PROFILING A SET OF PERSONALITY TRAITS OF TEXT AUTHOR: WHAT OUR WORDS REVEAL ABOUT US*

TATIANA LITVINOVA
Voronezh State Pedagogical University
centr_rus_yaz@mail.ru

PAVEL SEREDIN
Voronezh State University
paul@phys.vsu.ru

OLGA LITVINOVA
Voronezh State Pedagogical University
olga_litvinova_teacher@mail.ru

OLGA ZAGOROVSKAYA
Voronezh State Pedagogical University
olzagor@yandex.ru

Abstract
Authorship profiling, i.e. revealing information about an unknown author by analyzing their text, is a task of growing importance. One of the most urgent problems of authorship profiling (AP) is selecting text parameters which may correlate to an author's personality. Most researchers' selection of these is not underpinned by any theory. This article proposes an approach to AP which applies neuroscience data. The aim of the study is to assess the probability of self-destructive behaviour of an individual via formal parameters of their texts. Here we have used the "Personality Corpus", which consists of Russian-language texts. A set of correlations between scores on the Freiburg Personality Inventory scales that are known to be indicative of self-destructive behaviour ("Spontaneous Aggressiveness", "Depressiveness", "Emotional Lability", and "Composedness") and text variables (average sentence length, lexical diversity etc.) has been calculated. Further, a mathematical model which predicts the probability of self-destructive behaviour has been obtained.

Keywords: authorship profiling, neurolinguistics, language personality, computational stylometry, discourse production

* The study was funded by the RF President's grants for young scientists N° MK-4633.2016.6 "Predicting the Probability of Suicide Behavior Based on Speech Analysis".

1. Introduction

Authorship profiling (also referred to as AP), which is the process of revealing conjectural information about an unknown author (demographics, personality, education, mental health, etc.), just by computer analysis of a given text (on lexical, morphological, syntactical etc. levels), is a task of growing importance – for national security, criminal investigations, and market research. Scientists trying to address this task generally assume, as a given, the sociolinguistic observation that different groups of people speaking or writing in a particular genre and in a particular language use that language differently. To solve the authorship profiling problem, text corpora are used for which author details (gender, age, psychological testing results, etc.) are known, and numerical values of particular text parameters (content-based parameters – e.g., proportions of certain vocabulary groups; and style-based parameters – e.g., proportions of prepositions, conjunctions and other function words) are calculated. Correlations between text and personality parameters are identified; based on these, mathematical models are designed by means of mathematical statistical methods (regression methods, computer learning methods) with the input data being numerical text parameters and the output data being personality parameters. This is now a common approach. For example, in (Argamon et al., 2009) the researchers show that the right combination of linguistic features and statistical methods enable an automated system to effectively determine the gender, age, native language, and level of neuroticism of an anonymous author.

There has been a growing interest of late in the Author Profiling (AP) task (Argamon et al., 2009; Chung and Pennebaker, 2009; Koppel, Argamon and Shimoni, 2003; Litvinova, 2014; Litvinova, Seredin and Litvinova, 2015; Noecker, Ryan and Juola, 2013; Pennebaker, 2011; Pennebaker, Mehl and Niederhoffer, 2003; Rangel et al., 2014; Rangel et al., 2015; Schler et al, 2006; Tausczik and Pennebaker, 2010). This is mostly due to a rapid increase in volumes of internet communications, and thus a growing demand for methods which enable the identification of internet communicators. For example, in (Koppel, Argamon and Shimoni, 2003) the authors studied the problem of automatically determining an author's gender by the use of combinations of simple lexical and syntactic features; they achieved an accuracy of about 80% via this means. Schler et al. (2006) studied the effect of age and gender on the writing-style exhibited in blogs; the authors gathered over 71,000 blogs and determined from them a set of relevant stylistic features (e.g. use of non-dictionary words, parts-of-speech, function words and hyper-links), and content features (such as word unigrams with the highest information gain). Analysing the texts using these features, they obtained an accuracy of about 80% for gender identification and about 75% for age identification.

