The Use of Alphanumeric Symbols in Slovene Tweets

Dafne Marko
Faculty of Arts, University of Ljubljana
E-mail: dafne.marko@gmail.com

Abstract

This paper deals with the use of alphanumeric symbols in Slovene tweets. We use the JANES corpus, a large corpus of internet Slovene containing tweets, forum and blog posts, and comments on news articles and on Wikipedia discussion and user pages. We analyze the use of words consisting of alphabetic and numeric symbols as a means of both creative writing and as a word-shortening strategy. We investigate which alphanumeric features are most frequently used in Slovene tweets and identify the numerals substituting the letters. The results are compared with other subcorpora in the JANES corpus as well as with the Kres corpus, a collection of standard Slovene texts. Furthermore, we compare the distribution of alphanumeric features according to user type and text standardness.

Keywords: alphanumeric symbols, letter/number homophones, Slovene tweets, computer-mediated communication

1. Goal of the Paper

The main goal of the paper is to research the occurrence of a CMC-specific linguistic feature – words consisting of alphabetic and numeric characters. We focus on the subcorpus of Slovene tweets, but also compare the distributions with other subcorpora in the JANES corpus (forum posts, blog entries, comments on news articles, Wiki talk) as well as with the Kres corpus, a corpus of standard Slovene with a balanced genre structure. We predict to find no or very few occurrences in the Kres corpus, proving that it is, indeed, a CMC-specific linguistic feature. We also predict the Twitter subcorpus to be most abundant with this sort of writing, whereas no significant difference between other subcorpora is expected. We research whether gender (male vs. female) or user type (corporate vs. private) influences the use of alphanumeric characters in words. Furthermore, our goal is to carry out a detailed analysis of the most frequently used words with alphanumeric symbols in Slovene tweets. We try to investigate which numeric symbols (numerals) are used to substitute the letters and whether they are used phonetically (e.g., *ju3*, translated as *2morrow*, with a number 3 pronounced as /tri/, the same as in the word *jutri*) or graphically (e.g., *g33k*, where the number 3 represents the letter "e"). With our analysis, we present a linguistic phenomenon which could be described as a type of creativity in writing, and – according to the fact that the length of a single tweet is limited to 140 characters – also as a word-shortening strategy.

2. Related Work

So far, little research has been done on the use of alphabetic and numeric characters in tweets. Most researchers deal with text messaging or *texting* as a relatively new writing medium. It has been pointed out that "/t/here are a great deal of apparent similarities between Twitter and text messaging" since "they are both a medium via which friends and acquaintances can communicate with one another, and they both fall under the broad banner of technology mediated communication" (Denby, 2010). Another shared characteristic is character limitation – Twitter imposes an explicit message length limit of 140 characters, and in text messaging, the limitation is 160 characters. Although there are several differences between the aforementioned writing media (public vs. private, cost of specific service, device used for text messaging and Twitter posts, etc.), research on texting could help us get a deeper insight into the characteristics of another CMC phenomenon – specific linguistic features in Twitter posts.

Most of the researchers claim that "shortenings are presented to be the one major characteristic of text messaging that is assumed to be technologically determined by the limited number of permitted characters and the cumbersome input via the small cellular phone keypad" (Bieswanger, 2006). Language used in texting, or what Crystal (2001) refers to as *Netspeak*, is assumed to be "heavily abbreviated" (Thurlow, 2003), although Thurlow reports "relatively few (n = 73) examples of language play using letter-number homophones (e.g. Gr8 'great', RU 'are you'), which, in popular representations at least, have become the most definitive feature of text-messaging".

Some authors claim that Twitter posts, which fall into the category of *microblogs* (Moseley, 2013) or *microtexts* (Gouws et al., 2011), are rich with abbreviations "solely to conserve space within a text" (Alkawas, 2011), when others observe that "SMS language seems to have evolved into a fashionable and stylish way of writing where the way of writing is as important as the content" (Kirsten Torrado, 2014).

