

# Dynamics of an Idiostyle of a Russian Suicidal Blogger

Tatiana A. Litvinova<sup>1</sup> Olga A. Litvinova<sup>1</sup> Pavel V. Seredin<sup>1,2</sup>

<sup>1</sup>RusProfiling Lab,  
Voronezh State Pedagogical University  
86 Lenina ul., Voronezh, 394043, Russia  
centr\_rus\_yaz@mail.ru

<sup>2</sup>Department of Solid State Physics and Nanostructures,  
Voronezh State University  
1 Universitetskaya pl., Voronezh, 394018, Russia  
paul@phys.vsu.ru

## Abstract

Over 800000 people die of suicide each year. It is estimated that by the year 2020, this figure will have increased to 1.5 million. It is considered to be one of the major causes of mortality during adolescence. Thus there is a growing need for methods of identifying suicidal individuals. Language analysis is known to be a valuable psychodiagnostic tool, however the material for such an analysis is not easy to obtain. Currently as the Internet communications are developing, there is an opportunity to study texts of suicidal individuals. Such an analysis can provide a useful insight into the peculiarities of suicidal thinking, which can be used to further develop methods for diagnosing the risk of suicidal behavior. The paper analyzes the dynamics of a number of linguistic parameters of an idiostyle of a Russian-language blogger who died by suicide. For the first time such an analysis has been conducted using the material of Russian online texts. For text processing, the LIWC program is used. A correlation analysis was performed to identify the relationship between LIWC variables and number of days prior to suicide. Data visualization, as well as comparison with the results of related studies was performed.

## 1 Introduction

The development of Internet communication has paved the way for extensive studies into the reflection of personality traits, mental state, moods, and emotions in writing. One of the characteristic features of recent studies of the issue has been the collaborations of computational linguists and psychologists. A distinctive example of such an interaction is Computational Linguistics and Clinical Psychology Workshop held annually since 2014 and aimed at bringing together “computational linguistics researchers with clinicians to talk about the ways that language technology can be used to improve mental and neurological

health” (<http://clpsych.org>). One of the important problems in the field is to develop methods of identifying individuals with high suicide risks based on the analysis of their written texts including online texts, i.e. forums (Desmet and Hoste, 2018), tweets (Burnap et al., 2015; Fodeh et al., 2017), blogs (Guan et al., 2015) etc. The main idea of such work is to use automatic text classification to detect suicide-related content (see Gomez, 2014 for review).

There is no doubt as to its significance, however most studies rely on manual annotation of training material from the point of view of estimating suicidal behavior risks of authors of texts. However, as was rightfully pointed out by Homan et al. (2014), “the mental state of another individual, observed from a few lines of text often written in an informal register is necessarily hard to discern and, even under less noisy conditions, extremely subjective ... This makes annotation quite a challenge, and does not reveal in an objective fashion a tweeter’s true mental state” (p. 114).

One of the promising areas of research is analysis of social media texts by people who publicly stated that they have tried to take their own life (Wood et al. 2014; Coppersmith et al., 2016). However, it is questionable if it is possible to generalize obtained findings regarding behavior of suicide attempters to the completers (DeJong et al., 2010).

It also should be noted that only limited number of works in this booming line of language-related suicide risk detection consider dynamics of language variables and/or mental state of individuals. For example, Choudhury et al. (2016) proposed methodology to infer which individuals could undergo transitions from mental health discourse to

suicidal ideation. The authors showed a number of markers characterizing these shifts including social engagement, manifestation of hopelessness, anxiety and impulsiveness based on a small subset of Reddit posts. Coppersmith et al. (2016) examined data from Twitter users who have attempted to take their life and provide an exploratory analysis of patterns in language and emotions prior to their attempt. One of the interesting results found in this study is the increase in the percentage of tweets expressing sadness in the weeks prior to a suicide attempt, which is then followed by a noticeable increase in anger and sadness emotions the week following a suicide attempt.

It should be emphasized that most research in language-based suicide risk detection has employed English language materials with texts in other languages not being sufficiently addressed, with few exceptions (Desmet and Hoste, 2014; Guan et al., 2015; Litvinova et al., 2017).

Corbitt-Hall et al. (2016) analyzed Facebook users' (namely college students) abilities to notice, recognize, and appropriately interpret suicidal content and their willingness to intervene and found out that college students are responsive to suicidal content on Facebook. It is obvious that it is viable to get new insights into the language of suiciders and share this knowledge with a wider audience of social media users in order to facilitate suicide prevention for different language and cultures.

In order to develop methods of evaluating suicidal risks based on linguistic analysis, it is extremely important to analyze texts by people who died by suicide. However, such an analysis is made more complicated due to limited access to relevant data. Texts of suicide notes have long been employed in corresponding studies as well as literary texts by individuals who died by suicide (Baddeley et al., 2011; Stirman and Pennebaker, 2001). However, as stated by Litvinova et al. (2017), "there are certain restrictions associated with the nature of texts and their authors' personalities, which prevents the results from being extrapolated into the entire population" (p. 247). However, the development of Internet communications (publicly accessible blogs, tweets or Facebook) resulted in the fact that scholars have been able to access very valuable linguistic data containing texts by individuals who died by suicide as well as new data sources for the study of suicidal behavior.

Texts of blogs as a prevalent form of communication in expressing emotion and sharing information are particularly significant. However, studies of online texts by individuals who died by suicide are still very limited in number (Li et al., 2014). Besides, the dynamics of linguistic parameters as the author's death approached has not been sufficiently investigated while the analysis of the dynamics of an idiosyncrasy would allow a more profound insight into a psychological state of a suicidal individual resulting in the development of diagnostic tools.