There are international competitions which have been instituted in order to reveal the most accurate methods of authorship profiling using numerical parameters of texts (Rangel et al., 2014; Rangel et al., 2015). The prize winners

of the 15th evaluation lab on digital text forensics PAN 2015, which was held in a bid to find the most accurate ways of identifying the gender, age, and psychological traits in accordance with the Five Factor Theory (extroversion, emotional stability/neuroticism, agreeableness, conscientiousness, openness to experience) of Twitter users (Rangel et al., 2015) applied two types of features. The first type features are content-based (bag of words, words n-grams, term vectors, TF-IDF n-grams, named entities, dictionary words, slang words, ironic words, sentiment words, emotional words). The second type features are style-based (frequencies, punctuations, POS, verbosity measures and various other general statistics as well as Twitter specific ones such as numbers of mentions, hashtags, and urls). In this way they obtained models which yielded an accuracy of up to 95% (depending on the personality parameters and language).

There remains a great deal which needs to be addressed. One of the major issues facing researchers dealing with text-based personality detection is that of which text parameters are to be analysed. Most studies provide no explanation of the correlations between quantitative text parameters and personality traits; there is thus no theory supporting the choice of any particular parameter (Nini, 2014). A lot of research has revealed a special significance to the frequencies of certain FW (prepositions, pronouns, conjunctions, particles, etc.) and certain parts of speech (POS) and their sequences (POS n-grams) in relation to identifying certain personality traits (Argamon et al. 2009; Chung and Pennebaker, 2009; Koppel, Argamon and Shimoni, 2003; Litvinova, 2014; Litvinova, Seredin and Litvinova, 2015; Nini, 2014; Pennebaker, 2011; Pennebaker, Mehl and Niederhoffer, 2003; Schler et al, 2006; Tausczik and Pennebaker, 2010). These text parameters are common and not affected by the content and thus are very useful in research. However, numerous researchers do not seem to identify any reasons for these correlations between the texts and the personality parameters of their authors. Instead, they focus on making their mathematical models more accurate mostly by using not only context-independent parameters (POS, FW) but also frequencies of words or particular vocabulary groups (even though it is clear that the performance of the models obtained using these will be effective only on texts of the same genre as those within the learning selection. Additionally, most of these studies have been conducted only on English texts.

Along with personality and demographic traits, research in authorship profiling has addressed a variety of clinical problems (Demjen, 2015; Rude, Gortner and Pennebaker, 2004) including suicide. It is estimated that each year 800,000 people die by suicide worldwide (Pestian et al, 2010). The language of suicide notes has often been analyzed (Handelman and Lester, 2007; Pestian et al, 2010). In relation to this, text corpora have been designed (see (Marcicczuk, Zaньko-Zielicska and Piasecki, 2011) for more details). Structural characteristics (average sentence length, parts of speech) and content variables (length of communication, instructions, active state, explanation provided, locus of control) of these texts have been identified and analyzed for their predictive value, using modern computer linguistics methods. Another approach has focused on the

semantic content of words used in suicide notes by grouping words into linguistic variables (e.g. positive and negative emotions, hearing, references to people, time, religion) (Pestian et al., 2010). There have been attempts made to design models to distinguish between real and fake texts of the genre (Jones and Bennell, 2007).

Although suicide notes are crucial to this kind of study, they are generally not long enough to provide sufficient insight into the speech production of suicide victims; there has been little research, so far, dealing with the dynamics of their idiostyle manifesting before an impending tragic end. Poetry written by individuals who eventually commit suicide (Stirman and Pennebaker, 2001) as well as song lyrics (Lightman et al., 2007) and diary entries have been analyzed. The paper (Lester, 2014) offers a profound perspective on this problem. The studies suggest that it is not only the semantic parameters but also the structural parameters of texts which need to be focused on.

To the best of our knowledge, the problem of text-based detection of the risk of self-destructive behaviour (suicidal behaviour being an extreme form of this) still has not been adequately addressed.

The objective of this paper is to set forth a new approach to AP which involves choosing text parameters in such a way as to employ neuroscience and personality neurobiology data in the analysis. The approach will be presented in terms of profiling the risk of self-destructive behaviour.