The linguistic phenomenon discussed in this paper is often referred to as *letter/number homophones* (comp. Bieswanger, 2006; Kirsten Torrado, 2014; Frehner, 2008; Kadir et al., 2012; Elizondo, 2011; Farina and Lyddy, 2011; Thurlow, 2003; Kul, 2007; Alkawas, 2011) or (*alphanumeric*) *rebus writing* (Halmetoja, 2013; Danet and Herring, 2007), but we can also find wider, more generic expressions, e.g., *complex abbreviations*

(Filipan-Žignić et al., 2012) or *textism* (Grace et al., 2012; Bushnell et al., 2011). Crystal (2001) refers to this phenomenon as a “rebus-like potential” of letters and numbers “whose pronunciation is identical with words or parts of words” and “are used to replace words or letter sequences”. Frehner (2008) differentiates between *letter homophones* – the use of a single letter whose phonological content is equated with a word, e.g., “u” for “you”, *number homophones* – the use of a numeral whose phonological content is equated with a word, e.g., “4” for “for”, and a combination of letters and numerals, forming *letter/number homophones*, e.g., “b4” for “before”.

Such shortenings can also be observed in Slovene. Michelizza (2008) talks about a special group of abbreviations, “typical for the language of SMS, where parts of a word are substituted by a mathematical symbol or numeral, pronounced the same or at least similar to the part of the word it substitutes (e.g., *ju3*)”[1]. Dobrovoljc (2008) lists the most frequently used letter/number homophones in Slovene (ju3 = “jutri”, pr8 = “prosim”, 5er = “Peter”, 1x =“enkrat”) and English (4yeo = “for your eyes only”, j4f = “just for fun”, 2 much = “*too much*”), but concludes that there is a low percentage of such writings in Slovene texting language. Logar (2006) names this kind of linguistic feature a “combination of various writing symbols”, which are commonly observed in Slovene SMS. In her research, she showed that “/a/mong the more than 450 examples of SMS abbreviations that had been submitted to the site by 11 January 2002, more than 60% were some type of abbreviation, while the rest of the material (160 examples) was made of, for example, the following: :-) ‘zadovoljen’, :) ‘veselje’, :(... ‘jočem’, :x ‘poljubček’, :D ‘širok nasmešek’, mi2 ‘midva’, ju3 ‘jutri’, 2mač ‘preveč’, sk8ar ‘skejtar’, 8-) ‘Nosim očala’, <>< ‘ribica’, {*} ‘objemček, poljubček’, *+* ‘vidim te’, @x@ ‘maš mačka?’, @->-- ‘vrtnica’, _/0 ‘A greš na kavo?’, =:x ‘zajček’.” Since most of the researches mentioned above focused only on the use of alphanumeric symbols in texting, we will investigate how often they occur in Slovene tweets.

3. Dataset and Methodology

For our research, we used the JANES v0.4 corpus[2], a large corpus of Slovene tweets, forum posts, blog texts, comments on news articles and on Wikipedia pages and users, which contains over 175 million words or 9 million documents, published between 2002 and 2016 (Fišer et al., 2016). We focused on the biggest subcorpus, the Slovene tweets, which consists of 90.180.337 words from 7.503.199 different Twitter posts.

Using the concordancer SketchEngine[3], we searched for all occurrences of words consisting of alphanumeric symbols, where the numerals appear at the beginning, in the middle, or at the end of the word. To achieve that, appropriate regular expressions were used for querying the corpus: [word="[0-9]+[a-zA-ZščžŠČŽ]+"], [word="[a-zA-ZščžŠČŽ]+[0-9]+[a-zA-ZščžŠČŽ]+"], and [word="[a-zA-ZščžŠČŽ]+[0-9]+"]. Prior to further analysis, irrelevant results had to be manually selected and excluded from the list. This was done because there are numerous examples of words which consist of alphanumeric symbols but represent a proper name/part of a proper name, a chemical symbol, a unit of measurement, or some other abbreviation (e.g., *A4*, *CO2*, *C4*, *TEŠ6*, *m2*, etc.). With these examples, no transformation from numerals to letters can be made in the written form (e.g. A4 → *A-štiri, CO2 → *CO-dva, etc.). After a quick overview of the concordances, we created a frequency list of all the words with alphabetic and numeric symbols which appear in Slovene tweets.

