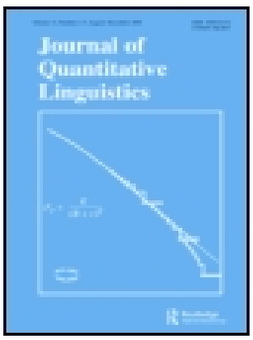

Journal of Quantitative Linguistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/njql20

Linguistic Accommodation in Teenagers' Social Media Writing: Convergence Patterns in Mixed-gender Conversations

Lisa Hilte, Reinhild Vandekerckhove & Walter Daelemans

To cite this article: Lisa Hilte, Reinhild Vandekerckhove & Walter Daelemans (2020): Linguistic Accommodation in Teenagers' Social Media Writing: Convergence Patterns in Mixed-gender Conversations, Journal of Quantitative Linguistics, DOI: 10.1080/09296174.2020.1807853

To link to this article: https://doi.org/10.1080/09296174.2020.1807853

Published online: 06 Sep 2020.

JOURNAL OF QUANTITATIVE LINGUISTICS
https://doi.org/10.1080/09296174.2020.1807853

Routledge
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Linguistic Accommodation in Teenagers' Social Media Writing: Convergence Patterns in Mixed-gender Conversations

Lisa Hilte, Reinhild Vandekerckhove and Walter Daelemans

CLiPS Research Center, University of Antwerp, Antwerp, Belgium

ABSTRACT

The present study analyzes the phenomenon of linguistic accommodation, i.e. the adaptation of one's language use to that of one's conversation partner. In a large corpus of private social media messages, we compare Flemish teenagers' writing in two conversational settings: same-gender (including only boys or only girls) and mixed-gender conversations (including at least one girl and one boy). We examine whether boys adopt a more 'female' and girls a more 'male' writing style in mixed-gender talks, i.e. whether teenagers converge towards their conversation partner with respect to gendered writing. The analyses focus on two sets of prototypical markers of informal online writing, for which a clear gender divide has been attested in previous research: expressive typographic markers (e.g., emoticons), which can be considered more 'female' features, and 'oral', speech-like markers (e.g., regional language features), which are generally more popular among boys. Using generalized linear-mixed models, we examine the frequency of these features in boys' and girls' writing in same- versus mixed-gender conversations.

Patterns of convergence emerge from the data: they reveal that girls and boys adopt a more similar style in mixed-gender talks. Strikingly, the convergence is asymmetrical and only significant for a particular group of online language features.

1. Introduction

A range of studies have reported on the correlation between adolescents' socio-demographic profiles and their online writing style, demonstrating how youths with distinct profiles (e.g. in terms of age or gender) tend to favour certain markers of online writing to different extents (De Decker & Vandekerckhove, 2017; Hilte et al., in press, 2018a, 2018b, 2018c; Varnhagen et al., 2010; Verheijen, 2015). It has, for instance, been demonstrated that teenage girls generally show a greater preference for expressive typographic markers such as emoji than boys, whereas teenage boys score higher than girls for the use of colloquial speech markers and regional

CONTACT Lisa Hilte lisa.hilte@uantwerpen.be

markers in informal online writing (Hilte et al., in press). A factor that remains under-researched in this domain, however, is the interlocutor's impact. While the phenomenon of accommodation, i.e. the adaptation of one's communicative behaviour to that of one's conversation partner, is widely investigated for 'traditional' face-to-face communication (see Section 2), it is seldom the object of research on online/digital communication. The present paper aims to fill that gap: In a large corpus of instant messages, we will investigate whether Flemish (i.e. living in Flanders, Dutch-speaking Belgium) teenagers adapt their writing style depending on their conversation partner's gender. More precisely, we wish to examine whether patterns of *gender convergence* emerge, with girls sounding less prototypically 'female' and boys less prototypically 'male' when interacting with someone of the opposite sex.

The paper is structured as follows. First, Section 2 gives an overview of related research. Next, Section 3 describes the materials and methods of the present study. Finally, the results of the analyses are presented and discussed in Sections 4, 5 and 6, respectively.

2. Related Research

The adaptation of one's communicative behaviour to one's conversation partner has been studied from different perspectives (e.g. sociolinguistic versus socio-psychological) and has been labelled differently, depending on the scientific field – e.g. *accommodation*, the term we will be using, but also *matching, alignment, mimicry* and *synchrony* (for an overview, see Burgoon et al., 2017). Our main point of reference will be the sociolinguistic framework 'Communication Accommodation Theory' (CAT), developed by Howard Giles in the 1970s and since then refined and elaborated multiple times (see Dragojevic et al., 2015 for an overview). Within CAT, accommodation is considered to be driven by a desire to facilitate interaction and to regulate social distance with respect to one's interlocutor (Dragojevic et al., 2015, p. 10), with common strategies consisting in adjusting one's communicative behaviour to appear more similar (convergence) or dissimilar (divergence) to others (Giles & Ogay, 2007, pp. 295–296). Although divergence is generally said to be evaluated more negatively and convergence more positively, full convergence is seldom desired; rather, optimal levels of (dis)similarity seem to exist, which are both individually and socio-culturally determined (Burgoon et al., 2017; Dragojevic et al., 2015, pp. 13, 15). The idea that 'speakers accommodate their style to their audience' (Bell, 1984, p. 162) is also key in Bell's related framework of Audience Design that aims to explain style variation. An additional distinction that is made in Audience Design compared to CAT but that falls out of the scope of the present research, concerns the hierarchy of audience roles (ranging from addressee

to eavesdropper), which is said to modify different audience members' impact on the speaker's style (Bell, 1984, pp. 159–160).

While the *propensity of alignment* – i.e. the inherent personal inclination to adapt one's communication to others – may differ from one person to another (Jones et al., 2014; Xu & Reitter, 2015), several patterns of accommodation appear to be quite robust. These patterns concern communicative adaptations related to aspects of interlocutors' socio-demographic and psychological profiles, such as social power/status and personality type (Dragojevic et al., 2015; Muir et al., 2016). With respect to gender, asymmetrical convergence appears to be common, with women adapting their language use more strongly to men than vice versa (Palomares et al., 2016, p. 133 and references therein). However, other studies reveal patterns of *speech complementarity*, i.e. mutual divergence by men and women in order to emphasize social roles (Burgoon et al., 2017; Dragojevic et al., 2015, p. 15). Finally, in the context of computer-mediated communication or CMC (see also below), yet another pattern has been attested: with respect to emoticon use, men appear to converge much more strongly to an 'expressive female standard' in mixed-gender online conversations than vice versa (Wolf, 2000).

Accommodation is most often analysed in dyadic (i.e. one-on-one), spoken, face-to-face interactions, but some studies have examined to which extent findings translate from such 'traditional' settings to the context of online communication (e.g. Doyle et al., 2016; Riordan et al., 2013; Scissors et al., 2009, 2008). However, as of yet, there do not exist any large-scale studies on accommodation in CMC corpora that truly mirror spontaneous face-to-face interactions, as existing studies are either carried out on small corpora (e.g. De Siqueira & Herring, 2009; Wolf, 2000), on public, asynchronous conversations (e.g. Danescu-Niculescu-Mizil et al., 2011; Dino et al., 2009; Doyle et al., 2016), or on synchronous chat conversations between strangers and/or in lab-based settings (e.g. Gonzales et al., 2010; Muir et al., 2017; Niederhoffer & Pennebaker, 2002; Scissors et al., 2009, 2008). In addition, most previous studies concern adult participants only, although the desire to obtain social approval may be much stronger among teenagers, since adolescent peer group behaviour tends to be determined by a 'need of acceptance' and 'fear of rejection' (Taylor, 2001, p. 298). The present study aims to fill these gaps with respect to gender accommodation.

The research design (see Section 3) allows to broaden the scope of accommodation research by incorporating multiple factors that may influence gender accommodation and – crucially – also their potential interaction. In a large corpus of Flemish adolescents' informal online interactions, we investigate which patterns of accommodation emerge with respect to 'gendered' writing: i.e. do teenage boys and girls adapt their online writing style to each other in mixed-gender conversations (to a similar extent)? The linguistic variables concern two sets of prototypical CMC markers for which

a clear gender divide has been attested (Hilte et al., in press): (predominantly typographic) 'expressive' markers, such as emoticons, and 'oral' or speech-like markers, such as regional language features (see below). Many previous studies only include one or two linguistic variables, e.g. word or phrase repetition (Scissors et al., 2008) or sequence length and duration (Riordan et al., 2013). However, the inclusion of a wide range of linguistic features may be crucial since convergence and divergence can occur simultaneously for distinct markers (Gasiorek, 2016, p. 27) and since people's (e.g. negative or positive) perception of accommodation may depend on the actual feature (Scissors et al., 2009).

