

Graphic Means for Expressing Emphasis in Online Communication of Youth *

Die informelle schriftliche Kommunikation von Jugendlichen in der Online-Umgebung zeichnet sich durch ihre Ausdruckskraft aus, die oft durch die Hervorhebung bestimmter Teile der Nachricht durch spezielle, unkonventionelle grafische Mittel unterstützt wird. Das Ziel des Beitrags ist es, eine Analyse dieser grafischen Mittel vorzulegen, insbesondere ihrer Arten, der Häufigkeit ihres Auftretens oder möglicher Gründe für die Wahl eines bestimmten Modus, unter Verwendung von Material aus Gruppenchats, das in einem speziellen Korpus gesammelt wurde.

The informal written communication of youth in the online environment is characterized by expressiveness, which is often supported by emphasising certain parts of the message through special, unconventional graphic means. The aim of the paper is to present an analysis of these graphic means, specifically their types, frequency of occurrence, or possible reasons for choosing a particular technique; using material of group chats, collected in a specialized corpus.

Communication of youth in the online environment often takes place via text messages, sent through applications intended for this purpose, which are usually part of the application interface of social networks. For a more dynamic, lively way of conversation, users attach multimedia elements to their messages, especially graphic ones: most often emoticons, then classic pictures or videos, or so-called gifs

* This paper is dedicated to the project MUNI/A/1249/2022 Lexikon a gramatika češtiny III – 2023, <https://www.muni.cz/vyzkum/projekty/69573> (accessed: 6 May 2025).

or stickers. However, they often make do solely with modifying the text itself, even though they have a limited repertoire of options in this case, since in most applications the text cannot be formatted, e.g. by changing the type, weight, slope or colour of the font. Instead, they use variations of standard writing (most often so-called reduplication of characters), which they apply to a certain passage, typically a single word. This paper describes these methods in more detail, giving examples of them, their frequency of occurrence and possible motivations for their use. The aim is to present the phenomenon as evidence of the specificity of the conversational style of youth in the online environment.

Source of Material and Characteristics of the Researched Community

The data used in the research were sourced from so-called group chats (online group conversations) that were active between 2020 and 2021 on the social networks Instagram and Facebook. These groups were created to bring together peers around the age of 15 who share a common interest, typically the same school, hobby or opinion. The conversations there take place in the Czech language.

Since anyone can become a member of the group if they request it from the administrator of the community within which the groups were created, it is not an entirely private correspondence, but at the same time, the groups are not publicly viewable or accessible. Thanks to this, the communication within these groups is very informal, relaxed and spontaneous, which is related to a high degree of emotionality and liveliness. This is reflected, among other things, in the presence of various expressions of emphasis.

An example of an authentic conversation using graphic text emphasis, specifically the techniques of reduplication and writing in all capital letters (underlined):

A: jn zajimavej jak svine

B: jnnn

C: POCKEJ NIKUSO

D: je mu podobnej holky?

B: myslim ze sara muze byt statna ze snim muze byt ve tride

D: co myslitel?

A: holky??????????????????

(...)

E: Hmmmm, no¹

All material stored in our own dedicated corpus, which contains data from 713 group chats where almost 10,000 members participated in varying amounts of content, was analysed. The corpus contains approximately 4.3 million words.

Character Reduplication

The most common way of graphically emphasising text is called character reduplication or repetition. It is not a new phenomenon; it has been present in electronic communication since its beginnings, as evidenced by some older studies (e.g. Bader 2002: 105; Kalman / Gergle 2014), including research from the Czech context (Hladká et al. 2005: 62; Jandová et al. 2006: 173). Interestingly enough, its popularity was justified by a simple technical solution, which consists in a long key press on the physical keyboard of the computer (cf. Anis 2002; cf. Hladká et al. 2005: 62). Nowadays, young people mostly use mobile devices for chatting, which do not have this function and it is necessary to press the virtual key repeatedly, but this obviously does not prevent the still frequent use of this technique.

Table 1 below shows the percentage ratio of words with multiple consecutive identical letters in our corpus (Ko) compared to the values in the largest Czech web corpus *csTenTen17* (Suchomel 2018) and in the corpus *syn v9* (Křen et al. 2021), which consists of representative texts from public print media: fiction, non-fiction, and especially journalism.