The analysis of extensive scientific literature shows that self-destructive behaviour is not just one particular personality trait but a complex personality feature functioning and manifesting itself on different levels. Self-destructive behaviour is associated with spontaneous aggressiveness, high levels of anxiety, and depressiveness (Angst and Clayton, 1986).

The analysis of current studies shows that self-destruction is partially biologically predetermined (Joiner, Brown and Wingate, 2005; Rozanov, 2004). It has been established that genetically induced conditions in biochemical systems of individuals showing aggressive and particularly suicidal behaviour can cause some deficits or hyper-reactivity of their links thus resulting in challenging combinations of aggressiveness, impulsivity, and anxious instability (Yegorov, 1999). Thus it has been shown that in individuals displaying self-destructive behaviour there is increased activation in right-hemisphere based cognitive problem-solving (both verbal and audio-spatial), caused by left prefrontal dysfunction (Yegorov and Ivanov, 2007).

The brain/cognitive activity of an individual is known to manifest itself in their texts – as intellectual products (see Sedov, 2007 for review). Therefore, we have to assume the following. If we compare texts by individuals displaying self-destructive behaviour and similar texts (of the same genre and topic) written by individuals displaying no such behaviour, we will see that there will be language elements controlled by the right hemisphere and fewer those controlled, principally, by the left hemisphere.

Research involving temporary inactivation of brain hemispheres has established which parts of the brain are responsible for certain discourse elements (e.g., abstract nouns, function words, complex syntactic structures) (Bloom et al., 1994; Long et al., 2012; Sedov, 2007). Some scientists even speak of two texts grammars – a right hemispheric and left hemispheric one:

> modelled structures of mechanisms of intellectual speech activity of the left and right hemisphere are two language grammars (…) that are differently oriented but complimentary. Using a somewhat crude "physiological" metaphor, they can be called right hemispheric and left hemispheric ones. Normal speech reveals a dynamic balance of the two grammars. A severe deviation from the balance causes speech disorders and speech pathologies as the worst-case scenario (Sakharniy, 1994).

Therefore the text parameters for the analysis were selected primarily in regard to the available neurolinguistics data on the language functions of the cerebral hemispheres as identified using the method of temporary inactivation of one hemisphere summarized in (Sedov, 2007). According to this data, as the left hemisphere is inactivated (the right hemisphere is active), individuals experience a reduction in vocabulary (i.e. this becomes less diverse), and his/her speech becomes more clichéd resulting in fewer abstract nouns and words in sentences; fewer complex syntactic structures and less textual cohesion also result. As the right hemisphere is inactivated (the left hemisphere remains active), there is an increase in vocabulary as well as in abstract nouns, and the text remains cohesive but less integral (Sedov, 2007). Since our research is pilot, we selected a limited range of the available text parameters for the analysis – those which are easily quantifiable (indices of readability, lexical diversity, POS, etc.). In order to abstract from any the dependence on the text length, only relative values, i.e. ratios, of the parameters were selected.

This paper looks at ways of identifying the correlations between the formal parameters of Russian-language texts and personality traits which are determinants of self-destructive behaviour (spontaneous aggressiveness, depressiveness, emotional lability, composedness) using personality neurobiology. Further, it looks at the design of a mathematical model which allows identification proneness to self-destructive behaviour in authors of written texts.

2. Materials and methods

2.1. Participants

For this study, we used the "RusPersonality Corpus", which consists of Russian-language texts of different genres which are samples of natural written speech (e.g. description of a picture, essays on different topics, etc.) labelled with

information on their authors (gender, age, results of psychological tests, and so on).

For the purpose of the current study, each respondent (N= 721, 422 female, mean age – 19.8, SD = 3.3, students of Russian universities, all native Russian speakers) was tested using the Freiburg Personality Inventory (FPI) (Litvinova, 2014; Litvinova, Seredin and Litvinova, 2015). This questionnaire was chosen as being capable of accurate measurement of determinants of self-destructive behaviour (Angst and Clayton, 1986). According to a current view, high scores on "Spontaneous Aggressiveness", "Depressiveness", "Emotional Lability" and low scores on "Composedness" were used as being indicative of self-destructive behaviour (Yegorov, 1999).