3. Materials and Methods

Below, we present the dataset (Section 3.1) and the linguistic variables (Section 3.2). Next, we describe the data preprocessing and the statistical analyses (Sections 3.3 and 3.4, respectively).

We note that the present study focuses on accommodation from a synchronic perspective, comparing teenagers' writing in different conversational settings (see below). We leave the diachronic perspective, including a temporal dimension – i.e. the analysis of *the course of* particular conversations – for future work.

3.1. Corpus

The dataset contains 433,731 social media posts (>2.5 million tokens) produced by 1384 teenagers living in Flanders, i.e. northern Dutch-speaking Belgium. The teenagers, aged 13–20, are all students in secondary education. The dataset contains their spontaneous, private instant messages, produced in Dutch on Facebook Messenger and WhatsApp, mainly between 2015 and 2016 (88% of the posts). Region or dialect area is no variable, as the vast majority (96%) of the teenagers live in the central Flemish province of Antwerp.

The data collection was carried out in collaboration with several secondary schools in the province of Antwerp. After informing the students about our research project, we invited them to voluntarily donate chat conversations that were produced before our school visits, in order to exclude observer's paradox. The students also provided the relevant metadata (see below). We asked the participating pupils' (and for minors also their parents') permission to store and linguistically analyse their anonymized utterances.

We allowed two submission formats, depending on the platform on which the conversations were produced. For WhatsApp, the students could easily forward entire conversations as plain text files via the app's export setting.

For Messenger, the students were instructed to copy their conversations from the Facebook website and paste them to a submission website we created. These pasted texts were automatically converted to a plain text format[1] too. Finally, we wrote Python scripts to automatically extract the chat conversations from the submission files and add them in a structured way to one final corpus file. In this file, each line contains one social media post along with an anonymized author identifier (instead of the participant's actual name), socio-demographic information on the author (see below), and additional conversational information (e.g. the platform on which the interaction took place). Apart from the participants' names, personal (contact) information that was mentioned in the chat utterances was removed and replaced by an anonymized placeholder too. We automatically detected persons' names and names of towns using predefined lists (e.g. name lists published by the Belgian government, and a list of all Flemish towns). Phone numbers, email addresses, urls and street names – which all tend to have a fixed and recognizable format – were detected automatically through pattern recognition (i.e. regular expressions). For more detailed information on the data collection and processing, see Hilte (2019).

The participants provided metadata on their social profile: therefore, for all participants, information on age, gender and educational track is available. Furthermore, all conversations in the dataset were categorized with respect to the interlocutors' gender as either same-gender (including only boys or only girls) or mixed-gender (including at least one boy and one girl). Finally, the number of interlocutors was added as a binary variable too: we distinguish between one-on-one conversations (two interlocutors) and group chats (more than two interlocutors). Table 1 shows the relevant distributions in the corpus with respect to gender accommodation, i.e. in terms of author gender and conversational setting. Table 2 presents the potential confounding factors: the teenagers' age and educational track, and the number of interlocutors in a conversation.

We note that gender is operationalized as the distinction between boys and girls, since a non-binary approach (e.g. operationalizing gender as a continuum) was infeasible with the profile information we had access to.

For age, we distinguish between two groups of high school students too: younger teenagers (13–16 years old) and older teenagers or young adults (17–20 years old). Age is treated as a categorical rather than

Table 1. Gender distributions in the corpus (in terms of tokens).

		Gender		
		Girls	Boys	Total
Conversational setting	**Same-gender**	1,188,236	454,586	1,642,822 (65%)
	Mixed-gender[2]	444,065	440,791	884,856 (35%)
	Total	1,632,301 (65%)	895,377 (35%)	2,527,678

Table 2. Distributions in the corpus (in terms of tokens) w.r.t. confounding factors.

Variable	Variable levels	Tokens
Educational track	General secondary education	739,795 (29%)
	Technical secondary education	1,150,600 (46%)
	Vocational secondary education	637,283 (25%)
Age	Younger teenagers (13–16)	1,360,159 (54%)
	Older teenagers/young adults (17–20)	1,167,519 (46%)
Number of interlocutors	One-on-one (2 interlocutors)	1,762,390 (70%)
	Group chat (>2 interlocutors)	765,288 (30%)
Total		2,527,678

a continuous variable, as previous sociolinguistic studies suggest that teenagers' non-standard language use does not evolve linearly as they age, but 'peaks' during mid-puberty: it increases until the age of 15–16, and then decreases again. This phenomenon is referred to as the 'adolescent peak' (e.g. Holmes, 1992, p. 184).

All participants attend one of the three main types of Belgian secondary education. These range from the theory-oriented general secondary education, where students are prepared for higher education, to the practice-oriented vocational secondary education, where students are taught a specific, often manual, profession. The technical secondary education holds an intermediate position on this continuum (Flemish Ministry of Education and Training & Flemish Ministry of Education and Training [FMET], 2017, p. 10).

3.2. Linguistic Variables

The research design includes two sets of prototypical markers of informal online writing. Many 'chatspeak' markers can be linked to one of three 'maxims' or implicit rules of linguistic conduct of informal online interaction: the principles of expressive compensation, orality and brevity (as described by e.g. Androutsopoulos, 2011, p. 149). In the present study, the principle of brevity, which consists in maximizing typing speed/comfort, was left out as features relating to this principle (e.g. acronyms/abbreviations) are often of a highly functional nature and appear not to be subject to social variation to the same extent as the other types of features (De Decker & Vandekerckhove, 2017, p. 278; Hilte et al., 2018c). Below, we describe the two maxims that are included in the research design along with their related linguistic features, which are illustrated with examples from the corpus. The selection of these particular features was based on related research (e.g. Varnhagen et al., 2010; Verheijen, 2015).

The orality maxim concerns the fact that in many forms of CMC, the register is to a large extent 'conceptually oral', reflecting oral communication and typical speech patterns rather than classical written communication. In our dataset of Flemish teenagers' Dutch instant messages, this maxim 'results' in the insertion of different kinds of non-standard Dutch lexemes:

- dialect/regiolect words:
  e.g., *tot seffes* (std. Dutch: *tot straks*, 'see you later')

- colloquial words or 'slang':
  e.g., *negeer die gast* (std. Dutch: *negeer die jongen*, 'ignore that dude')

- orthographic renderings of non-standard pronunciation or morphology:
  e.g., *da weet ik ni* (std. Dutch: *dat weet ik niet*, 'I don't know that')
  e.g., *komde gij ook?* (std. Dutch: *kom jij ook?*, 'Are you coming too?')

Furthermore, Flemish teenagers often incorporate English words or phrases that are part of Dutch adolescent speech. These insertions include:

- English words rendered in their 'original' form:
  e.g., *ben home alone nu* ('am home alone now')

- English words adapted to Dutch (in terms of e.g., spelling or morphology):
  e.g., *naais* ('nice')
  e.g., *haar cadeaus sucken* ('her gifts suck')

We note that the base language in the dataset is always Dutch, as entire chat conversations in another language were excluded. Furthermore, English loan words that have been integrated in Dutch for a considerable time to the extent that they are generally considered part of standard Dutch vocabulary and included in Dutch dictionaries (e.g. *computer*), are not counted as English lexemes in the analysis, but are labelled as Dutch.

The maxim of expressive compensation relates to the application of a wide range of (predominantly typographic) strategies to compensate for the absence of certain expressive cues in written communication, such as intonation, volume or facial expressions. The following features are included in the research design:

- emoticons and emoji:
  e.g. 😂😂😂😂😂

- words or phrases rendered in capital letters ('allcaps'):
  e.g., *WIL ALLES WETEN!!* ('Want to know everything!!')

- deliberate repetition of letters and punctuation marks ('flooding'):
  e.g., *jaaaaaaaaaaaaa!!!!* ('yeeeeessss!!!!')

- combinations of question and exclamation marks:
  e.g., *echt?!* ('really?!')

- onomatopoeic rendering of laughter:
  e.g., *Hahahahah*

- typographic rendering of kisses and/or hugs through combinations of the letters 'x' and 'xo':
  e.g., *doei xxxxxx* ('bye xxx')
  e.g., *dankjewel xoxo* ('thank you xoxo')

In previous work, we repeatedly attested a clear gender divide with respect to these two sets of chatspeak features: while expressive markers have a higher frequency in girls' writing than in boys' writing, for the oral markers we attest the opposite pattern (Hilte et al., in press). However, hitherto, the profile of the interlocutor was not included as a variable. Consequently, it seems worth investigating whether the gender patterns are affected by the interlocutors' gender, and whether systematic linguistic adaptation occurs in mixed-gender contexts.