¹ Transl. by the author:

»A: *yeah interesting as heck*

B: *yeahhhh*

C: *WAIT NIKUSA*

D: *does he look like him girls?*

D: *i think sara's lucky that she can be in the same class as him*

D: *what do you mean?*

A: *girls??????????????????*

(...)

E: *uhhhh, well«.*

	KO	<i>csTenTen17</i>	<i>syn v9</i>
words with two or more identical letters in a row (including standard spelling)	3,30 %	1,10 %	0,78 %
words with three or more identical letters in a row	0,87 %	0,047 %	0,007 %
words with four or more identical letters in a row	0,31 %	0,0040 %	0,00028 %
words with five or more identical letters in a row	0,12 %	0,0021 %	0,00012 %
words with ten or more identical letters in a row	0,01 %	0,0007 %	0,000015 %

Tab. 1: Cumulative relative frequency of letter reduplication in different corpora.

In the case of two or more identical letters in a row, the data itself within a single corpus is not very relevant, as it includes words that commonly occur with two identical letters in a row in standard Czech, e.g. *vinný* (›guilty‹), *měkký* (›soft‹), *oddělený* (›separated‹), or it may be a typographical error, so it is not an emphasis. These cases cannot be easily filtered out or determined how many occurrences they make up. Nevertheless, it is the two-character reduplication as a function of emphasis that is the most frequent (due to the simple typing). This can already be well documented by comparison with other corpora, where the majority of such occurrences are probably standard Czech words. It is noticeable that in our corpus the ratio of words with two or more identical letters in a row is several times higher. An even greater difference is observed in the case of three or more repeated characters, which are not present at all in standard Czech (except for certain abbreviations). In other corpora, such occurrences are only recorded to a negligible extent, with *csTenTen17* reaching slightly higher values than *syn v9*. However, the factor of text production on the Internet is evidently not as strong as the factor of the chat and/or young age of the chat participants.

The frequency distribution of particular words with reduplicated letters, with the first fifteen entries shown in table 2 below, reveals that this phenomenon most commonly occurs in very short words (up to four characters in standard spelling), and especially interjections, which corresponds to the emotional nature of emphasis. Typically, the reduplicated letters are the final ones (80% of all instances in the corpus) and vowels (85% of the cases). Due to the user's choice of the reduplication rate (double, triple, quadruple, ...), such cases may serve as a more convenient and variable way of the acute accent which is used for the standard written representation of lengthening (often emphatic) in Czech. Thus, they are getting closer to the spoken form of the language. After all, there are more elements of spokenness in the online

communication of young people, often blurring the distinctions between written and spoken forms (cf. Jandová et al. 2006: 160).

rank	word	frequency
1	<i>joo</i> (< <i>jo</i> (>yeah<))	2 603
2	<i>noo</i> (< <i>no</i> (>well<))	1 609
3	<i>nee</i> (< <i>ne</i> (>no<))	1 506
4	<i>jooo</i> (< <i>jo</i> (>yeah<))	1 308
5	<i>ahaa</i> (< <i>aha</i> (>I see<))	1 057
6	<i>ahooj</i> (< <i>ahoj</i> (>hi<))	997
7	<i>neee</i> (< <i>ne</i> (>no<))	952
8	<i>takyy</i> (< <i>taky</i> (>too<))	905
9	<i>hmm</i> (< <i>hm</i> (>hm<, >um<))	565
10	<i>nooo</i> (< <i>no</i> (>well<))	563
11	<i>jaa</i> (< <i>já</i> (>I/me<))	472
12	<i>taak</i> (< <i>tak</i> (>so<, >such a<))	469
13	<i>anoo</i> (< <i>ano</i> (>yes<))	430
14	<i>ahaaa</i> (< <i>aha</i> (>I see<))	398
15	<i>hmmm</i> (< <i>hm</i> (>hm<, >um<))	392

Tab. 2: Cumulative absolute frequency of the most frequent words with reduplicated letters.

Although the predominant objects of reduplication have been clearly described, there is no limitation of their formal characteristics, and other, though less frequent, examples show that any letter in any position and in any long word can be reduplicated, e.g. *cccccccccoze* (< *čože* (>what<)), *uuurcitee* (< *určité* (>sure<)), *slysseet* (< *slyšet* (>to hear<)). The variety is evident even within a single word; e.g. *olomouuuc*, *olomooouc*, *olomouccc* (< *Olomouc*, the name of a Czech city).