All of the above were the prerequisite for the objective of the paper which is to investigate the dynamics of linguistic parameters of a Russian-language blog of a software engineer from Moscow, the creator of the website *mysuicide.ru*, one of Russia's largest suicide websites, who died by suicide at the age of 30, in order to attempt to sketch the suicidal process. To be consistent with a unified classification method, the language patterns of the blog were analyzed using the Russian version of the Linguistic Inquiry and Word Count (LIWC) program (Pennebaker, 2007), a text analysis software program that provides over 80 psychologically meaningful language variables, such as emotion and self-referencing words.

## 2 Material and methods

### 2.1 Material

The material of the study were *LiveJournal* blogs by the user *light\_medelis* (<http://light-medelis.livejournal.com/>) The user also had a name *lm\_diary* (<http://lm-diary.livejournal.com/>) The accounts belong to Sergey Makarov, the creator of the website *mysuicide.ru*, one of Russian Internet's (Runet) largest websites, containing suicide-related content. Blog entries used as a data source for this study are publicly available. These blog entries are extracted from the corpus of Russian texts *RusSuiCorpus*<sup>1</sup> which consists of the blogs written by individuals who died by suicide. It currently contains texts by 45 Russian individuals aged from 14 to 30. The total volume of the corpus is about 200 000 words. All the texts are manually collected from publicly available source and represent blog posts by individuals who died by suicide (blogs from *LiveJournal*) (Litvinova, 2016). The fact that suicides had actually took

<sup>1</sup> Currently the corpus is available by request at [centr\\_rus\\_yaz@mail.ru](mailto:centr_rus_yaz@mail.ru)

place was checked by analyzing friends' comments, media texts, etc.

Sergey died by suicide on December 12, 2005, which became known based on his friends' comments on *LiveJournal* and media. The website *mysuicide.ru* was shut down after its creator as well as a few other regular visitors died by suicide. The events got a wide media coverage.

We took a look at S. Makarov's two blogs as they were both on different topics. The blog *lm\_diary* is more personal and looks like a personal diary as the author describes his feelings and suffering (for further reference it will be called **PD1**). The blog *light\_medelis* dealt with discussion of suicide-related content, depression, etc. (for further reference it will be called **PD2**). Both blogs were updated almost up to the day of the author's death, but PD1 was being updated from July 28, 2004 till December 11, 2005, PD2 from June 13, 2003 till December 11, 2005. For a correct comparison of the obtained data we chose the texts written over the same time period, i.e. the PD2 entries starting from July 28, 2004 were analyzed. All the author's texts (blog entries as well as author's comments) written on the same day were entered into the same file named according to the entry date. That was done separately for each PD1 and PD2. The texts not written by the author (citations, including "hidden" ones, for example, news without quotes, links, etc.) were removed manually.

## 2.2 Methods

The texts were processed using the LIWC2007 software with Russian dictionary (Kailer and Chung, 2011). Apart from a standard dictionary, we developed a set of our own ("users") dictionaries in accordance with LIWC2007 manual:

- a dictionary of demonstrative pronouns and adverbs - *Deictic*,
- a dictionary of intensifiers and downtoners - *Intens*,
- a dictionary of perception vocabulary - *PerceptLex*,
- a dictionary of pronouns and adverbs describing the speaker (self-references) - *Ego*,
- a dictionary of emotional words - *Emo* (negative and positive);
- a dictionary of pronouns with subcategories (personal, indicative etc.) - *Pronouns*;
- a dictionary of Russian most frequent words - *Freq.*, etc.

The users' dictionaries were compiled using available dictionaries and Russian thesauri. As a

Russian dictionary that came with the software was a translation of a corresponding English dictionary, we have to check it manually and make some corrections.

The values of 142 text parameters were extracted. Further we chose the frequency parameters, i.e. those differing from zero in more than 50 % of the texts (in both blogs). At this point of the analysis the number of the text parameters went down to 66. Pearson's correlation analysis was then carried out to identify the correlation between each of the chosen LIWC variables and the number of days prior to the death.

## 3 Results and discussion

As a result of correlation analysis, 8 out of the chosen text parameters (LIWC variables) correlated with the number of days prior to the death in PD1:

- common verbs;
- personal pronouns;
- the overall pronouns;
- words describing social processes (*mate, talk, they, child*);
- prepositions;
- preposition 'with';
- numerals;
- pronoun 'I'.

As for PD2, 9 out of the chosen text parameters correlated with the number of days prior to the death:

- the percentage of words describing the writer ("I", "my" and its forms; the expression "in my opinion", etc.) ("Ego");
- words describing affect (*happy, cried, abandon*);
- the conjunction "and";
- personal pronouns;
- the overall pronouns;
- words describing positive emotions;
- conjunctions;
- words describing achievements (earn, hero, win);
- pronoun 'I'.

All the correlations are positive (with Pearson's  $r$  0.2-0.3,  $p < 0.05$ ), i.e. as the date of the death approached, the values of the above parameters drop. In both types of blogs there is a dependence between the number of days prior to the death and the proportions of personal pronouns, overall pronouns, "I" pronouns, words describing positive emotions.

As we can see, a considerable part of the correlations is made up of the parameters associated with the frequency of pronouns. The significance of the analysis of pronouns in written documents as an unobtrusive way of assessing underlying psychological processes has been described a lot (Tausczik and Pennebaker, 2010).