3.3. Data Preprocessing

For the quantitative analyses, a participant-level 'summary' of the dataset was created, with each line containing information about one participant belonging to either the younger or older adolescents, in one of the potential conversational settings (in terms of interlocutors' gender and number of conversation partners). Consequently, every participant can occur in the dataset on multiple lines (e.g. at different age points, or in different conversational setups). These 'repeated observations' are corrected for in our statistical models through the addition of a random effect for participant (see below).

Each line in the dataset contains the participant's meta-information (a unique, anonymized, identifier as well as gender, age category and educational track), conversational meta-information (mixed/same-gender and one-on-one/group chat) along with the total number of tokens as well as the number of oral and expressive features produced in this particular context by this participant.

The feature occurrences were detected automatically in the dataset with Python scripts. For the oral markers, a dictionary-based approach was used (relying on existing standard Dutch and standard English word lists as well as

on handcrafted lists of colloquial Flemish Dutch lexemes). The expressive markers were detected through pattern recognition (i.e. with regular expressions). The scripts' output was compared to a human annotator's performance for a test set of 200 randomly selected posts (1,257 tokens). The software reached a satisfying average precision and recall of 91% resp. 88% for all features combined. While the precision score indicates the share of detected occurrences of a feature that are indeed valid occurrences of that feature, the recall score presents the share of all occurrences in the (test) corpus of a feature that are detected as such by the software. The evaluation metrics were also sufficiently high for each individual feature, which indicates that the software is reliable and its output suitable for further linguistic analysis. For an extensive discussion and error analysis of the feature extraction procedure, see Hilte et al. (in press).

3.4. Model Fitting

We modelled the teenagers' degree of 'expressive' and 'oral' online writing with generalized linear-mixed models (GLMMs) with a Poisson distribution, as implemented in the 'lme4′ package for R (Bates et al., 2017). These models are typically recommended for the analysis of count data (Harrison, 2014, p. 2; Ismail & Jemain, 2007, p. 105), as the Poisson distribution is considered the 'simplest distribution for modeling count data' (Zeileis et al., 2008, p. 5).

GLMMs enable the simultaneous analysis of different predictors or fixed effects as well as of their potential interactions with each other. In the present study, we are specifically interested in the impact of author gender (male/female) and conversational setting (same-gender/mixed-gender) on the response variable, i.e. the counts for expressive or for oral markers. We will also take potential confounding factors into account: the authors' age and educational track, and the number of interlocutors in a chat conversation.

As mentioned above, the models can take into account the impact of individual chatters and correct for repeated observations as a random effect for subject was added. This way, the models can 'link' observations from one and the same participant in distinct conversational contexts to each other, and deal with individual 'writing styles' (for instance: certain people may simply *always* write in a more expressive way than others). The inclusion of a random slope for subject*conversational setting would allow us to take into account differences among the teenagers with respect to their 'propensity of alignment' (Jones et al., 2014; Xu & Reitter, 2015), or, in other words, to correct for the fact that some people may just be more likely to adapt their language to their interlocutor. However, we did not have sufficient data points to include this random effect. Finally, the models can handle differences in sample size between participants by adding an 'offset' for the logarithm of the number of tokens per chatter.

In order to deal with troubles of overdispersion (i.e. the variance of the response variable exceeding the mean – see Hilte et al., in press), which, if left unaddressed may lead to unreliable outcomes (Harrison, 2014, pp. 1, 2, 17–18 and references therein; Ismail & Jemain, 2007, p. 103), we added an observation-level random effect (OLRE), i.e. a random effect for each observation in the data, as recommended by Harrison (2014, p. 1).

In the results section below, we always discuss the best possible model (including the best subset of predictors or fixed effects), i.e. the model that resulted in the best fit for the data. We experimentally determined this best fit through a backwards stepwise procedure, i.e. starting from a complex model, and step-by-step deleting irrelevant factors.

4. Results

Our previous research demonstrated how Flemish teenagers' age, gender and educational track significantly influence expressive and oral/speech-like online writing (Hilte et al., in press). The present study aims to complement these findings with potential patterns of gender accommodation, operationalized as the interaction between author gender and conversational setting. In addition, the effect of the number of interlocutors in a conversation will be examined.

Sections 4.1 and 4.2 present the best models for oral and expressive writing, respectively. Plots will only be provided for patterns that directly relate to the phenomenon of gender accommodation. Other (confounding) tendencies are described in the text only.

4.1. Orality

The best model for orality takes all predictors into account except number of interlocutors, and includes multiple interactions. Tables 3 and 4 present the fixed effects and the Anova, respectively.

Table 3. Orality: fixed effects (Reference group: younger boys, general education, same-gender one-on-one chats).

	Estimate	Std. Error	z value	Pr(>|z|)	Signif.
(Intercept)	−1.732333	0.026261	−65.97	< 2e-16	***
GenderFemale	−0.131950	0.029288	−4.51	6.63e-06	***
AgeOlder	−0.001621	0.026875	−0.06	0.951909	
SettingMixed-gender	−0.014147	0.026926	−0.53	0.599293	
EducationTechnical	0.102364	0.026921	3.80	0.000143	***
EducationVocational	0.133557	0.027763	4.81	1.51e-06	***
GenderFemale:AgeOlder	−0.103818	0.033058	−3.14	0.001687	**
AgeOlder:SettingMixed-gender	−0.059212	0.029133	−2.03	0.042108	*
GenderFemale:SettingMixed-gender	0.071584	0.030381	2.36	0.018462	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Table 4. Orality: Anova.

	Chisq	Df	Pr(>Chisq)	Signif.
Gender	45.0600	1	1.911e-11	***
Age	22.6676	1	1.926e-06	***
Setting	0.0571	1	0.811143	
Education	27.1262	2	1.287e-06	***
Gender:Age	9.8628	1	0.001687	**
Gender:Setting	5.5518	1	0.018462	*
Age:Setting	4.1309	1	0.042108	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

A significant interaction can be observed between author gender (female/male) and conversational setting (same-gender/mixed-gender). This interaction is visualized in Figure 1. First of all, the figure clearly confirms that oral markers are a primarily 'male' feature (Hilte et al., in

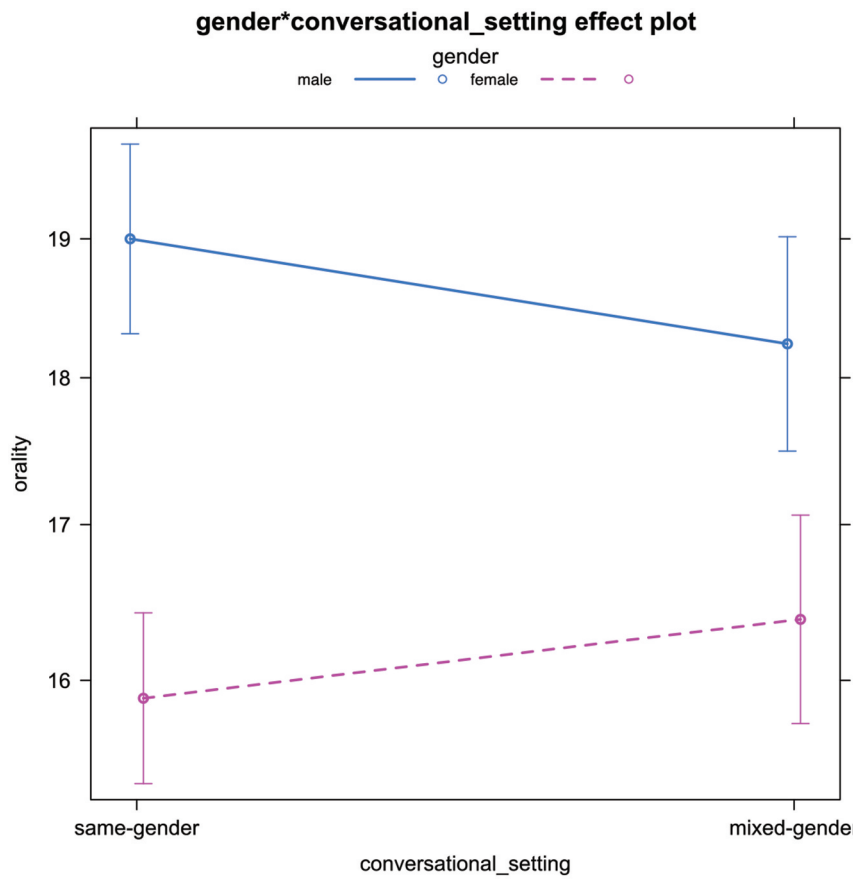

Figure 1. Effect plot for gender accommodation w.r.t. orality (counts per 100 tokens).