To investigate which users (corporate or private; male or female) incorporate such shortenings into their tweets, we used the appropriate filters and compared the results. We also compared the frequency of letter/number homophones in tweets according to their technical and linguistic standardness[4].

Furthermore, the distributions of letter/number homophones in all JANES subcorpora were compared to that of the Kres corpus[5], a collection of standard written Slovene with a balanced genre structure (Logar Berginc et al., 2012).

In the second part of our study, the letter/number homophones found in Slovene tweets were analyzed in more detail. Among the most frequently used numerals in the shortenings discussed, we investigated:

- where they appear (at the beginning, in the middle, or at the end of a word);
- what they substitute (a string of letters, a single letter);
- whether they are used phonetically or graphically (*2morrow* vs. *g33k*).

4. Results

Surprisingly, the first query returned no results, which means that there are no words with numerals at the beginning of a word appearing in Slovene tweets represented in the JANES corpus. Thus, we used only

[1] The paragraph was translated into English by the author.

[2] Description available at http://nl.ijs.si/janes/ (access: June 12, 2016).

[3] Available at https://www.sketchengine.co.uk/ (access: June 10, 2016.

[4] All texts in the JANES corpus are annotated according to their level of standardness. “The score for technical text standardness focuses on word capitalisation, the use of punctuation, and the presence of typos or repeated characters in the words. The score for linguistic standardness, on the other hand, takes into account the knowledge of the language by the authors and their more or less conscious decisions to use non-standard language, involving spelling, lexis, morphology, and word order” (Ljubešič et al., 2015. Tweets are annotated “using a score between 1 (standard and 3 (very non-standard, with 2 marking slightly non-standard texts” (Ljubešič et al., 2015).

[5] Description available at http://www.slovenscina.eu/korpusi/kres (access: August 20, 2016).

the remaining two regular expressions for our further research.

4.1 Numeral at the End of the Word

The total number of concordances for all tokens ending in a numeral is 58.794. However, as mentioned before, words which represent a part of a proper name, a chemical symbol, a unit of measurement, etc., had to be manually selected and excluded from the list to get the actual result. The remaining tokens and their absolute frequencies are represented in Table 1.

Token	Frequency
ju3	1173
Mi2	593
mi2	371
Ju3	337
s5[6]	292
MI2	119
vi2	110
hi5	97
tr00	77
zju3	50
Hi5	47
na1	36
gr8	36
Tr00	31
Mi3	31
me2	27
str8	26
Vi2	20
Gr8	17
Me2	11
h8	11
u3	10
sk8	7
Zju3	5
mi3	5
H8	3
TR00	1

Table 1: Words with numerals at the end of the word.

As seen in the table above, 27 different tokens with 15 different lemmas[7] were found in the corpus of Slovene tweets, altogether representing a relative frequency of 33.1 per million tokens. Nine out of 27 tokens, which are written in bold, represent alternative written forms of different personal pronouns. The relatively high frequency can be explained by the fact that the numeral in the personal pronoun does not only have the same phonetic content as the letters (e.g., *s5* → /spet/, where 5 is pronounced as /pet/), but emphasizes the number of people a specific personal pronoun is denoting (*mi2* → two people; *mi3* → 3 people, etc.).

It is also interesting that 11 tokens are actually English homophones, frequently used in Slovene tweets as well.

4.2 Numeral in the Middle of the Word

Interestingly, the list of different letter/number homophones with numerals appearing in the middle of the word is significantly longer, whereas the relative frequency in much lower (9.97 per million tokens). This kind of shortening technique proves to be very productive, but still appears less frequently than the one with numerals at the end of the word. After excluding all the proper names and other irrelevant words, 117 tokens with approximately 50 different lemmas were found in Slovene tweets.