Note that in the study by Litvinova et al. (2017) using the material of RusSuiCorpus it was shown that Russian online texts by suicidal individuals contain more function words, verbs, conjunctions, cognitive words, commas, fewer prepositions, comparison words and pronouns compared to the texts by the control group (with no consideration of the time factor). These texts appear to be more abstract and contain fewer spatial references. Texts by suicidal individuals were also found to contain more words for negative emotions and fewer of those describing social relations and perception (particularly visual), which is indicative of these people being more preoccupied with their own thoughts and isolated from the outside world. As we can see from the example of an individual whose texts are part of the corpus, some of the above parameters also correlate with the number of days prior to the death.

For a detailed analysis of the behavior of the chosen text parameters the data was visualized. We designed the dependencies of the intensity of posting (in terms of the number of words per day) for both blogs on the number of days prior to the death in the same graph (Fig. 1).

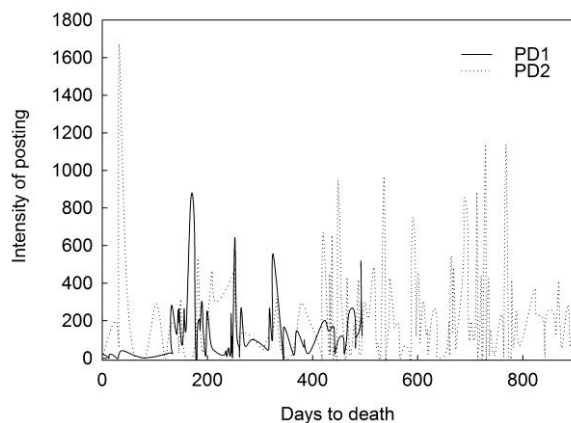


Fig. 1. Graph of the dependency of the intensity of posting (in words) on the number of days prior to the death for both blogs

As can be clearly seen from the experimental data presented in Fig. 1, several periods of peaks and drops in the intensity of posting are typical of both blogs. At certain points the intensity is identical for both blogs. For a further analysis we

chose five periods when there is a peak in the intensity for both blogs at a time. We then calculated the average values of the above text parameters at the specified peaks. The obtained results are presented graphically (Fig. 2-9) with the averaged values of a text parameter in the analyzed periods along with the standardized dependence of the intensity of posting (for PD1 and PD2). To build the dependencies, we have performed min-max normalization of the intensity of posting in the chosen periods (number of words per day).

Let us take a closer look at some of the parameters that were commonly used for other languages in studies of the dynamics of the parameters of a suicidal individual's idiosyncrasy using the LIWC software. In these studies (see the review of the results in paper by Li et al., 2014) the researchers relied on the existing conceptions of suicidogenesis according to which a suicide is associated with a growing social isolation (the sociological concept), feeling of hopelessness, sadness, and despair (the psychological conceptions of suicide). Therefore a special attention is paid to the analysis of the frequency of the pronouns "I" and "we", words describing social processes; the number of words describing positive and negative emotions.

In some studies it was shown that as the date of the suicide approaches, the frequency of the pronouns "I" increases while the number of the pronouns "we" decreases; there are fewer words describing social processes as well as positive emotions and more words describing negative emotions. However, in some other studies the results were the opposite (Li et al., 2014).

Since the parameters "Percentage of Words Describing the Writer (self-references)" and "Percentage of the Pronouns "I"" are closely related, we are considering them together (Fig. 2-3).

In the personal diary PD1 the percentage of the words of the above category is consistently high at the peak periods, but during the last period the number of such words drops significantly as well as the intensity.