press): boys use them more often than girls in any setting. However, an important gender difference can be attested with respect to linguistic adaptation related to interlocutor's gender: boys use fewer oral markers in a mixed-gender than a same-gender setting, whereas the opposite tendency emerges for the girls. So boys seem to adopt a 'less male' writing style and girls a 'less female' writing style when interacting with each other. This pattern can be interpreted as *convergence* since conversation partners adjust their communicative behaviour to appear more similar to each other (see Giles & Ogay, 2007, pp. 295–296). We recall that convergence is generally evaluated positively (Burgoon et al., 2017; Dragojevic et al., 2015, p. 13, 15). However, the female nor the male adaptation is statistically significant. So, while the interaction reveals that boys and girls adapt the frequency of oral markers in opposite ways, the extent to which they actually modify their writing style in this respect is quite limited (this is confirmed by the effect sizes, which equal only 1.04 and 1.03 for the linguistic adaptation by boys resp. girls). Consequently, the linguistic gender difference remains significant in both settings, although it is slightly larger in same-gender than mixed-gender conversations (effect size 1.20 vs 1.12).

Some additional patterns emerged from the model that are not (closely) related to gender accommodation, but that do significantly impact the teenagers' use of oral features on social media. First of all, students in more practice-oriented educational tracks appear to integrate more oral markers than their peers in more theory-oriented tracks. Second, age and gender interact: while all teenagers tend to use fewer oral markers as they age, this decrease is much stronger for girls (see Hilte et al., in press). The third and final additional pattern concerns the interaction of age and conversational setting. Younger teenagers write more speech-like than older teenagers in every setting, but the two age groups also seem to adapt to the setting differently: while older teenagers use fewer oral markers in mixed-gender compared to same-gender talks, the opposite pattern emerges for young teenagers. These findings suggest a potential attitudinal difference: younger adolescents might for instance, want to sound 'cooler' in mixed-gender talks, whereas older teenagers might want to emphasize other personality traits (e.g. being 'adult').

4.2. Expressiveness

The best model for the use of expressive markers includes all five predictors as well as multiple interactions. Tables 5 and 6 present the fixed effects and the ANOVA, respectively.

A significant three-way interaction can be attested between author gender (female/male), conversational setting (same-gender/mixed-gender) and

Table 5. Expressiveness: fixed effects (Reference group: younger boys, general education, same-gender one-on-one chats).

	Estimate	Std. Error	z value	Pr(>|z|)	Signif.
(Intercept)	−3.16620	0.06983	−45.34	< 2e-16	***
GenderFemale	0.98889	0.07236	13.67	< 2e-16	***
SettingMixed-gender	0.42084	0.08786	4.79	1.67e-06	***
InterlocutorsGroup-chat	0.02942	0.08622	0.34	0.732942	
AgeOlder	−0.09484	0.05805	−1.63	0.102299	
EducationTechnical	−0.02973	0.07040	−0.42	0.672755	
EducationVocational	0.07998	0.07328	1.09	0.275066	
GenderFemale:SettingMixed-gender	−0.69160	0.09655	−7.16	7.88e-13	***
GenderFemale:InterlocutorsGroup-chat	−0.19973	0.09425	−2.12	0.034069	*
SettingMixed-gender:InterlocutorsGroup-chat	−0.43044	0.11092	−3.88	0.000104	***
GenderFemale:AgeOlder	−0.35153	0.07734	−4.55	5.48e-06	***
SettingMixed-gender:EducationTechnical	0.13570	0.08749	1.55	0.130883	
SettingMixed-gender:EducationVocational	0.25201	0.09123	2.76	0.005736	**
InterlocutorsGroup-chat:EducationTechnical	−0.12955	0.08502	−1.52	0.127560	
InterlocutorsGroup-chat:EducationVocational	0.27554	0.09149	3.01	0.002599	**
GenderFemale:SettingMixed-gender: InterlocutorsGroup-chat	0.43487	0.14698	2.96	0.003090	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Table 6. Expressiveness: Anova.

	Chisq	Df	Pr(>Chisq)	Signif.
Gender	148.8763	1	< 2.2e-16	***
co.nversational_setting	5.7226	1	0.0167476	*
nr_interlocutors	14.0141	1	0.0001814	***
Age	55.6098	1	8.838e-14	***
Education	31.7088	2	1.302e-07	***
gender:conversational_setting	47.1050	1	6.728e-12	***
gender:nr_interlocutors	0.1117	1	0.7381945	
conversational_setting:nr_interlocutors	6.3095	1	0.0120092	*
gender:age	20.6615	1	5.481e-06	***
conversational_setting:education	7.6916	2	0.0213695	*
nr_interlocutors:education	19.9356	2	4.688e-05	***
gender:conversational_setting:nr_interlocutors	8.7537	1	0.0030898	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

number of interlocutors (one-on-one/group-chat). This interaction is visualized in Figure 2. It confirms that expressive features are primarily 'female' markers (Hilte et al., in press): regardless of the setting or the number of interlocutors, girls use these markers more often than boys. But we also see how girls always use fewer expressive markers in a mixed-gender than a same-gender setting, while the opposite is true for boys. In other words: girls seem to write in a more prototypically 'male' (or less 'female') and boys in a more prototypically 'female' (or less 'male') style when interacting with each other, which means that their writing styles are converging.

Figure 2 shows that the convergence is more outspoken in one-on-one conversations (left panel) than in group chats (right panel). In one-on-one talks, the female adaptation is only borderline significant (and loses

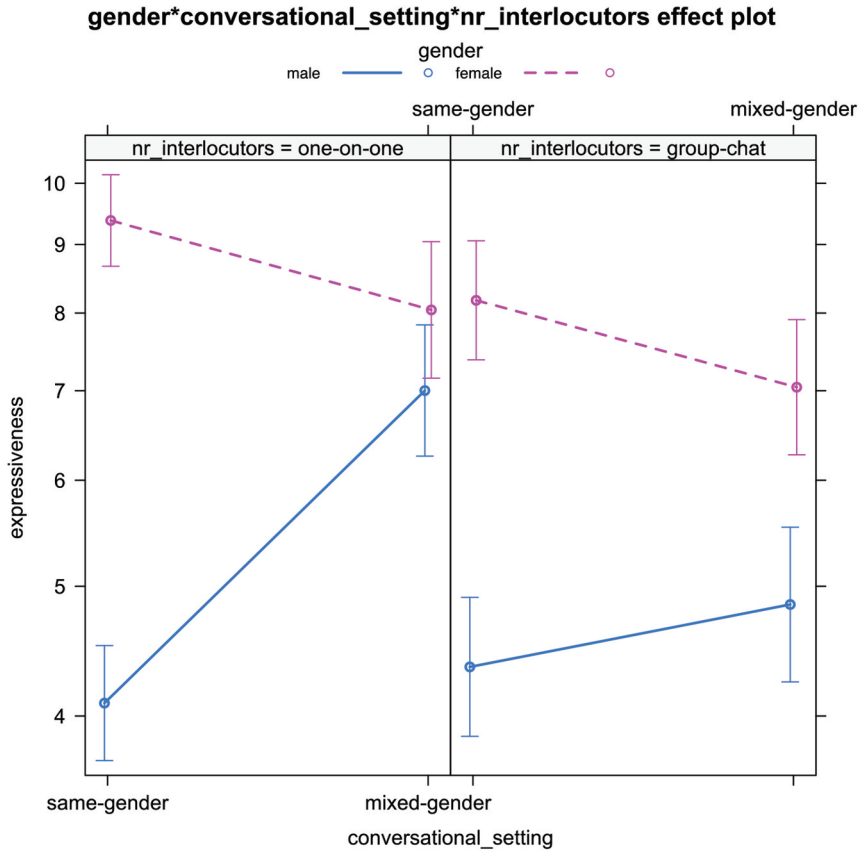

Figure 2. Effect plot for gender accommodation w.r.t. expressiveness (counts per 100 tokens).

significance after Bonferroni correction for multiple testing), whereas the boys significantly adapt their writing style depending on their interlocutor's gender. The effect sizes also reveal that the latter modify the use of expressive markers in one-on-one talks to a much greater extent than their female peers (effect size 1.73 vs 1.15). Consequently, the gender divide is much larger in same-gender than in mixed-gender one-on-one talks (effect size 2.25 vs 1.13). In mixed-gender settings, the linguistic gender difference even becomes insignificant, which indicates that in those contexts the frequency of expressive markers is comparable in boys' and girls' writing.