In some cases, it was difficult to identify whether we were dealing with a typographical error or whether the word was intentionally written in that form. We used the proximity of the letters and numbers on the keyboard to identify and exclude possible typographical errors (e.g., *v0lilcev* → number 0 and the letter "o" are very close on the keyboard, so we identified it as a typographical error vs. v8dja → probably intentionally written as such). As expected, the majority of homophones represent English words, where the preposition "to" is typically substituted by number 2 (e.g., *B2B*, *p2p*, *coffee2go*, *up2date*, etc.). There are, however, also numerous Slovenian words written with both alphabetic and numeric symbols. It is important to emphasize that we did not exclude the homophones which represent a phrase consisting of two or more words, e.g., *mi3je* = mi trije; *še1x* = še enkrat, etc. We decided not to consider them as typographical errors, but as a decision of Twitter users to write them as one word, similar to the English multi-word phrases listed above (*coffee2go*, *up2date*, etc.).

Token	Frequency
B2B/b2b	205/41
w00t/W00t	66/39
d00h/d0h/D0h/d000h	51/48/26/4
pr0n/Pr0n	49/6
g33k/g33ki/g33kov/g33ka/G33k	35/9/6/5/4
na1x	30
n00b/n00be	24/4
B2C	21
s3ksi/S3ksi	19/4
p2p/P2P	19/18
B4B	19
p0rn[8]	18

[6] Token *S5* was excluded from the list because it was almost exclusively used in the proper name *Galaxy S5*.

[7] Since the texts in the JANES corpus are normalized, the tokens *Ju3* and *ju3* would both have the same lemma – *jutri*.

[8] The motivation for writing specific words with numerals

Za1x	13
mi3je	12
še1x	11
ju3šnji/ ju3snji/ Ju3šnji/ ju3šnjega/ ju3snjem	11/4/4/3/3

Table 2: Most frequently used words with numerals in the middle of the word.

4.3 The Use of Alphanumeric Symbols According to User Gender

The comparison of frequency of letter/number homophones according to user gender shows that the use of words with alphanumeric symbols is far more frequent among male users (roughly 80% of all the occurrences were written by male users). However, we must take into account that the comparison was made using all the occurrences returned by the regular expressions, since we could not filter out all irrelevant examples.

4.4 The Use of Alphanumeric Symbols According to User Type

The comparison according to user type (corporate vs. private) shows a strong tendency of private users to incorporate such writing into their tweets. From 65.042 occurrences, 45.682 (≈ 70%) were written by private users.

4.5 The Use of Alphanumeric Symbols According to the Level of Text Standardness

Regarding the level of text standardness, an assumption can be made that less standard tweets (from both linguistic and technical perspective) contain more letter/number homophones than those written according to grammatical and orthographic rules. We compared all 9 possibilities of text standardness available in the JANES corpus (from L1T1 = linguistic 1, technical 1 to L3T3 = linguistic 3, technical 3 with all median possibilities). Words with alphabetic and numeric symbols are most frequently used in tweets annotated as very non-standard (L1T1) or linguistically very non-standard and technically slightly non-standard (L1T2).

4.6 Comparison with the Kres Corpus

In the figure bellow, relative frequencies of all letter/number homophones (regardless of the position of the numeral) in the JANES subcorpora and the Kres corpus are presented. As evident from the chart, letter/number homophones are far most frequent in Twitter posts (43.07 per million), followed by the forum posts (18.19 per million). Blog entries, comments on news, and wiki talk show a fairly similar distribution (2.61, 3.37, and 2.94 per million, respectively). Surprisingly, letter/number homophones can also be found in the Kres corpus (1.36 per million). A total of 12 different examples were found, 10 of them with numerals in the middle of the word (e.g., *l33t*, *cig4ni*, *za1x*, *pr0n*), one with numerals ending a word (*ju3*), and also one with numerals starting a word (*4ever*). However, all of these examples were found in the texts obtained from the web pages and from the Slovenian magazine *Joker*, which is primarily a computer gaming magazine with a distinctive writing style.