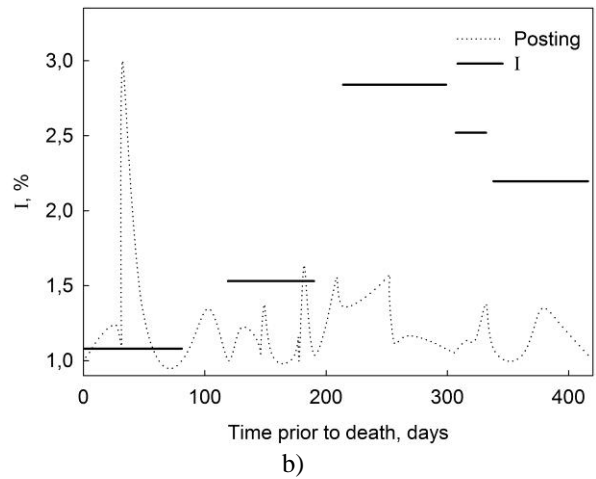
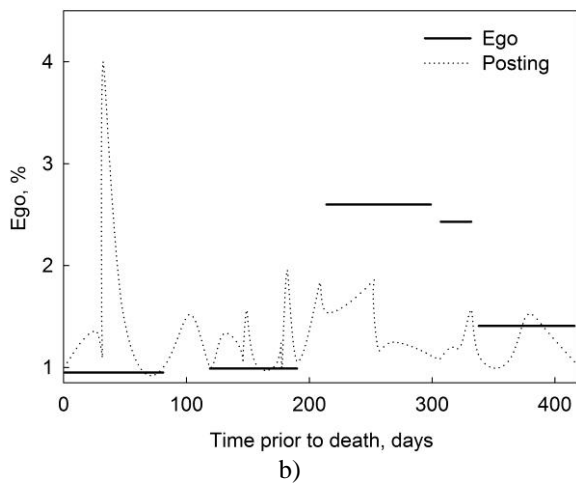
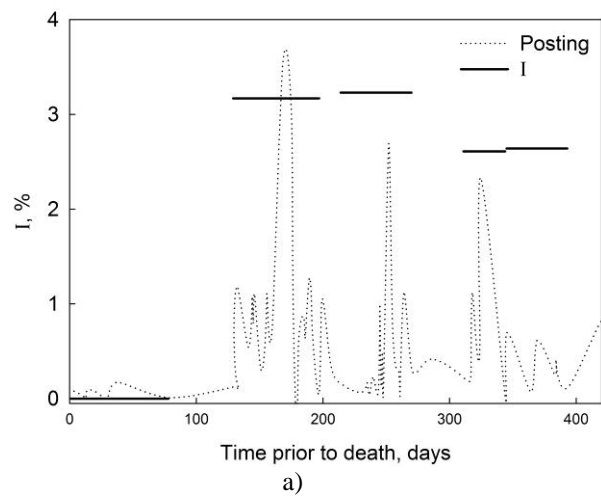
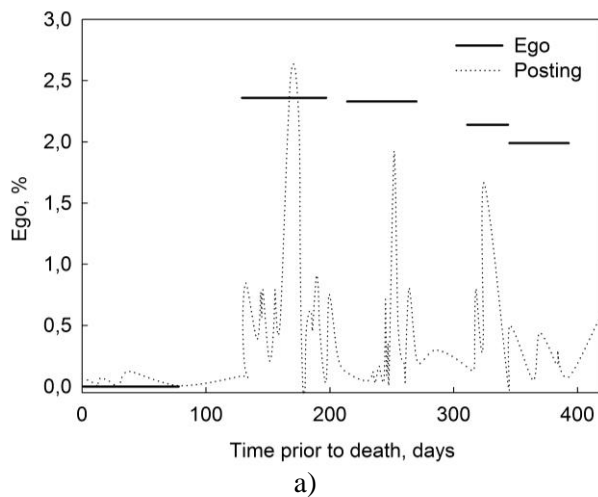


Fig. 2. Graphs of changes in the parameter “Percentage of Words Describing the Writer (“Ego”): a – PD1, b – PD2

However, in the texts in PD2 despite a peak during the last period there is also a drop in the frequency of linguistic units that describe the author, which does not agree with the results showed in some studies using literary texts but is consistent with the results obtained in paper by Li et al. (2014) where the methodology and material (blog texts were examined over a year prior to the author’s death) are most similar to those we chose to employ. When we analyzed texts we have noticed an increasing use of impersonal sentences describing writer feelings and states in this period, but this fact needs further investigation.

Fig. 3. Graphs of changes in the parameter “Percentage of the Pronouns “I””: a – PD1, b – PD2

The results of the analysis of the behavior of the parameter “Percentage of Words Describing Social processes” (Fig. 4) in the texts we have analyzed are in good agreement with those obtained in other studies: immediately prior to the death the proportion of such words in texts drops, which is consistent with the sociological conception of suicidogenesis (Stirman and Pennebaker, 2001; see also Choudhury et al., 2016, for similar finding in reduced social engagement as a marker of shift to suicidal ideation).

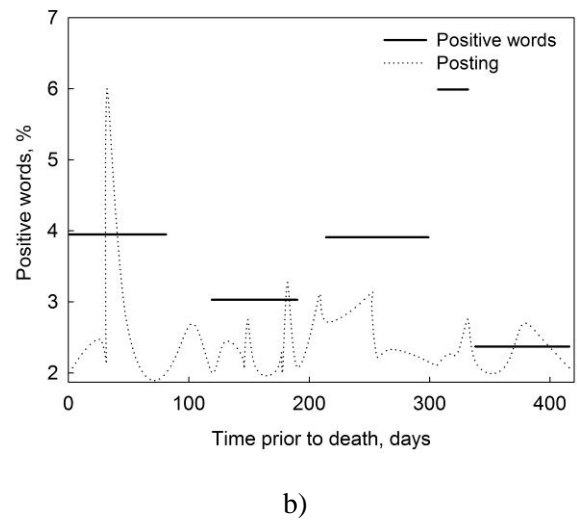
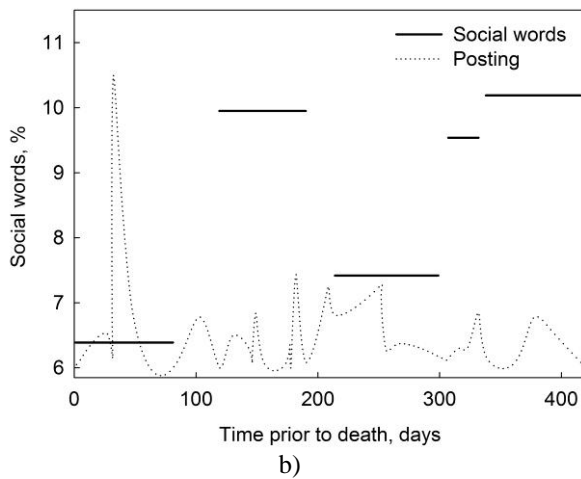
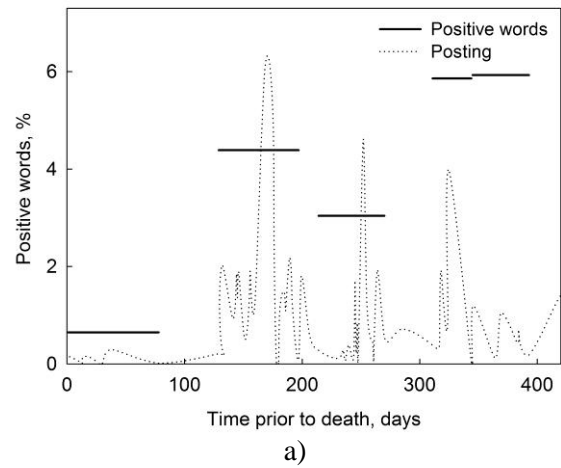
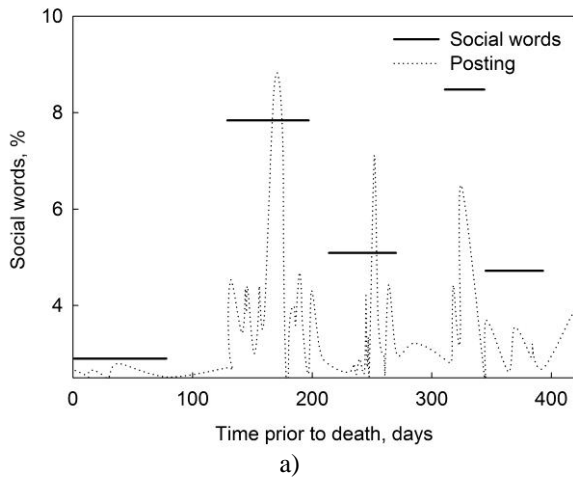