In group chats (Figure 2, right panel), however, neither girls nor boys significantly adapt their writing style on the basis of their interlocutors' gender. Consequently, the gender difference remains significant in both same- and mixed-gender group chats, although it is more outspoken in the former (in same-gender settings, the effect size of the gender difference

equals 1.85, versus 1.43 in a mixed context). This 'weaker' convergence in group chats may have two explanations. First of all, one-on-one conversations tend to be of a more personal and intimate nature than group talks, and trust may facilitate linguistic convergence (Riordan et al., 2013). Second, linguistic 'mirroring' is more straightforward when there is only one other interlocutor. Group conversations are more problematic in this respect as they might pose 'accommodative dilemmas' (Dragojevic et al., 2015, p. 17): to whom does one adapt? Do multiple interlocutors converge to one central/dominant person in the group, as has been suggested previously (Dino et al., 2009; Doyle et al., 2016; Jones et al., 2014; Noble & Fernández, 2015)? Or can the group be 'divided' into smaller units (e.g. pairs) of interlocutors who converge towards each other? We note that most previous research on accommodation concerns one-on-one interactions only, although the analysis of group conversations has been highlighted as an important path for further research (Dragojevic et al., 2015, p. 17).

We can conclude that in one-on-one interactions, boys adapt their online writing style to a more expressive 'female' standard when interacting with girls, and vice versa (although girls' adaptation is only marginally significant). For group chats, a similar convergence pattern seems to emerge from Figure 2, but the adaptation by girls nor boys is statistically significant.

While classical gender differences with respect to expressive writing (i.e. girls/women using much more expressive markers) hold in group chats, irrespective of the conversational setting, this is not the case for one-on-one interactions: in one-on-one mixed-gender talks, this gender divide becomes insignificant. Strikingly, the adjustive behaviour is mutual but not symmetrical: boys adapt more strongly to a 'female' style than vice versa. This particular pattern contradicts previous observations for spoken conversations, since several studies have found that women adjust more strongly to men than vice versa (Palomares et al., 2016, p. 133 and references therein), while others lay bare patterns of mutual divergence as a way of emphasizing social roles (Burgoon et al., 2017; Dragojevic et al., 2015, p. 15). However, Wolf's (2000) findings on emoticon use in CMC are corroborated since she also found stronger accommodative behaviour for the male participants. The present study does not only confirm her findings, but generalizes them too, as a much larger corpus of instant messages was investigated and, apart from emoticons, a wide range of expressive features were included. In addition, the attested pattern appears to hold after correcting for confounding factors, such as the effect of other socio-demographic variables (e.g. age, educational track) and contextual parameters (e.g. the number of interlocutors).

Finally, some additional patterns emerged that are not (closely) related to gender accommodation. First of all, age and gender interact: all teenagers use fewer expressive markers in their online discourse as they age, but this decrease is much stronger for girls (see also Hilte et al., in press). A second

side-pattern concerns the interaction between educational track and conversational setting. More practice-oriented – and especially vocational – students appear to use much more expressive markers in mixed-gender settings, while students in general (theory-oriented) secondary education use slightly *fewer* expressive markers in mixed-gender interactions. This result may be related to one of our previous findings, i.e. that vocational students value expressive markers to a much greater extent than their peers in other tracks (Hilte et al., in press). Gaining social approval might be more of an issue for teenagers in mixed-gender settings, and vocational students may therefore use their own preferred markers even more in these conversations. Finally, education interacts with number of interlocutors. While students in the two more theoretical tracks use fewer expressive markers in group chats compared to one-on-one conversations, the opposite pattern is attested for vocational students, which suggests the existence of different group dynamics depending on the members' educational background and may once again be related to fishing for social approval by means of favoured features. We also note that the writing style of technical students' alternately show greater similarity to that of their vocational resp. general peers in these final two interactions. This aligns with conclusions from previous research on the present database. Teenagers in the technical track appear to be a hybrid and unpredictable group: while they hold a middle position on the educational continuum from theory to practice, they certainly do not (always) hold a middle position with respect to online writing practices (see Hilte et al., 2018, in press, 2018a, 2018c).

5. Discussion: Expressive Accommodation or Flirting?

The statistical analyses discussed in Section 4 reveal patterns of linguistic gender accommodation in Flemish teenagers' informal online interactions. Our findings show that the convergence predominantly manifests itself in the frequency of typographic expressive markers, i.e. at the level of the maxim of expressive compensation (see Section 3.2). We recall that, although accommodative adjustments were made by the teenagers with respect to oral (speech-like) markers too, this linguistic adaptation was not statistically significant for any of the two gender groups.

The insertion of oral versus expressive features in online discourse largely corresponds to two distinct ways of 'diverging' from formal, standard ways of writing, i.e. through the integration of 'old' and 'new' vernacular (Androutsopoulos, 2011, p. 146; see also Hilte et al., in press). Old vernacular features (i.e. the oral, speech-like markers in informal online writing) represent 'traditional' non-standard language that is not bound to the relatively new communicative practices of social media, whereas new vernacular features (e.g. the – predominantly typographic – expressive markers)

encompass new ways of diverging from formal writing practices, related to social media and informal computer-mediated communication (Hilte et al., in press). These two types of vernacular tend to have quite different connotations: while the use of old vernacular, and especially dialect and regional language features, might evoke a sense of informality, localness and in some cases also toughness, new vernacular features are rather indexical of dynamism and trendiness (Grondelaers & Speelman, 2013, p. 178), and in the case of expressive markers also of involvement and empathy. From previous research we concluded that teenage boys are typically more drawn to the former and girls to the latter for building social capital (Hilte et al., in press). Since accommodation is considered to be (at least partially) driven by a need for social approval (Dragojevic et al., 2015), it might not be surprising that this approval is primarily pursued by means of markers of social and emotional involvement rather than by markers of colloquial speech.

We note that the expressive typographic markers included in the present study can serve as 'tools' for flirting (see also Hilte et al., 2018b). This is especially true for the three markers that appear to be most susceptible to accommodative change in terms of odds ratio: emoji/emoticons (e.g.:), 😂, 🔥), deliberate letter repetition or 'flooding' (e.g. *suuuuuper*) and the typographic rendering of kisses (e.g. *xxxxx*). This raises the following conceptual question: does linguistic gender accommodation equal flirting? Below, we will discuss the findings of some exploratory analyses.

The outcomes of our analyses suggest that linguistic gender accommodation and flirting are related yet distinct phenomena. We argue that expressive markers are not solely used for flirting, but that they are truly part of a primarily female online writing style. This is supported by quantitative as well as qualitative findings. On a quantitative level, the analyses discussed in Section 4.2 show how teenage girls use more expressive markers when chatting to each other (i.e. in all-girls conversations) than when interacting with boys. Consequently, only boys use significantly more 'flirty' expressive features in a (one-on-one) mixed-sex context. If these linguistic adaptations were purely related to flirting/romance, we would expect an increase in expressive markers for both girls and boys, and not for boys only. These quantitative results can be complemented by the findings of a more fine-grained (yet small-scale) analysis in which we compared teenage girls' and boys' top-50 most frequently used emoticons and emoji in same-gender versus mixed-gender conversations. While much more 'romantic' emoji (e.g. heart-emoji, faces with heart-shaped eyes, kissing faces, etc.) figure in the boys' top-50 when they are interacting with girls versus with other boys, the opposite pattern can be observed for girls: girls seem to 'temper' their use of these particular emoji in a mixed-gender setting, as several love-related emoji only figure in their top-50 for all-girl talks and not for mixed talks. For instance, some emoticons/emoji that only occur in the boys' top-50 in mixed

settings and not in same-gender talks are <3, 😘, and 💕, while some examples that no longer figure in girls' top-50 in mixed-gender settings but that do in same-gender talks, are 💖, 💗, and 💞.

On a qualitative level, we conducted an exploratory analysis by manually inspecting the use of heart-emoji in a random subset of the data for which the relationship between the interlocutors was annotated. The qualitative analysis that is discussed below demonstrates how heart-emoji can serve multiple functions in online discourse (see e.g. Dresner & Herring, 2010). Interestingly, clearly distinct gender patterns can be observed that are highly relevant with respect to making a distinction between flirting and gender accommodation.