For a further study respondents with severe risk of self-destructive behaviour were chosen, i.e. those scoring high (7-9) on 3 of 12 scales of FPI: "Spontaneous Aggressiveness" (individuals scoring high on this, display high psychotisation levels resulting in growing impulsive behaviour risks), "Depressiveness" (individuals scoring high on psycho pathological depressive syndrome), "Emotional Lability" (high scores are indicative of an unstable emotional condition with affective reactions), and low (1-3) on "Composedness" (low scores are indicative of low stress resistance), N = 33 (16 females, 17 males, average age is 20, SD = 2.3), and with low self-destructive behaviour risks, i.e. those scoring low (1-3) on 3 scales of FPI: "Spontaneous Aggressiveness", "Depressiveness", "Emotional Lability", and high (7-9) on "Composedness", N = 27 (13 females, 14 males, average age is 19.5, SD = 2.2).

2.2. Procedure

Each respondent (N=60) was asked to produce two texts which were then analyzed as one text: a letter to a friend about things happening lately, and one to an imaginary employer explaining why the respondent was good for a particular job. Respondents were instructed to write as much as possible: whatever first came into their minds. There was a time limit of 40 minutes. An average text was 176 words long, SD = 54 words.

As stated in the introduction, the text parameters for the analysis were selected primarily based on the available neurolinguistics data on the language functions of the cerebral hemispheres as identified using the method of temporary inactivation of one hemisphere. According to a review in (Sedov, 2007), inactivation of the left hemisphere causes the following:

- a severe decline in active lexicon, more clichés in speech;
- fewer verbs and abstract nouns;
- worse performance of short-term verbal memory;
- text integrity not affected but lesser text;
- no more complex syntactic structures;
- fewer function words;

- new use patterns of content words: fewer verbs, more nouns and adjectives;
- more deitic elements.

Inactivation of the right hemisphere results in the following:

- a considerably smaller active lexicon,
- more abstract nouns;
- more speech activity, highly eloquent speech;
- worse performance of visual memory;
- text coherence not affected but integrity compromised.

For the analysis, only quantifiable parameters which can be automatically retrieved from texts were selected.

The following were on the list:

1. Indices of the readability of the texts:

1.1. Flesch readability index. For the Russian language, this is calculated according to the formula (Oborneva, 2005):

$$\text{Flesch index} = 206.835 - 1.3\,\frac{(totalofwords)}{(totalofsentences)} - 60.1\,\frac{(totalofsyllables)}{(totalofwords)},$$

the calculation was performed using an online service http://audit.te-st.ru/tests/readability/.

1.2. Hanning Index (or Fog Index). For the Russian language, this was calculated using the formula (Oborneva, 2005):

$$\text{Hanning index} = 0.4\,[0.78(\frac{(totalofwords)}{(totalofsentences)}) + 100\,(\frac{(totalofcomplexwords)}{(totalofwords)})],$$

where the total of complex words is a number of words with more than 4 syllables; 0.78 is a correction coefficient for Russian. The index was automatically calculated using the service http://audit.te-st.ru/tests/readability/

1.3. Average sentence length in words. This was calculated as the ratio of the total of words to the total of sentences.

2. Index of lexical diversity in the text, i.e. the ratio of the number of different word forms to the total of word forms abstracted to the range 0 to 100. This index was automatically calculated using special software Novel Score
http://sourceforge.net/projects/novelscore/

3. Frequencies of different parts of speech.
3.1. Frequencies of function words in a text.
3.1.1. Frequencies of prepositions in a text. The percentage of prepositions was calculated as a ratio of the total of prepositions to the total number of words in a text. Here and further on, the calculation of the frequencies of different parts of speech was made automatic using the designed Python script and polymorphy2.
3.1.2. *Frequencies of conjunctions in a text.* The percentage of conjunctions in a text was calculated as a ratio of the total number of conjunctions to the total number of words in a text.
3.1.3. Frequencies of particles in a text. The percentage of particles in a text was calculated as a ratio of the total number of particles to the total number of words in a text.
3.1.4. Coefficient of coherence. This was calculated using the formula (particles + conjunctions + prepositions)/3N· sentence (Fotekova and Akhutina, 2002).
3.2. Frequencies of pronouns.
3.2.1. Frequencies of personal pronouns in a text. This was calculated as a ratio of the total number of prepositions to the total number of words in a text.
3.2.2. *Pronominalization index*. This was calculated as a ratio of the total number of pronouns to the total number of nouns.
3.3. Coefficient indicating the ratio of the total number of verbs and pronouns to the total number of nouns and adjectives was calculated using the formula (verbs + personal pronouns)/(nouns + adjectives).