Fig. 4. Graphs of changes in the parameter “Percentage of Words Describing Social processes”: a – PD1, b – PD2

Analyzing words describing emotions is an essential part of studying texts by suicidal individuals (Fig. 5).

As our analysis showed no correlations between the percentage of words describing negative emotions in a text and the number of days prior to the death, only the behavior of the parameter “Percentage of Words Describing Positive Emotions” was visualized.

In the personal diary PD1 the percentage of words describing positive emotions drops as so does the intensity of posting. In the texts in PD2, however, in the last period the percentage of words of the above group rises as so does the intensity of posting.

Fig. 5. Graphs of changes in the parameter “Percentage of Words Describing Positive Emotions”: a – PD1, b – PD2

An increase in the proportion of words describing positive emotions in the period prior to suicide was identified in 4 out of 9 studies analyzing the writing of suicidal individuals using LIWC (Li et al., 2014), which may be associated with an improvement in the author’s psychological state following the decision to die.

Let us examine the dynamics of some other parameters that have not been dealt with in studies of changes in an idiosyncrasy of suicidal individuals. In both blogs we can see a drop in the number of verbs in the time in the run up to the suicide (Fig. 6) as well as the number of personal pronouns (Fig. 7).

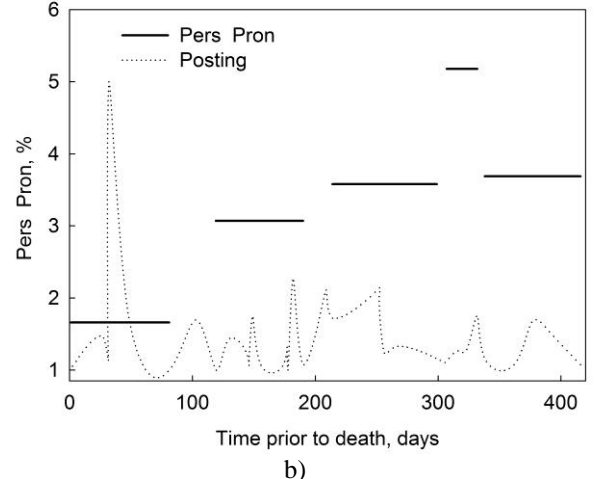
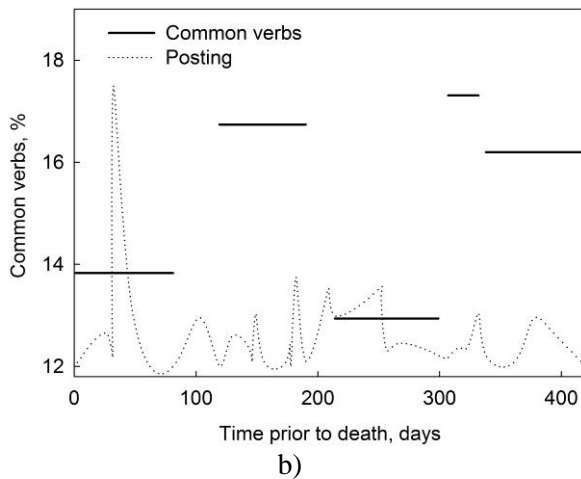
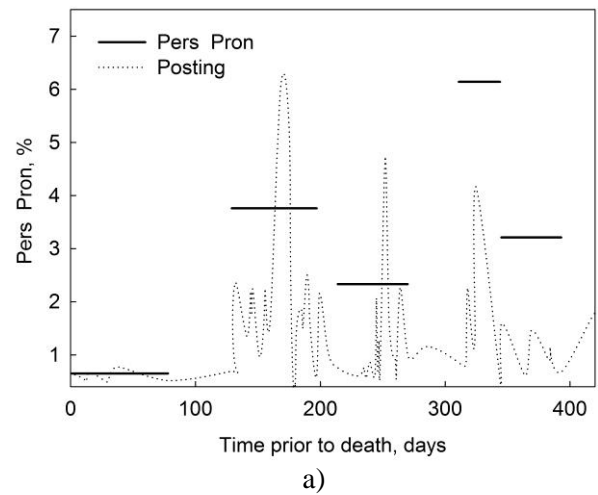
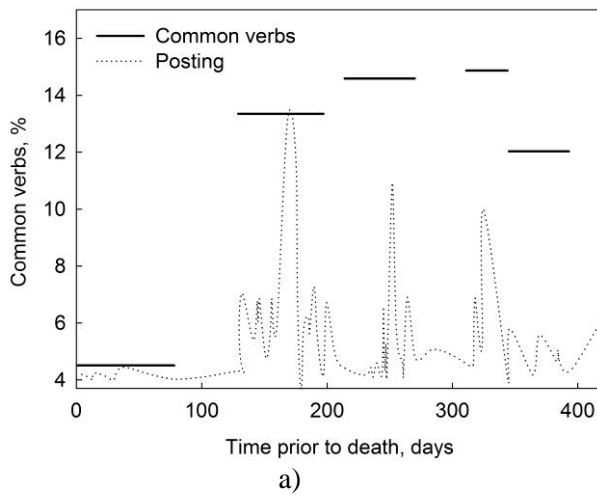