First of all, emoji can act as *structural* or *syntactic* markers, i.e. 'serving as punctuation in place of traditional punctuation marks' (Spina, 2017, p. 27, 2019). We note that in informal online interaction, messages that end in full stops tend to be perceived as unkind or grumpy by youths, whereas emoji tend to be interpreted as more neutral/friendly post endings – a tendency that suggests the existence of alternative writing standards and conventions on digital media compared to 'classical' written genres (see Hilte et al., 2019). In the selected subcorpus, many teenagers indeed systematically end their messages with one or more emoji, in a punctuation-like way, and this also holds for heart-emoji. These emoji often do not seem to express anything in particular with respect to the content of the utterances they accompany: they rather seem to be conventionalized expressions of social closeness between the interlocutors, i.e. emoji serving as 'social markers of familiarity' or, on a more general level, as 'relational icons' (Spina, 2019, p. 346). In addition, they may mitigate the utterance they accompany and consequently serve as indicators of illocutionary force (Dresner & Herring, 2010), i.e. as pragmatic markers. In example (1), two lovers, a boy and a girl, are discussing a rather neutral (non-romantic) subject, while systematically ending their utterances with multiple heart-emoji:

(1)

girl: Kga da vragen en al gedoucht? ❤️❤️❤️	*I'll ask and did you shower yet?*❤️❤️❤️
boy: Nee ik ga nu en vraag da dan nu schat ❤️❤️	*No I'm going to do that now and ask it now honey*❤️❤️
girl: Die zyn ni thii	*They are not hoem*
girl: Thuis❤️❤️	*Home*❤️❤️
boy: Stalk die bel die sms die❤️❤️❤️	*Stalk them call them text them*❤️❤️❤️
girl: Hahahhaha❤️❤️	*Hahahhaha*❤️❤️

The hearts do not relate to the content of the messages, but continuously visualize the romantic relationship. They are relational icons that in this particular case do have a flirty connotation. So they both have a structural and a social-relational function. Strikingly, the use of hearts or other symbols of romantic intimacy (e.g. kisses) as structural markers of post endings is nearly absent in mixed-gender conversations between boys and girls who are 'just' friends. Example (2) is symptomatic in that respect: the girl immediately corrects herself after accidentally sending a heart to her male friend, and the boy comments upon it:

(2)

girl: Aah ke leuk leuk❤️	*Aah okay nice nice*❤️
girl:😂moest da zijn	*I meant*😂
boy: gy stuurt heel graag hartjes zeker	*looks like you really like sending hearts*
girl: Ni na u	*Not to you*
boy: ne das waar	*no that's true*
girl:😂😂😂	😂😂😂

Example (3) presents a comparable case: the girl who accidently sends kisses ('xx') to a male friend explains her mistake in terms of an automatism she retains from interacting with a female friend:

(3)

girl: Haha best ja xx	*Haha that's for the best xx*
girl: Oeps die 'xx' was perongeluk😉	*Oops that 'xx' was by accident*😉
boy: ja ja	*yeah sure*
girl: Haha jaah k ben ook me n vriendin aant sture en daar zerk heel de rijdxx achter en dus doe k da nu automatisch😜	*Haha yes I am also texting with a female friend and then I put xx at the end all the time so I do it automatically now* 😜

This spontaneous meta-comment perfectly describes a pattern that can be attested repeatedly in the dataset: female friends frequently (and often excessively) insert heart-emoji as structural, punctuation-like markers that also have a clearly social function (i.e. implying closeness/friendship). See example (4) below:

(4)

girl1: Nja hoe ist op sgool😍	*How is school*😍
girl2: Gooeeeeed bij u ?❤️💋	*Goooooooooood, and for you ?*❤️💋

girl1: Suuuuuuuuuuuupppper goed😍❤️💕	*Suuuuuuuuupppper good😍❤️💕*
girl2: Vertel💋❤️	*Tell me💋❤️*
girl1: Kheb kei goeie punten😍😂😂	*I have really good grades😍😂😂*
girl2: Aaaah ik ook hahaha❤️💋😂😂	*Aaaah me too hahaha❤️💋😂😂*

In conversations among *male* friends, however, this particular use of heart-emoji is nearly absent, and when it does occur, it is judged highly inappropriate:

(5)

boy2: ik zal zelf ook niks zegge over u hoor anders vemoort gij mij 😛😛😛❤️❤️❤️❤️❤️❤️❤️❤️😛😛😛❤️❤️❤️❤️❤️❤️❤️❤️	*I won't say anything about you or else you'll kill me*
boy1: ja e eigelijk wel en da zijn verdomd veel hardjes homo! xD	*yeah that's true and that's a hell of a lot of hearts you faggot! xD*
boy2: euh sorry vor die harte da was vo GIRLS_NAME sorry gewoonte	*uhm sorry for those hearts that was for GIRLS_NAME sorry habit*
boy2: lol	*lol*
boy1: TIS FUCKING TE HOPE xD	*I FUCKING HOPE SO xD*

Consequently, the insertion of heart-emoji as structural markers with a (secondary) social dimension seems to truly be a part of an adolescent female, and not male, online writing style.

However, heart-emoji may also explicitly reinforce or 'boost' utterances of a more intimate romantic nature. This frequently occurs in conversations among (both hetero- and homosexual) lovers:

(6)

boy: 💗💗💗💗💗💗💗	💗💗💗💗💗💗💗
girl: Khou vnuu😱💏💏	*Love youu😱💏💏*
boy: Koook vnuuuu💋💋🙀	*Love youuuu toooo💋💋🙀*
girl: King🙌💏😍	*King🙌💏😍*
boy: Nnnnjjjjjjjaaaaaaa 💋💋😍😍💞	*Yeeeeessss💋💋😍😍💞*
[…]	
boy:😍	😍
girl: loveyou!!!😍😍😍😍😍😍 😍😍😍😍😍😍😍😍😍 😍😍😍😍😍😍😍😍😍 😍😍😍	*loveyou!!!😍😍😍😍😍😍😍😍😍😍😍*

boy: I love you too en en sorry schat !!😍😍😍😍😍	*I love you too and and sorry honey !!* 😍😍😍😍😍

And:

(7)

boy1: Kmis u 💋💋💋💋	*I miss you* 💋💋💋💋
boy2: Ik jou ook zie ke vandaag	*Me too will I see you today*
boy2: Hou Van je	*Love you*
boy2:❤️❤️❤️❤️❤️	❤️❤️❤️❤️❤️
boy2: Xxxxxx	*Xxxxxx*
boy1: Hou ook van jou	*Love you too*
boy1:❤️❤️❤️❤️❤️❤️	❤️❤️❤️❤️❤️❤️

In examples (6) and (7), the emoji are 'intensifying the sentiment already conveyed by sentiment-carrying words' (Hogenboom et al., 2015, p. 27): the hearts primarily stress the message of love that is expressed verbally. This particular use of heart-emoji is nearly absent in mixed-gender conversations among friends. However, it once again frequently occurs in conversations among *female* friends. The examples from the dataset, e.g. example (8) below, thus suggest that that girls are much less reluctant than boys to explicitly express their intimate feelings of friendship/appreciation towards each other, and reinforce those messages with all kinds of heart-emoji and kisses:

(8)

girl2: Ik zweer ik ga er echt altij voor u zijn enzo	*I swear I'm really always going to be there for you*
ik zie u echt dood graag x	*I really love you to death x*
girl1: ooohn ku ook meis xx	*oooh me too girl xx*
girl2: [. . .] ik vind u echt een zalige griet ni normaal xx	*[. . .] I really think you're an awesome chick for real xx*
girl1: hahah ku ook gy bent egt een van mijn beste vriendinnen geworden x	*hahah me too you've really become one of my best friends x*
girl2: Ja gij ook echt van mij echt meer enzo x	*Yes and you one of mine and more x*
girl1: hahaha oohn twins xx	*hahaha ooh twins xx*
girl2: Lyyy❤️❤️	*Lyyy*❤️❤️
girl1: Too😍😍😍	*Too*😍😍😍

We did not find any examples of this particular use of heart-emoji in conversations among *male* friends. Consequently, these types of insertions of

hearts also seem to be part of an adolescent female, and definitely not a male, online writing style.