3. Results

The data on numerical values of text parameters and scores on certain scales of FPI was exported into SPSS Statistics and a correlation analysis was performed of numerical values of the selected text parameters and scores of the test scales (for each scale: “Spontaneous Aggressiveness”, “Depressiveness”, “Emotional Lability”, “Composedness”), $p < 0.05$. The correlation (see Tables 1-4, Fig. 1) revealed that a lot of text variables correlate with several psychological personality traits at a time. In our view, this fact is accounted for by a high mutual dependence of personality traits themselves due to the shared neurobiological roots.

In order to detect self-destructive tendencies (as noted above, a set of personality traits) by means of the obtained correlation coefficients considering multicollinearity, a regression model, which was a system of linear equations (for each personality trait associated with self-destructive behaviour), was designed. Generally the system of linear equations looked like the following:

$$c_1 = a_1b_{11} + a_2b_{21} + a_3b_{31} + a_4b_{41} + a_5b_{51}$$
$$c_2 = a_1b_{12} + a_2b_{22} + a_3b_{32} + a_4b_{42} + a_5b_{52}$$
$$c_3 = a_1b_{13} + a_2b_{23} + a_3b_{33} + a_4b_{43} + a_5b_{53}$$
$$c_4 = a_1b_{14} + a_2b_{24} + a_3b_{34} + a_4b_{44} + a_5b_{54}$$

where a_1 is the average sentence length; a_2 is the index of lexical diversity; a_3 are frequencies of prepositions; a_4 are frequencies of conjunctions; a_5 are frequencies of personal pronouns; c_1 is spontaneous aggressiveness; c_2 is depressiveness; c_3 is composedness; c_4 is emotional lability.

The obtained equation system can be handily represented as a matrix with text parameters as the input parameters (row vector A) and personality traits as the output parameters (column vector C), *B* is a matrix model:

$$A \times B = C,$$

A x B = C,

$$\begin{pmatrix} a_1 a_2 a_3 a_4 a_5 \end{pmatrix} x \begin{bmatrix} b_{11} b_{12} b_{13} b_{14} \\ b_{21} b_{22} b_{23} b_{24} \\ b_{31} b_{32} b_{33} b_{34} \\ b_{41} b_{42} b_{43} b_{44} \\ b_{51} b_{52} b_{53} b_{54} \end{bmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix}$$

where a_1 is the average sentence length; a_2 is the index of lexical diversity; a_3 are frequencies of prepositions; a_4 are frequencies of conjunctions; a_5 are frequencies of personal pronouns; c_1 is spontaneous aggressiveness; c_2 is depressiveness; c_3 is composedness; c_4 is emotional lability.

The solution for this model can be easily found using a Mathematica software by finding a global minimum of the function f(*B*) = |*AB*-*C*|. The matrix elements at which the above system reaches its minimum were identified based on the numerical values of all the text parameters that are included in the study corpus and personality traits of authors.

The matrix model identified using minimization for calculating the personality traits based on the selected text parameters is as follows:

$$B = \begin{bmatrix} 0.501 & 0.475 & 0 & 0.267 \\ 1.527 & 0.396 & 7.772 & -1.38 \\ -12.744 & 0 & 9.643 & 0 \\ 0 & 0 & 0 & 45.163 \\ 0 & 0 & -19.473 & 0 \end{bmatrix}$$

The minimization model was proved to be highly efficient. The average deviation from the test results was 2 points (on a 10 scale) for each personality trait (spontaneous aggressiveness; depressiveness; composedness; emotional lability). The accuracy of the model for predicting self-destructive behaviour is about 80%.