Fig. 6. Graphs of changes in the parameter “Common verbs”: a – PD1, b – PD2

Fig. 7. Graphs of changes in the parameter “Personal pronouns”: a – PD1, b – PD2

Let us look at the dynamics of such parameters as the proportion of conjunctions (Fig. 8) and prepositions (Fig. 9).

As can be seen, the behavior of the category “Conjunctions” was different in the two diaries. While in PD2 the number of conjunctions was dropping in the time in the run up to the suicide, in contrast, in PD1, as the analysis suggests, it was on the rise mainly due to a high frequency of the conjunction “and”.

The proportion of prepositions was dropping in the last period on both diaries. As was already noted, in the study comparing blogs of suicidal individuals and texts by the control group (Litvinova et al., 2017), it was found that on average texts by the former contain more function words in total, verbs, conjunctions but fewer prepositions. It is of interest that as was shown in (Litvinova et al., 2016) using texts by healthy individuals (students who had done psychological tests), overall for texts by individuals with high risks of autoaggressive behavior (according to the results of psychological tests), a lower lexical diversity, fewer prepositions, more pronouns overall, particularly personal ones with a higher index of logical cohesion (created due to more conjunctions) are typical.

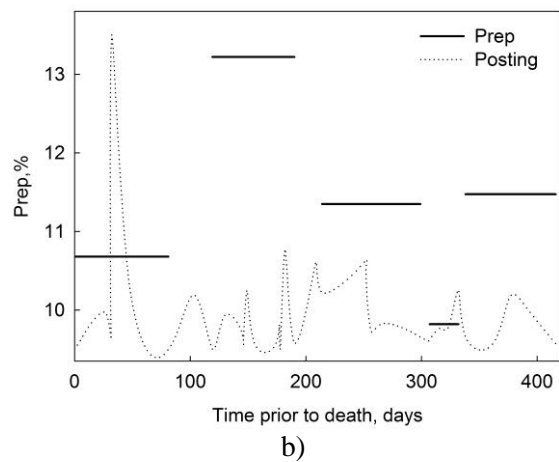
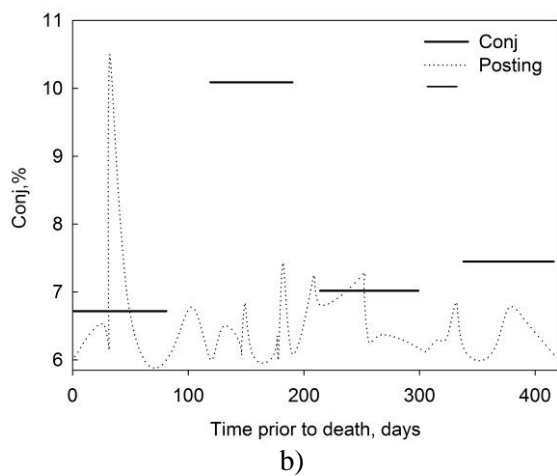
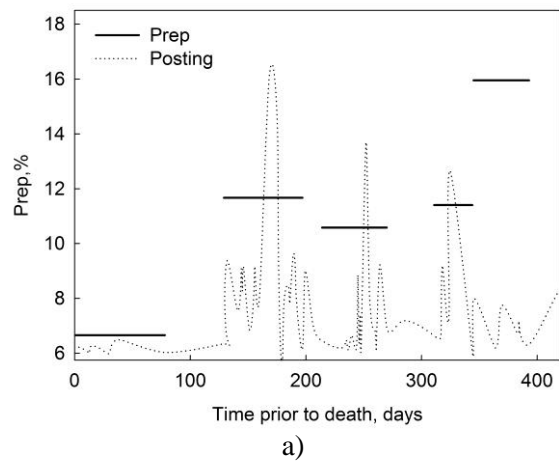
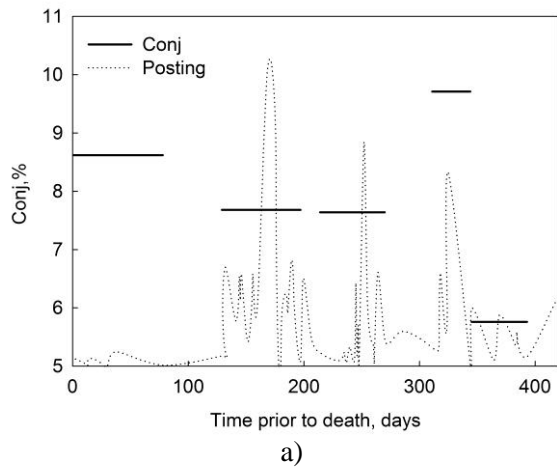