Finally, heart-emoji occur in the analysed subcorpus as 'emotional markers', i.e. to express a sentiment or emotion with respect to what is being discussed, rather than with respect to one's interlocutor. In this case they often add a sentiment that is 'not conveyed by any clearly positive or negative words in a text segment' (Hogenboom et al., 2015, p. 27). When hearts are inserted in mixed-gender conversations among friends, they generally serve this particular function. For instance, in example (9), the sentiment of appreciation expressed by the faces with heart-shaped eyes does not concern the interlocutor, but relates to what is being discussed, i.e. a favourite meal:

(9)

boy: Wais u lievelingseten😂	*What is your favorite meal😂*
girl: Da heb ik ni😂😂😈	*I don't have one😂😂😈*
girl: En jij	*And you*
boy: Macaroni😍😍😍	*Macaroni😍😍😍*

Furthermore, when hearts occur in conversations among *male* friends – which they rarely do –, they most often fulfill this particular role. Consequently, the insertion of hearts as an emotional marker rather than as a social or structural marker in some cases does appear to be accepted among male peers who are not romantically involved. In example (10), for instance, boy number 1 expresses appreciation of a famous soccer player through heart-emoji:

(10)

boy1: Die spits scoort alles carrasco	*That striker scores everything carrasco [i.e. soccer player]*
wauuww echt nice spelers man	*woww really good players man*
boy2: nice	*nice*
boy1: Idd	*Indeed*
boy1:😍😍😍	😍😍😍

In conclusion, the quantitative analyses in Section 4 as well as the exploratory qualitative analysis discussed in the present section have demonstrated that while expressive markers, and heart-emoji in particular, can certainly be inserted in online interactions as 'flirty' or 'romantic' features (see also Hilte et al., 2018b), they are also clear markers of an adolescent female online writing style, in both romantic and non-romantic conversations. So the use of particular emoji need not involve flirting, at least not in girls' chat. Finally, the qualitative and

quantitative analyses with respect to expressive markers yield complementary results: they reveal that boys make a much stronger quantitative adaptation to a more female style than vice versa (e.g. by inserting much more emoticons and emoji when talking to girls than when talking to other boys), but with respect to the nature of the inserted expressive markers, they are less inclined to adopt typically 'female' features (e.g. heart-emoji), except when flirting is involved. Girls, on the other hand, adapt their writing style to a lesser extent than boys do on a quantitative level (i.e. in terms of feature frequency). However, they do make a stronger adaptation with respect to the nature of the inserted expressive markers, as they tend to avoid the use of certain typically female emoji when interacting with boys.

6. Conclusion

The present paper investigated the phenomenon of linguistic gender accommodation in an online setting. A large corpus of Flemish teenagers' (Dutch) informal online interactions was investigated. The analyses showed how the teenagers adapt their writing style in mixed-gender talks which consequently converges to that of their conversation partner. While no significant adaptation could be observed with respect to (more prototypically 'male') oral chatspeak markers, such as regional language features, significant patterns of convergence were attested for (more prototypically 'female') typographic expressive markers, such as emoticons. We found that in one-on-one talks girls insert significantly *fewer* and boys significantly *more* expressive markers when conversing with someone of the opposite sex. However, this pattern of gender convergence proved to be asymmetrical, with boys converging much more strongly to a 'female' writing style than vice versa. Strikingly, this is not in line with previous research on spoken interactions, in which it has either been suggested that women are more inclined to adapt their communicative style than men, or that men and women would mutually diverge in order to consolidate social (gender) roles (Burgoon et al., 2017; Dragojevic et al., 2015, p. 15). Our results do confirm the findings of Wolf (2000) with respect to emoticon use in online writing, since she also observes stronger patterns of accommodative behaviour among the male informants. However, the present study covers a much wider range of expressive markers and relies on a more extensive corpus of private conversations, which strengthens the generalizability of the observed tendencies. The conclusion is not simply that accommodation is not restricted to face-to-face oral interactions but extends into online written interactions, but also that classical views on gender accommodation are challenged when studying social media conversations. While we observe a very classical gender divide in terms of

favouring 'speech-like writing' versus 'expressive writing', we do not observe a classical gender pattern in terms of gender accommodation, since in terms of feature frequency the men outperform the women in adjusting their writing style.

Since significant linguistic adaptation occurred for expressive chatspeak markers only, which can serve as 'tools' for flirting, we investigated to which extent linguistic gender accommodation can be obscured by flirting. Based on the quantitative analyses and an exploratory qualitative analysis on the use of specific typographic markers with a connotation of love or romance, we argue that they are related yet distinct phenomena, and that even these types of expressive markers are not solely used for the purpose of flirting, since they are truly part of a general female adolescent online writing style. While women show less accommodative behaviour in terms of feature frequency, they seem to take in to account male aversion for particular features (e.g. heart emoji) by avoiding them in mixed-gender conversations. Men accommodate in terms of general feature frequency (e.g. overall emoji frequency), but, as far as the pilot study shows, much less so in terms of adopting particular female features (e.g. heart emoji), except when flirting is involved. The qualitative analysis needs to be extended in future research, but these preliminary results suggest that we need to distinguish between more quantitative and more qualitative accommodation.

In the present research design, three confounding factors with respect to gender accommodation were included: the teenagers' age and educational track, and the number of interlocutors in a conversation. An interesting follow-up study could be conducted with respect to the contextual variable by focusing on group chats (i.e. including more than two interlocutors) in particular. They have a different conversational dynamic compared to one-to-one interactions and challenge the analysis of accommodation since it is harder to find out who mirrors whom. For instance, do multiple interlocutors adapt their writing style to one network-central/dominant person? Or do other patterns emerge, such as mutual convergence between pairs of interlocutors? Both quantitative and qualitative analyses may shed more light on this topic.

Another path for future work consists in the systematic inclusion in the research design of the relationship between interlocutors, in order to verify whether different patterns emerge for e.g. friends versus lovers. We note that the relationship between conversation partners was found to impact accommodation in previous research: convergence e.g. tends to be stronger among friends than among strangers (Riordan et al., 2013). Furthermore, such a follow-up analysis could enable the investigation of the distinction between linguistic gender accommodation and flirting on a larger scale.

Finally, follow-up research with a diachronic focus could complement the present synchronic study. The inclusion of the temporal dimension of

accommodation (e.g. the chronology of adjustments made by interlocutors throughout a conversation) may shed light on the actual process of convergence by revealing e.g. linear or more irregular patterns with respect to adaptation and linguistic similarity.

Notes

1. We note that the plain text format keeps special characters such as emoji – which are important in the present research – intact.
2. We note that, naturally, all mixed-gender talks include both female and male interlocutors. We split up the row in a 'girls' and 'boys' column to indicate the share of tokens within such mixed-gender conversations that were uttered by girls versus by boys. Furthermore, we wanted to be able to offer a total count of the number of tokens in the corpus produced by girls versus boys (i.e. bottom row of the table).

Acknowledgments

We thank Ella Roelant (StatUA) for statistical advice.

Disclosure Statement

No financial interest or benefit has arisen or will arise from the direct applications of this research. There are no conflicts of interest.

Funding

This research was funded by the FWO (research foundation Flanders) under Grant [12U2620N].

Data Availability Statement

In order to protect the participants' privacy, and following the guidelines of our university's ethical committee as well as GDPR regulations, the collected dataset cannot be made publicly available. For more information on the database, see chapter 1 in https://repository.uantwerpen.be/docman/irua/948a9a/159941.pdf

References

Androutsopoulos, J. (2011). Language change and digital media: A review of conceptions and evidence. In T. Kristiansen & N. Coupland (Eds.), *Standard languages and language standards in a changing Europe* (pp. 145–161). Novus.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2017). *Package 'lme4'*. CRAN (Comprehensive R Archive Network). https://cran.r-project.org/web/packages/lme4/lme4.pdf

Bell, A. (1984). Language style as audience design. *Language in Society*, *13*(2), 145–204. https://doi.org/10.1017/S004740450001037X

Burgoon, J. K., Dunbar, N. E., & Giles, H. (2017). Interaction coordination and adaptation. In A. Vinciarelli, M. Pantic, N. Magnenat-Thalmann, & J. K. Burgoon (Eds.), *Social signal processing* (pp. 78–96). Cambridge University Press.

Danescu-Niculescu-Mizil, C., Gamon, M., & Dumais, S. (2011). Mark my words! Linguistic style accommodation in social media. In S. Srinivasan, K. Ramamritham, M. P. Arun Kumar, E. B. Ravindra, & R. Kumar (Eds.), *Proceedings of the 20th international conference on world wide web* (pp. 745–754). ACM.

De Decker, B., & Vandekerckhove, R. (2017). Global features of online communication in local Flemish: Social and medium-related determinants. *Folia Linguistica*, *51*(1), 253–281. https://doi.org/10.1515/flin-2017-0007

De Siqueira, A., & Herring, S. C. (2009). Temporal patterns in student-advisor instant messaging exchanges: Individual variation and accommodation. In *Proceedings of the 42nd Hawaii international conference on system sciences*. IEEE.