A similar situation applies to other (non-alphabetic) characters, especially punctuation at the end of a sentence. More than three full stops, exclamation marks or question marks in a row occur in 0.049% of cases (in the compared corpora the ratio is significantly lower: 0.034% (*csTenTen17*) and 0.001% (*syn v9*)). The most common (80% of cases; a total of 1,971 instances) are full stops; the longest sequence consists of 36 full stops in a row.

Character reduplication is abundant in text emoticons, which allows to express a greater degree of particular emotion. For example, the right parenthesis in the :)

emoticon was reduplicated in 1,177 instances, and the left parenthesis in the :(emoticon in 1,479 instances. The most common text emoticon is *xd*, which has been reduplicated in 15,130 instances in 53 variants (without distinguishing between uppercase and lowercase). The basic two-character form of *xd* (without any reduplication) accounts for only 65% of occurrences of this emoticon.

Words Reduplication

In addition to individual characters, reduplication also applies to entire words that are written consecutively without any graphic separation. The most frequent entries are listed in table 3 below.

rank	word	frequency
1	<i>jojo</i> (‹ <i>jo</i> ›yeah‹)	1 813
2	<i>nene</i> (‹ <i>ne</i> ›no‹)	1 002
3	<i>okok</i> (‹ <i>ok</i> ›)	216
4	<i>anoano</i> (‹ <i>ano</i> ›yes‹)	62
5	<i>caucau</i> (‹ <i>čau</i> ›hi‹)	49
6	<i>taktak</i> (‹ <i>tak</i> ›like this‹)	37
7	<i>coco</i> (‹ <i>co</i> ›what‹)	29
8	<i>ahaaha</i> (‹ <i>aha</i> ›I see‹)	27
9	<i>cuscus</i> (‹ <i>čus</i> ›hi‹)	20
10	<i>takytaky</i> (‹ <i>taky</i> ›too‹)	19
11	<i>yesyes</i> (‹ <i>yes</i> ›)	18
12	<i>dikdik</i> (‹ <i>dik</i> ›thanks‹)	17
13	<i>cscs</i> (‹ <i>cs</i> › <i>čus</i> ›hi‹)	17
14	<i>jesjes</i> (‹ <i>jes</i> › yes ›)	15
15	<i>okioki</i> (‹ <i>oki</i> › <i>ok</i> ›)	14

Tab. 3: Cumulative absolute frequency of the most frequent reduplicated words.²

2 The list omitted words that formally correspond to the phenomenon being searched for but where reduplication does not fulfil the function of emphasis, e.g. *toto* ›this‹, *mama* (‹ *máma* ›mum‹), or where they represent established forms of interjections recorded in dictionaries of the standard language (e.g. *haha* (interjection of laughter), *papa* (interjection of greeting)).

This kind of reduplication is not as common as the one involving individual characters; after all, its function can be fulfilled by the (more standard) graphical representation of multiple identical words separated by spaces. Moreover, it almost always occurs only in the lowest, twofold reduplication (doubled). Comparison with other corpora again reveals that this phenomenon is an order of magnitude more abundant in terms of cumulative relative frequency in those corpora: 0.21% (Ko), 0.04% (*csTenTen17*), 0.015% (*syn v9*).³ Another characteristic is also similar to the previous case: the reduplication mostly affects short words (faster typing may be a decisive factor here) and interjections, although there is no general limitation: words not fitting this pattern are also reduplicated, e.g. *pravdapravda* (< *pravda* ›truth‹), *nevadinevadi* (< *nevadí* ›does not matter‹), *nechapunechapu* (< *nechápu* ›I do not understand‹).

In addition, abbreviations consisting of the same letters symbolizing the same words are frequent. Formally, they belong to the category of letter reduplication, but essentially, they represent an emphasis on the entire word. The most common forms are *jj* (< *jo jo* ›yes yes‹), *nn* (< *ne ne* ›no no‹) and *kk* (< *ok ok*).

Some forms have been recorded in both categories of reduplication because they represent a combination of these two forms, e.g. *joojoo* (< *jo jo* ›yes yes‹), *dobryydobryy* (< *dobry dobry* ›good good‹).