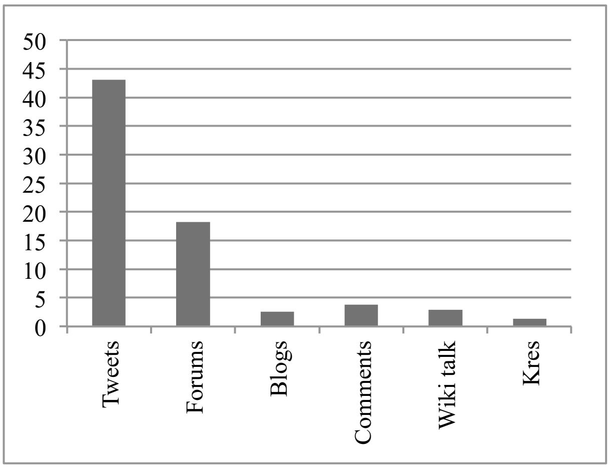

Figure 1: Relative frequencies of letter/number homophones in the JANES subcorpora and the Kres corpus.

5. Qualitative Analysis of Extracted Letter/Number Homophones

In this section, a qualitative analysis of the most frequently used letter/number homophones is presented, along with interpretations of the numeric features and their functions.

5.1 Most Frequent Numerals Used in Letter/Number Homophones

If we analyze the extracted words with alphanumeric symbols in more detail, it is evident that numerals are used only in the middle of the word (117 tokens) or at the end of a word (27 tokens). As mentioned before, no tokens with numerals starting a word were found in the corpus. Numerals used in letter/number homophones are definitely not randomly picked by the users. Each numeral has a specific meaning or interpretation and substitutes either a single letter or a string of letters in a word. In our corpus, 9 numerals were identified in letter/number homophones, i.e. 0, 1, 2, 3, 4, 5, 7, and 8. Among them, numerals 2, 3, 8, and 0 are the most frequent ones. Since the same numeral can appear in different words and even have different functions, it is interesting to identify which letter/letters are substituted

instead of letters could be to avoid censorship carried out by the moderators of forums, blogs, or comment sections. The same pattern appears in the word *cig4n* (= cigan) found in the Kres corpus.

by them. In Table 3, all numerals and their interpretations are presented, together with the examples from the corpus.

Numeral	Interpretation	Example
1	“ena” “i”	*na1* = “na ena” *BRA71L* = “Brazil”
2	”dva” “dve” “to”	*mi2* = “midva” *me2* = “medve” *up2date* = “up to date”
3	“tri” “e”	*ju3 = “jutri”* *s3njam = “strinjam”* *g33k = “geek”*
4	“for” “a”	*t4t = “training for trainers”* *G4ME = “game”*
5	“pet” “five”	*s5 = “spet”* *hi5 = “high five”*
7	“z”	*BRA71L = “Brazil”*
8	“eat” “aight” “ate”	*gr8 = “great”* *str8 = “straight”* *h8 = “hate”* *l8r = “later”*
0	“o”	*n00b = “noob”* *p0rn = “porn”* *w00p = “woop”*

Table 3: Numerals used in letter/number homophones with their interpretations and examples.

5.2 Phonetic vs. Graphic Function of Numerals

As evident from the table above, there is a striking difference between numerals that are used phonetically (e.g. *s5* = “spet”, where their pronunciation is identical with a part of the word, enabling them to replace the letter sequence) and graphically (e.g. G4ME = “game”, where the numeral 4 has a similar visual appearance as the letter “A”). All numerals used at the end of the words (see Table 1) are used phonetically; the only exception is the word *tr00* (= “true”). This example is especially interesting because numerals do not only substitute a string of letters (*00* = “oo”), but the pronunciation of the substituted letters is similar or the same as the original pronunciation of the word string (“ue” in *true*). In other words, 3 transformations are needed to identify the “original” word, namely *tr00* → troo → /truː/ → “true”. Numerals which appear in the middle of the word can be used either graphically or phonetically. Numerals which substitute letters based on their appearance rather than their pronunciation tend to be duplicated (e.g. *g33k*, *w00p*, *n00b*, etc.)[9].