Dino, A., Reysen, S., & Branscombe, N. R. (2009). Online interactions between group members who differ in status. *Journal of Language and Social Psychology*, *28*(1), 85–93. https://doi.org/10.1177/0261927X08325916

Doyle, G., Yurovsky, D., & Frank, M. C. (2016). A robust framework for estimating linguistic alignment in social media conversations. In J. A. Jacqueline Bourdeau, R. N. Hendler, I. Horrocks, & B. Y. Zhao (Eds.), *Proceedings of the 25th international conference on world wide web* (pp. 637–648). Montreal: ACM.

Dragojevic, M., Gasiorek, J., & Giles, H. (2015). Communication accommodation theory. In C. R. Berger & M. E. Roloff (Eds.), *The international encyclopedia of interpersonal communication* (Vol. 1, pp. 1–21). Wiley Blackwell.

Dresner, E., & Herring, S. C. (2010). Functions of the nonverbal in CMC: Emoticons and illocutionary force. *Communication Theory*, *20*(3), 249–268. https://doi.org/10.1111/j.1468-2885.2010.01362.x

[FMET] Flemish Ministry of Education and Training. (2017). Structuur en organisatie van het onderwijssysteem. In Flemish Ministry of Education and Training (Ed.), *Statistisch jaarboek van het Vlaams onderwijs. Schooljaar 2015-2016* (pp. 8–18). Department of Education and Training.

Gasiorek, J. (2016). Theoretical perspectives on interpersonal adjustments in language and communication. In H. Giles (Ed.), *Communication accommodation theory. Negotiating personal relationships and social identities across contexts* (pp. 13–35). Cambridge University Press.

Giles, H., & Ogay, T. (2007). Communication accommodation theory. In B. B. Whaley & W. Samter (Eds.), *Explaining communication. Contemporary theories and exemplars* (pp. 293–310). Lawrence Erlbaum.

Gonzales, A. L., Hancock, J. T., & Pennebaker, J. W. (2010). Language style matching as a predictor of social dynamics in small groups. *Communication Research*, *37*(1), 3–19. https://doi.org/10.1177/0093650209351468

Grondelaers, S., & Speelman, D. (2013). Can speaker evaluation return private attitudes towards stigmatised varieties? Evidence from emergent standardisation in Belgian Dutch. In T. Kristiansen & S. Grondelaers (Eds.), *Language (de)standardisation in late modern Europe: Experimental studies* (pp. 171–192)). Novus.

Harrison, X. A. (2014). Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ*, *2*, e616. https://doi.org/10.7717/peerj.616

Hilte, L. (2019). *The social in social media writing: The impact of age, gender and social class indicators on adolescents' informal online writing practices*. University of Antwerp (doctoral dissertation).

Hilte, L., Daelemans, W., & Vandekerckhove, R. (2018). Predicting adolescents' educational track from chat messages on Dutch social media. In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (pp. 328–334), Association for Computational Linguistics.

Hilte, L., Vandekerckhove, R., & Daelemans, W. (2018a). Adolescents' social background and non-standard writing in online communication. *Dutch Journal of Applied Linguistics*, *7*(1), 2–25. https://doi.org/10.1075/dujal.17018.hil

Hilte, L., Vandekerckhove, R., & Daelemans, W. (2018b). Expressive markers in online teenage talk: A correlational analysis. *Nederlandse Taalkunde*, *23*(3), 293–323. https://doi.org/10.5117/NEDTAA2018.3.003.HILT

Hilte, L., Vandekerckhove, R., & Daelemans, W. (2018c). Social media writing and social class: A correlational analysis of adolescent CMC and social background. *International Journal of Society, Culture and Language*, *6*(2), 73–89. http://www.ijscl.net/article_31069_c69cbd62ef2442ed11f698c8d8ff9cc4.pdf

Hilte, L., Vandekerckhove, R., & Daelemans, W. (2019). Adolescents' perceptions of social media writing: Has non-standard become the new standard? *European Journal of Applied Linguistics*, *7*(2), 189–224. https://doi.org/10.1515/eujal-2019-0005

Hilte, L., Vandekerckhove, R., & Daelemans, W. (in press). Modeling adolescents' online writing practices: The sociolectometry of non-standard writing on social media. In *Zeitschrift für Dialektologie und Linguistik*. Franz Steiner Verlag.

Hogenboom, A., Bal, D., Frasincar, F., Bal, M., De Jong, F., & Kaymak, U. (2015). Exploiting emoticons in polarity classification of text. *Journal of Web Engineering*, *14*(1–2), 22–40. https://dl.acm.org/doi/10.5555/2871254.2871257

Holmes, J. (1992). *An introduction to sociolinguistics*. Longman.

Ismail, N., & Jemain, A. A. (2007). Handling overdispersion with negative binomial and generalized Poisson regression models. Casualty actuarial society forum. *Citeseer*, 103–158. https://www.casact.org/pubs/forum/07wforum/07w109.pdf

Jones, S., Cotterill, R., Dewdney, N., Muir, K., & Joinson, A. (2014). Finding Zelig in text: A measure for normalising linguistic accommodation. In J. Tsujii & J. Hajic (Eds.), *Proceedings of COLING 2014, the 25th international conference on computational linguistics: Technical papers* (pp. 455–466), ACL.

Muir, K., Joinson, A., Cotterill, R., & Dewdney, N. (2016). Characterizing the linguistic chameleon: Personal and social correlates of linguistic style accommodation. *Human Communication Research*, *42*(3), 462–484. https://doi.org/10.1111/hcre.12083

Muir, K., Joinson, A., Cotterill, R., & Dewdney, N. (2017). Linguistic style accommodation shapes impression formation and rapport in computer-mediated communication. *Journal of Language and Social Psychology*, *36*(5), 525–548. https://doi.org/10.1177/0261927X17701327

Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology*, *21*(4), 337–360. https://doi.org/10.1177/026192702237953

Noble, B., & Fernández, R. (2015). Centre stage: How social network position shapes linguistic coordination. In T. O'Donnell & M. Van Schijndel (Eds), *Proceedings of the 6th workshop on cognitive modeling and computational linguistics* (pp. 29–38), ACL.

Palomares, N. A., Giles, H., Soliz, J., & Gallois, C. (2016). Intergroup accommodation, social categories, and identities. In H. Giles (Ed.), *Communication*

accommodation theory. Negotiating personal relationships and social identities across contexts (pp. 123–151)). Cambridge University Press.

Riordan, M. A., Markman, K. M., & Stewart, C. O. (2013). Communication accommodation in instant messaging: An examination of temporal convergence. *Journal of Language and Social Psychology*, *32*(1), 84–95. https://doi.org/10.1177/0261927X12462695

Scissors, L. E., Gill, A. J., Geraghty, K., & Gergle, D. (2009). In CMC we trust: The role of similarity. In D. R. Olsen (Ed.), *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 527–536), ACM.

Scissors, L. E., Gill, A. J., & Gergle, D. (2008). Linguistic mimicry and trust in text-based CMC. In B. Begole & D. W. McDonald (Eds.), *Proceedings of the 2008 ACM conference on computer supported cooperative work* (pp. 277–280), ACM.

Spina, S. (2017). Emoticons as multifunctional and pragmatic resources: A corpus-based study on Twitter. In E. W. Stemle & C. R. Wigham (Eds), *Proceedings of the 5th conference on CMC and social media corpora for the humanities* (pp. 25–29). https://hal.archives-ouvertes.fr/hal-01614310

Spina, S. (2019). Role of emoticons as structural markers in Twitter interactions. *Discourse Processes*, *56*(4), 345–362. https://doi.org/10.1080/0163853X.2018.1510654

Taylor, R. L. (2001). Adolescent peer group language. In R. Mesthrie (Ed.), *Concise encyclopedia of sociolinguistics* (pp. 297–302)). Elsevier.

Varnhagen, C. K., Peggy McFall, G., Pugh, N., Routledge, L., Sumida-MacDonald, H., & Kwong, T. E. (2010). Lol: New language and spelling in instant messaging. *Reading and Writing*, *23*(6), 719–733. https://doi.org/10.1007/s11145-009-9181-y

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

Wolf, A. (2000). Emotional expression online: Gender differences in emoticon use. *Cyber Psychology & Behavior*, *3*(5), 827–833. https://doi.org/10.1089/10949310050191809

Xu, Y., & Reitter, D. (2015). An evaluation and comparison of linguistic alignment measures. In T. O'Donnell & M. van Schijndel (Eds.), *Proceedings of the 6th workshop on cognitive modeling and computational linguistics* (pp. 58–67), ACL.

Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software*, *27*(8), 1–25. https://doi.org/10.18637/jss.v027.i08