Fig. 8. Graphs of changes in the parameter “Conjunctions”: a – PD1, b – PD2

Fig. 9. Graphs of changes in the parameter “Prepositions”: a – PD1, b – PD2

In this study a neuropsycholinguistic interpretation of the data is set forth. Therefore the analysis of conjunctions and prepositions in their dynamics is seen as essential for further studies of the dynamics of an idiostyle of suicidal individuals. Hence it was found that in blogs by the suicidal individual in the time in the run up to the suicide there are fewer self-references, words describing social interactions, verbs, prepositions, but (in one of the diaries) there is a stable high number of conjunctions (mostly the conjunction “and”) as well as words describing positive emotions.

We assume that the above indicates that there is a drop in the suicidal individual’s activity (a reduction in the proportions of self-references, verbs), growing isolation from the world (a reduction in the proportion of deictic elements – prepositions and pronouns) in the time immediately prior to the suicide.

Note that the above changes occur in the time of around three months prior to the suicide. There is a clear indication that the final decision had already been made. It is also worth noting that in this period the depression symptoms got more severe and the antidepressants that were being taken seemed to be working less.

#### 4 Limitations

As any case study, this work has a number of limitations. We only analyzed blogs of one person who suffered from depression and wrote a lot about his mental health and willingness to die by suicide. It is essential in future work to make comparison of his writing to the blogs by people who did not die by suicide and to the blogs by people who died by suicide but never discussed their plans concerning suicide. This could highlight some universal linguistic patterns of dynamics of idiostyle of suiciders.



## 5 Conclusions and future work

Our study extends the findings of psycholinguistic analysis of suicides to the online document form. Besides, this study analyzed Russian material, whereas most previous studies have only analyzed English material or material from other languages translated into English before analysis.

A unique aspect of the current study is that we used blog entries that were written in Russian and were analyzed by means of the Russian version of the LIWC. The results of our study that are certainly preliminary have proved that it is viable to use software, particularly LIWC with a Russian dictionaries, for processing a large massive of texts in order to identify stable and varying characteristics of idiostyle with respect to topic dimension. However, it will be necessary to verify and expand internal Russian dictionary and to create special dictionaries for suicide-related studies as it was done for Chinese (Lv et al., 2015). In addition, we are planning to extend the list of linguistic parameters and add linguistic complexity, syntactic parameters, etc. in particular.

We argue that it would be rational to perform multivariate analysis to reveal how different linguistic parameters best predict time course of suicide.

Based on the results of the data visualization, changes in the chosen text parameters are generally nonlinear. Therefore, while analyzing the dynamics of a suicidal individual's idiostyle, it is not sufficient to choose text parameters using only a correlation data analysis that involves searching for linear connections without visualizing the behavior of the text parameters over different periods. The contradictory results obtained in the existing research dealing with the character of the dynamics of linguistic parameters of texts by suicidal individuals, among other things, might be due to not enough attention being given to the behavior of each parameter at different periods.

In addition, the above contradictions might be accounted for by the fact that in the existing studies texts of different genres and mostly literary works are analyzed. As our study suggests, the differences in the behavior of text parameters might emerge even in an Internet blog that can obviously be represented by different subgenres. Besides, the above differences in the results of the study might be due to the fact that literary texts are mostly employed that were written over a long period of time and a character of changes in the text parameters might be affected by age as well. Thus the behavior of the pa-

rameters of texts by different authors written over the same time period, e.g., a year prior to the death, should be investigated in future studies. It also seems promising to seek to identify the correlations between the text parameters and the ordinary number of a text (entry), but not only the number of days prior to the death as we have done in the present study as changes in the behavior of linguistic parameters might be not only due to those in the author's state but also with some events in their lives that affect the intensity of posting.

Despite the above difficulties, the study indicates that it is searching for tendencies and analyzing the dynamics of the behavior of the text parameters that allows a more profound insight into the cognitive characteristics of suicidal individuals and a further development of predictive models of assessment of suicide risks based on a linguistic analysis employed for online texts as well. Studying such texts using modern methods of NLP and data mining would allow one to develop a new set of tools for identifying individuals with suicidal behavior tendencies. This could be instrumental for practicing psychologists in their daily work resulting in a screening system for monitoring publicly available messages on social media as well as to identify individuals with high risks of suicidal behavior.

## Acknowledgments

Funding of the projects N MK-4633.2016.6 "Predicting Suicidal Tendencies of Individuals Based on Their Speech Production", N MK-5718.2018.6 "Speech portrait of the extremist: corpus-statistical research (on the material of the extremist forum "Kavkazchat") from the RF President's grants for young scientists is gratefully acknowledged.

