness in Flemish online teenage talk: A corpus-based analysis of social and medium-related linguistic variation. Fišer, Darja and Michael Beißwenger (eds.): *Proceedings of the 4th Conference on CMC and Social Media Corpora for the Humanities*. Academic Publishing Division of the Faculty of Arts of the University of Ljubljana. 30–33.

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

Jacobs, Gloria, 2008. People, purposes, and practices: Insights from cross-disciplinary research into instant messaging. Coiro, Julie, Michele Knobel, Colin Lankshear and Donald J. Leu (eds.): *Handbook of Research on New Literacies*. New York, NY: Routledge. 469–490.

Labov, William, 1966. *The Social Stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

Oostdijk, Nelleke, Martin Reynaert, Véronique Hoste and Ineke Schuurman, 2013. The construction of a 500-million-word reference corpus of contemporary written Dutch. Spyns, Peter and Jan Odijk (eds.): *Essential Speech and Language Technology for Dutch: Results by the STEVIN Programme*. Heidelberg: Springer. 219–247.

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

Peersman, Claudia, Walter Daelemans, Reinhild Vandekerckhove, Bram Vande-kerckhove and Leona Van Vaerenbergh, 2016. The effects of age, gender and region on non-standard linguistic variation in online social networks. http://arxiv.org/abs/1601.02431. (Last accessed 29 June 2017.)

Plester, Beverly, Clare Wood and Puja Joshi, 2009. Exploring the relationship be-tween children's knowledge of text message abbreviations and school literacy outcomes. *British Journal of Developmental Psychology* 27/1. 145–161.

Proudfoot, Candice, 2011. *An Analysis of the Relationship between Writing Skills and 'Short Messaging Service' Language: A Self-Regulatory Perspective*. Disserta-tion Potchefstroom Campus, North-West University.

Robin F, 2014, August 8. De jongste taalfouten. *Onze Taal*. https://onzetaal.nl/nieuws-en-dossiers/weblog/de-jongste-taalfouten. (Last accessed 29 June 2017.)

Shaw, Philip, 2008. Spelling, accent and identity in computer-mediated com-munication. *English Today* 24/2. 42–49.

Silva, Cláudia, 2011. Writing in Portuguese chats :). A new wrtng systm? *Written Language & Literacy* 14/1. 143–156.

Stroop, Jan, 2010. *Hun hebben de taal verkwanseld: over Poldernederlands, 'fout' Nederlands en ABN*. Amsterdam: Athenaeum-Polak & Van Gennep.

TN, 2014, June 30. De jongste taalfouten. *Onze Taal*. https://onzetaal.nl/nieu-ws-en-dossiers/weblog/de-jongste-taalfouten. (Last accessed 29 June 2017.)

Truijens, Aleid, 2009, July 21. Straattaal: Algemeen Cool Nederlands. *De Volk-skrant*. http://www.volkskrant.nl/binnenland/straattaal-algemeen-cool-nederlands~a339199. (Last accessed 29 June 2017.)

Verheijen, Lieke, 2013. The effects of text messaging and instant messaging on literacy. *English Studies* 94/5. 582–602.

Verheijen, Lieke, 2016. De macht van nieuwe media: hoe Nederlandse jongeren com-municeren in sms'jes, chats en tweets. Van de Mieroop, Dorien, Lieven Buysse, Roel Coesemans and Paul Gillaerts (eds.): *De macht van de taal: Taalbeheersing-sonderzoek in Nederland en Vlaanderen*. Leuven / Den Haag: Acco. 275–293.

Verheijen, Lieke and Wessel Stoop, 2016. Collecting Facebook posts and What-sApp chats: Corpus compilation of private social media messages. Sojka, Petr, Aleš Horák, Ivan Kopeček and Karel Pala (eds.): *Text, Speech and Dialogue: 19th International Conference, TSD 2016, LNAI 9924*. Cham: Springer. 249–258.

Werry, Christopher, 1996. Linguistic and interactional features of Internet Relay Chat. Herring, Susan (ed.): *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives*. Amsterdam: Benjamins. 47–63.

Winzker, Kristy, Frenette Southwood and Kate Huddlestone, 2009. Investigating the impact of SMS speak on the written work of English first language and English second language high school learners. *Per Linguam* 25/2. 1–16.

Wolf, Alecia, 2000. Emotional expression online: Gender differences in emoticon use. *Cyberpsychology & Behavior* 3/5. 827–833.

Wood, Clare, Nenagh Kemp and Beverly Plester, 2013. *Text Messaging and Literacy: The Evidence*. London: Routledge.

W00t00w, 2015, July 20. 'Bijna driekwart Nederlandse tieners heeft Instagram en Snapchat op telefoon'. *Tweakers*. https://tweakers.net/nieuws/104306/bijna-driekwart-nederlandse-tieners-heeft-instagram-en-snapchat-op-telefoon.html. (Last accessed 29 June 2017.)