Writing in All Capital Letters

Another frequently used means is writing in all capital letters, also known from the earlier Internet era (see the older studies mentioned above). It may seem surprising that the frequency of words written entirely in capital letters in our corpus (0.71%) is – in contrast to the phenomenon of reduplication – similar in the compared corpora, with only a few hundredths of a percent difference. The main reason seems to be initial abbreviations such as *USA* (< *United States of America*), *EU* (< *Evropská Unie* ›European Union‹), *ODS* (< *Občanská demokratická strana* ›Civic Democratic Party, a Czech political party‹), which are not as common in youth communication as they are in texts of various genres on the Internet and in public print media. And when they do appear, their representation is simplified by writing them in lowercase,

3 As in the case of two identical letters in a row, these values also include undesirable cases (as mentioned in the previous footnote). However, these cases do not undermine the result of the comparison due to the assumption of their similar frequency across corpora.

since the proper capitalization is rarely respected by youth in the chat environment. To illustrate this, in the *csTenTen17* corpus, the occurrence of the abbreviation *USA* written properly in uppercase is 103 times more frequent than its lowercase representation, in the *syn v9* corpus it is even 1816 times more frequent, but in our corpus, it is only 4.5 times more frequent. It can also be inferred from this that capital letters in the conversations of youth primarily fulfil different functions, especially emphasis. 57% of the occurrences of words written in capital letters are followed by another word also written in capital letters, which usually means that the entire message was written in this way – the author intended to emphasise the whole text, not just a single word (e.g. *A ONO JE TO MEGA CUTE ŽE* (>and it's mega cute, isn't it?<); *POJĎ SE HADAT JINAM* (>let's argue somewhere else<)). However, there are also cases in which particular word within such a message is emphasised in addition, using the most popular technique – reduplication (e.g. *O KURVA NAZDAR MIROOOOOOOOO* (>oh fuck hi Mira<)). Again, we observe a combination of multiple means of emphasis.

Due to the fact that writing in all capital letters is often applied to the entire message, which contains commonly used words, including synsemantics, the frequency distribution does not provide as specific results as with the previous phenomena. Nevertheless, we can observe an increased occurrence of expressives: among the 75 most frequent entries, there are several variants of the emoticon *XD*, international emotive abbreviations like *OMG* (< oh my God), *LOL* (< laugh out loud), *WTF* (< what the fuck), interjections expressing (dis)agreement *ANO* (>yes<), *JO* (>yeah<) and *NE* (>no<), or the Czech vulgar expressions *KURVA* and *VOLE*. In addition, a considerable portion of the list is covered by pronouns referring to the first person, or verb forms in the first person singular, which may be related to the desire to emphasise oneself and to assert one's presence within the collective, e.g. *KÁMO JÁ JSEM LEGENDA* (>dude I'm a legend<); *Ehmmmm ME NIKDO NEPREKONA* (>ummmm nobody beats me<); *Je to sračka (MŮJ NÁZOR!!!!)* (>it's shit (my opinion!!!!)<)).

Alternating Caps

Among the lesser-used means is the alternation of uppercase and lowercase letters within a single word (in any combination except for the initial capital letter followed by lowercase letters, which is the usual beginning of a sentence or proper noun according to Czech spelling rules – although such cases can still be part of the emphasis of the entire message; see below). The formal characteristic of this phenomenon also

includes a noticeable number of undesired (nonemphasising) instances (e.g. *WiFi*, *McDonald*, abbreviations such as *LoLko* (< the computer game *League of Legends*), including typographical errors. It was therefore more difficult to filter out the actual number of instances of emphasising, which ended up being only a few hundred, corresponding to less than one-hundredth of one percent of all words in the corpus. The reason can be both a technically demanding process of typing, which requires switching the virtual keyboard after each character, and the general eccentricity of this style: while reduplication can replace the quantity of vowels and writing in all capital letters is typical in various commonly available texts (headings, titles, banners), the alternating caps has no special function and does not occur in common texts. Therefore, this technique is more suitable for drawing the attention of conversation partners during their quick scanning of chat messages. This corresponds to a certain form of exhibitionism, which is typical for young chat participants on the Internet (cf. Jandová et al. 2006: 167).

This phenomenon rarely applies to the entire message, probably due to the aforementioned demanding process of typing, e.g. *nEčUm Na Mě NeBo Tě RoZbljU* (>don't stare at me or I'll beat you up<); *nEjMeNoVaNaOsObO jAk sE mAšssS⁴* (>unnamedperson how are you<).