## References

- A. Kailer and Cindy K. Chung, 2011. *The Russian LIWC2007 dictionary*. Austin, TX: LIWC.net.
- Anthony Wood, Jessica Shiffman, Ryan Leary, Glen Coppersmith, 2016. Language signals preceding suicide attempts. In *CHI 2016 Computing and Mental Health Workshop*, San Jose, CA.
- Bart Desmet and Veronique Hoste, 2014. Recognising Suicidal Messages in Dutch Social Media, in *Proceedings of LREC 2014 – Ninth international conference on language resources and evaluation*, pages 830–835.
- Bart Desmet and Véronique Hoste, 2018. *Online suicide prevention through optimised text classifica-*

- tion, *Information Sciences*, 439–440: 61–78. <https://doi.org/10.1016/j.ins.2018.02.014>
- Christopher M. Homan, Ravdeep Johar, Tong Liu, Megan Lytle, Vincent Silenzio, Cecilia O. Alm, 2014. Toward macro-insights for suicide prevention: Analyzing fine-grained distress at scale, in *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology*, pages 107–117.
- Darcy J. Corbitt-Hall, Jami M. Gauthier, Margaret Taylor Davis and Tracy K. Witte, 2016. *College Students' Responses to Suicidal Content on Social Networking Sites: An Examination Using a Simulated Facebook Newsfeed*. *Suicide & life-threatening behavior*, 46(5): 609–624. <https://doi.org/10.1111/sltb.12241>
- Glen Coppersmith, Kim Ngo, Ryan Leary, Anthony Wood, 2016. Exploratory analysis of social media prior to a suicide attempt. In *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, pages 106–117.
- Guan L, Hao B, Cheng Q, Yip PS, Zhu T., 2015. *Identifying Chinese microblog users with high suicide probability using internet-based profile and linguistic features: Classification model*, *JMIR mental health*, May 12;2(2):e17. <http://dx.doi.org/10.2196/mental.4227>
- James W. Pennebaker, 2007. *The development and psychometric properties of LIWC2007*. Austin, TX: LIWC.net.
- Jenna L Baddeley, Gwyneth R Daniel, James W. Pennebaker, 2011. How Henry Hellyer's use of language foretold his suicide. *Crisis* 32(5): 288–292. <http://dx.doi.org/10.1027/0227-5910/a000092>
- John Pestian, Henry Nasrallah, Pawel Matykiewicz, Aurora Bennett, and Antoon Leenaars, 2010. *Suicide note classification using natural language processing: a content analysis*, *Biomed Inform Insights*, 3: 19–28.
- Jose M. Gomez, 2014. Language technologies for suicide prevention in social media, in *Proceedings of the Workshop on Natural Language Processing in the 5th Information Systems Research Working Days (JISIC)*.
- Li Guan, Bibo Hao, Qijin Cheng, Tingshao Zhu, 2015. *Identifying Chinese Microblog Users With High Suicide Probability Using Internet-Based Profile and Linguistic Features: Classification Model*. *JMIR Ment Health*, 2(2):e17, <http://dx.doi.org/10.2196/mental.4227>
- M Lv, A Li, T Liu, T Zhu, 2015. *Creating a Chinese suicide dictionary for identifying suicide risk on social media*, *PeerJ*. Dec 15;3: e1455. <http://dx.doi.org/10.7717/peerj.1455>
- Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. 2016. Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, pages 2098–2110. <https://doi.org/10.1145/2858036.2858207>
- Peter Burnap, Walter Colombo, Johnatan Scourfield, 2015. *Machine Classification and Analysis of Suicide-Related Communication on Twitter*, *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, pp. 75–84. <http://dx.doi.org/10.1145/2700171.2791023>.
- Samah Fodeh, Joseph Goulet, Cynthia Brandt, Al-Talib Hamada, 2017. *Leveraging Twitter to better identify suicide risk*, in *Proceedings of The First Workshop Medical Informatics and Healthcare held with the 23rd SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1–7.
- Shannon W. Stirman and James W. Pennebaker, 2001. *Word use in the poetry of suicidal and non-suicidal poets*. *Psychosomatic Medicine*, 63(4): 517–522.
- Tatiana A. Litvinova, 2016. *Corpus studies of speech of individuals who committed suicides*. *Russian Linguistic Bulletin*, 7(3): 133–136. <http://doi.org/10.18454/RULB.7.16>
- Tatiana A. Litvinova, Pavel V. Seredin, Olga A. Litvinova, Olga V. Romanchenko, 2017. *Identification of Suicidal Tendencies of Individuals Based on the Quantitative Analysis of Their Internet Texts*, *Computación y Sistemas*, 21(2): 243–252.
- Tatiana Litvinova, Olga Zagorovskaya, Olga Litvinova, Pavel Seredin, 2016. *Profiling a set of personality traits of a text's author: a corpus-based approach*, *Lecture Notes in Computer Science*, Vol. 9811, pp. 555–562. [http://doi.org/10.1007/978-3-319-43958-7\\_67](http://doi.org/10.1007/978-3-319-43958-7_67)
- Tim M. H. Li, Michael Chau, Paul S. F. Yip, and Paul W. C. Wong, 2014. *Temporal and Computerized Psycholinguistic Analysis of the Blog of a Chinese Adolescent Suicide*, *Crisis: The Journal of Crisis Intervention and Suicide Prevention*, 35(3), 168–175. <http://doi.org/10.1027/0227-5910/a000248>
- Timothy M. DeJong, James C. Overholser, and Craig A. Stockmeier, 2010. *Apples to oranges?: A direct comparison between suicide attempters and suicide completers*. *Journal of Affective Disorders*, 124(1–2), 90–97. <http://doi.org/10.1016/j.jad.2009.10.020>
- Yla R. Tausczik, James W. Pennebaker. 2010. *The Psychological Meaning of Words: LIWC and Computerized Text Analysis Method*. *Journal of Language and Social Psychology*, 29(1): 24–54. <https://doi.org/10.1177/0261927X09351676>