According to that, we can undoubtedly claim that letter/number homophones are not only used as a word-shortening strategy, but as a form of creative writing and a specific stylistic feature as well. Furthermore, graphically used numerals also prove that CMC is “essentially a mixed modality” which “resembles speech” but “looks like writing” (Baron, 2008), since the words, such as *g33k*, *tr00*, or *n00b*, have little significance if not visually represented.

6. Conclusion

This paper presents the use of so-called letter/number homophones in Slovene tweets as presented in the JANES corpus. The results show that a considerable amount of alphabetic and numeric symbols are used both in Slovene and in English words. This phenomenon, also described as a type of “neography” (Danet and Herring, 2007), proved to be characteristic for CMC, especially microtexts, such as Twitter and forum posts, since no letter/number homophones were found in the Kres corpus apart from the texts obtained from web pages. As expected, Twitter proved to be the richest subcorpus regarding this phenomenon, followed by the forum subcorpus with a relatively high frequency of the shortenings discussed. Numerals used in the middle or at the end of specific words substitute letters or strings of letters and make the texts either shorter or more interesting to read. Since a certain numeral can be used both graphically (*g33k*) and phonetically (*u3nek*), a creative writing style has emerged among new generations, which definitely deserves linguistic attention. For a more precise description of the phenomenon, a more detailed comparison with other CMC media (SMS, blogs, forums, etc.) would be necessary. Apart from that, an analysis of usernames would be very useful for investigating language creativity as observed in computer-mediated communication.

7. Acknowledgements

The work described in this paper was funded by the Slovenian Research Agency within the national basic research project “Resources, Tools and Methods for the Research of Nonstandard Internet Slovene” (J6-6842, 2014-2017).

8. References

Alkawas, S. (2011). *Textisms: The Pragmatic Evolution among Students in Lebanon and its Effect on English Essay Writing*. Master Thesis, Lebanese American University.

Baron, S. (2008). *Always On: Language in an Online and Mobile World*. Oxford University Press, Oxford.

Bieswanger, M. (2006). 2 abbrevi8 or not 2 abbrevi8: A contrastive analysis of different space-and time-saving strategies in English and German text messages. In Hallett, T., Floyd, S., Oshima, S. and Shield, A. (Eds.), *Texas Linguistics Forum Vol. 50*, Austin.

[9] This type of writing is also refered to as “l33t speak” or “l33t”, used mostly by players of video games “where numbers and symbol combinations are used to represent letters” (Sherblom-Woodward, 2002).

http://studentorgs.utexas.edu/salsa/proceedings/2006/Bieswanger.pdf

Bushnell, C., Kemp, N. and Heritage Martin, F. (2011. Text-messaging practices and links to general spelling skills: A study of Australian children. In *Australian Journal of Educational & Developmental Psychology*. Vol 11, pp. 27–38.

Crystal, D. (2001). *Language and the Internet*. Cambridge, Cambridge University Press.

Danet, B. and Herring, S. (Eds.). (2007). *The Multilingual Internet. Language, Culture, and Communication Online.* Oxford University Press, Oxford.

Denby, L. (2010). *The Language of Twitter: Linguistic Innovation and Character Limitation in Short Messaging.* Undergraduate dissertation, University of Leeds.

Dobrovoljc, H. (2008). Jezik v e-poštnih sporočilih in vprašanja sodobne normativistike. In Košuta, M. (Ed.), *Slovenščina med kulturami*, Slavistično društvo Slovenije, Celovec, pp. 295–314.

Elizondo, J. (2011). *Not 2 Cryptic 2 DCode: Paralinguistic Restitution, Deletion, and Non-standard Orthography in Text Messages*. Ph.D. thesis, Swarthmore College.