Letter Spacing

Another means of emphasis is letter spacing, i.e. separating the individual characters of a word with spaces. It has a similarly low frequency as alternating caps but is more often applied to the entire message. In addition, the distinction between spaces between letters and spaces between words is missing, making the message visually prominent but more difficult for the reader to perceive; e.g. *netusimcosetadyresia vlastnemetoaninezajima* (>i have no idea what's going on here and i really don't care<). Most often, letter spacing is combined with the writing in capital letters, e.g. *SLUŠNĚCHOVÁNÍCHLAPČEBYTINEUŠKODILO* (>good manners young man wouldn't hurt you<). Occasionally, individual characters are written separately in separate messages, creating the effect of vertical text which occupies more screen space and is therefore less easily overlooked.

4 In the last word of this instance, we can once again observe a combination with reduplication.

Other Means of Emphasis

There are several other means of emphasis, but they are only sporadically used. For example, there is the insertion of certain words between special symbols (most commonly asterisks (*) and hyphens (-)), which, however, primarily serve other functions than mere emphasis, such as to indicate the description of an activity in narration (**smích** (laughter); **potlesk** (applause); ...) or structuring text into bullet points.

The insertion of these symbols between individual letters has also been recorded, e.g. *p - o - m - o - c* (help), apparently again used to draw the reader's attention in general.

Conclusion

Emphasising certain messages or their parts with graphically non-standard forms of writing is a common element in online communication of youth, in contrast to other texts on the Internet (based on comparison with the *csTenTen17* corpus) or texts in public print media (based on comparison with the *syn v9* corpus). It can therefore be considered as a characteristic feature of the informal conversational style of youth in the online environment and one of the indications of its distinctiveness.

Despite formatting limitations, there is considerable variability in the graphic means used in the chat, which shows the linguistic and perhaps also aesthetic ingenuity of the young chat participants. However, the frequency of particular means (and their types) is based on the simplicity and speed of their typing. In most cases, a phenomenon called reduplication (or repetition) is used. Other techniques include writing in all capital letters, alternating caps, letter spacing, the use of special symbols or a combination of multiple means. The individual means differ in their frequency of use and scope of application; some are predominantly used for single words, while others are applied to entire sentences/messages. Certain tendencies (but never exclusive rules) can be observed in the formal characteristics of emphasised objects. For example, shorter words are more likely to be emphasised than longer ones, and vowels more than consonants. The essence of emphasis itself leads to its application in passages with an emotive tone, especially interjections, expressives and emoticons. In addition, thanks to some of the means, the spoken form of the language is manifested in a certain way.

Apart from the desire to draw attention to a part of one's message (or the entire message) that could easily go unnoticed in the large volume of other messages in the

group chat, the use of non-standard graphic means is also often caused by the desire to draw attention directly to oneself, to impress one's peers with unconventional expression, and to assert one's presence among them. Thus, this research shows that not only the content but also the form of the text can reveal certain communicative intentions of its authors.

References

- Anis, J. (2002): *Communication électronique scripturale et formes langagières*. Poitiers, <http://rhr.edel.univ-poitiers.fr/documentef73.html?id=547> (accessed: 15 September 2022).
- Bader, J. (2002): *Schriftlichkeit und Mündlichkeit in der Chat-Kommunikation*. Hannover, <https://doi.org/10.15488/2920>.
- Hladká, Z. et al. (2005): *Čeština v současné soukromé korespondenci. Dopisy, e-mail, SMS*. Brno.
- Jandová, E. et al. (2006): *Čeština na WWW chatu*. Ostrava.
- Kalman, Y. M. / Gergle, D. (2014): »Letter repetitions in computer-mediated communication: A unique link between spoken and online language«. In: *Computers in Human Behavior* (34), 187–193, <https://doi.org/10.1016/j.chb.2014.01.047>.
- Křen, M. et al. (2021): *Korpus SYN, verze 9 z 5. 12. 2021*. Praha, <https://www.korpus.cz> (accessed: 15 September 2022).
- Suchomel, V. (2018): »csTenTen17, a Recent Czech Web Corpus«. In: *Proceedings of the Twelfth Workshop on Recent Advances in Slavonic Natural Languages Processing, RASLAN 2018*. Brno, pp. 111–123.

Open Access

This paper is published under the Creative Commons Attribution-ShareAlike 4.0 International license (<https://creativecommons.org/licenses/by-sa/4.0>). Please note that individual, appropriately marked parts of the paper may be excluded from the license mentioned or may be subject to other copyright conditions. If such third party material is not under the Creative Commons license, any copying, editing or public reproduction is only permitted with the prior consent of the respective copyright owner or on the basis of relevant legal authorization regulations.