Farina, F. and Lyddy, F. (2011). The Language of Text Messaging: “Linguistic Ruin” or Recource? In *The Irish Psychologist*, Vol. 37, Issue 6, pp. 145–149.

Filipan-Žignić, B., Velički, D. and Sobo, K. (2012). SMS communication – Croatian SMS language features as compared with those in German and English Speaking Countries. In *Revija za elementarno izobraževanje*, št. 1. Pedagoška fakulteta, Maribor.

Fišer, D., Erjavec, T. and Ljubešić, N. (2016). Janes v0.4: korpus slovenskih spletnih uporabniških vsebin. *Slovenščina 2.0* (to appear).

Frehner, C. (2008). *Email, SMS, MMS: The Linguistic Creativity of Asynchronous Discourse in the New Media Age*. Peter Lang.

Gouws, S., Metzler, D., Cai, C. and Hovy, C. (2011). Contextual bearing on linguistic variation in social media. In *Proceedings of the workshop on language in social media* (*LSM 2011*), pp. 20–29. http://aclweb.org/anthology/W/W11/W11-0704.pdf.

Grace, A., Kemp, N., Martin, F. H. and Parrila, R. (2012). Undergraduates’ use of text messaging language: Effects of country and collection method. In *Writing Systems Research.* Taylor & Francis Online.

Halmetoja, T. (2013). *Gender-Reated Variation in CMC Language: A Study of Three Linguistic Features on Twitter*. BA thesis, Göteborgs Universitet.

Kadir, Z. A., Maros, M. and Hamid, B. A. (2012). Linguistic Features in the Online Discussion Forums. In *International Journal of Social Science and Humanity*, Vol. 2, No. 3, May 2012, pp. 276–281.

Kirsten Torrado, U. (2014). Development of SMS language from 2000 to 2010. In Cougnon, L. and Fairon, C. (Eds.), *SMS communication: A linguistic approach*. Benjamins Current Topics.

Kul, M. (2007). Phonology in text messages. In *Poznań Studies in Contemporary Linguistics 43(2)*, pp. 43–57.

Ljubešić, N., Fišer, D., Erjavec, T., Čibej, J., Marko, D., Pollak, S. and Škrjanec, I. (2015). Predicting the level of text standardness in user-generated content. In *Proceedings*, pp. 371–378, Hissar: [s.n.]. http://lml.bas.bg/ranlp2015/docs/RANLP_main.pdf.

Logar Berginc, N., Grčar, M., Brakus, M., Erjavec, T., Arhar Holdt, Š. and Krek, S. (2012). *Korpusi slovenskega jezika Gigafida, KRES, ccGigafida in ccKRES: gradnja, vsebina, uporaba*. Ljubljana: Trojina, zavod za uporabno slovenistiko; Fakulteta za družbene vede.

Logar, N. and Smith, J. (trans.). (2006). Stilno zaznamovane nove tvorjenke: tipologija = Stylistically marked new derivates: a typology. In Vidovič-Muha, A. (Ed.), *Slovensko jezikoslovje danes*, Slavistično društvo Slovenije, Ljubljana, pp. 87–101.

Michelizza, M. (2008). Jezik SMS-jev in SMS-komunikacija. In *Jezikoslovni zapiski: zbornik Inštituta za slovenski jezik Frana Ramovša*, Inštitut za slovenski jezik Frana Ramovša ZRC SAZU, Ljubljana, pp. 151–166.

Moseley, N. (2013). *Using word and phrase abbreviation patterns to extract age from Twitter microtexts*. Thesis, Rochester Institute of Technology.

Sherblom-Woodward, B. (2002). *Hackers, Gamers and Lamers: The Use of l33t in the Computer Sub-Culture.* http://www.swarthmore.edu/SocSci/Linguistics/papers /2003/sherblom - woodward.pdf.

Thurlow, C. (2003). Generation Txt? The sociolinguistics of young people's text-messaging. In *Discourse Analysis Online*, Sheffield.