4. Discussion

We made it our objective to develop a new approach to the selection of language parameters of texts for authorship profiling. Specifically, we were concerned with the assessment of the risk of self-destructive behaviour by authors, using quantitative parameters of their texts. Most previous studies on authorship profiling provide no comprehensive explanation of correlations which exist between text parameters and authors' personality traits. In our selection of text parameters in order to design a mathematical model for profiling self-destructive behaviour, we made use of neuroscience data. Neurolinguistics gives insights into cerebral mechanisms of speech production in the context of functional asymmetry. Neurobiology provides data as to how brains of individuals with self-destructive tendencies operate. We assume that as right hemisphere dominance is common (for individuals with self-destructive tendencies) in solving cognitive tasks, which is primarily due to the dysfunction of the prefrontal cortex of the left frontal lobe, texts by such individuals as opposed to those displaying no such tendencies, can be expected to contain more language elements controlled by the right hemisphere and fewer of those controlled by the left one respectively.

In order to test this hypothesis, we selected a corpus of texts by individuals with high and low risks of self-destructive behaviour (according to FPI data) and labelled it according to the list of parameters chosen based on the neurolinguistics data on speech production in the brain.

As correlation-regression analysis shows, texts produced by individuals with a greater likelihood of self-destructive behaviour (i.e. those who scored high on spontaneous aggressiveness; depressiveness; emotional lability and low on composedness according to FPI) typically show less lexical diversity, fewer prepositions, more pronouns overall (and particularly personal ones), a higher coefficient of coherence (due to more conjunctions and deictic particles), and a

higher average sentence length as compared to texts produced by people with less likelihood of self-destructive behaviour (i.e. those who scored low on spontaneous aggressiveness; depressiveness; emotional lability and high on composedness according to FPI).

The data are overall consistent with our hypothesis. Using the available data on the neurobiology underlying self-destructive behaviour, we suggested that texts by individuals with high risk of self-destructive behaviour could contain more text elements controlled by the right hemisphere and fewer those for which the left hemisphere is responsible than texts by individuals displaying no such behaviour. Indeed, a lower coefficient of lexical diversity in individuals with a greater likelihood of self-destructive behaviour is consistent with the data indicating less vocabulary in individuals with the activated right hemisphere. A lower percentage of prepositions in the above individuals is accounted for by insufficient activation of the left hemisphere areas known to be responsible for producing more abstract lexical units.

A higher pronominalization index, which is characteristics of written speech of people with greater likelihood of self-destructive behaviour, "is commonly observed in weaker paradigmatic language links relying on the cerebellum" (Fotekova and Akhutina, 2002, p. 82).

It is completely consistent with the neurobiological and neuropsychology data indicating that insufficient activation of the cerebellum is associated with aggressive and suicidal behaviour (Pennebaker and Stone, 2004).

Therefore, our study indicated that the identified correlations between text parameters and a set of personality traits associated with self-destructive tendencies are not random and can be accounted for using neurolinguistics data on the brain mechanisms of discourse production on one hand and neurobiology of personality on the other.

We identified correlations between text parameters and personality traits (represented by personality test scores) and designed a mathematical statistical model, which proved to be 80% accurate. Unlike most studies on AP, this study was concerned with language parameters which were selected on the basis of theoretical findings, i.e. neuroscience data. We argue that this current research can significantly inform further studies in authorship profiling as:

1) it proposes an approach to selecting text parameters while employing theoretical findings and behavioural data from neurolinguistics and neurobiology in particular.
2) it suggests that it is not a particular personality trait that needs to be analysed but a whole set of traits, as the neurobiology of personality indicates that self-destructive behaviour is based on a large number of personality traits which share neurobiological foundations and are mutually correlating;
3) a mathematical solution for profiling a set of personality traits using texts is set forth;

4) the problem is addressed using Russian language materials. This has not previously been extensively researched in relation to authorship profiling (Litvinova, 2014; Litvinova, Seredin and Litvinova, 2015);
5) a model which predicts the risk of self-destructive behaviour based on formal text parameters is proposed – although we are aware of certain limitations of the study due to the relatively small sample size, as well as the relatively few language parameters which were used for the analysis.

Of course, further research is necessary for a more comprehensive assessment of these results. This would involve more respondents and more text parameters. This would be particularly significant to validate the results in relation to the coefficients of coherence and the cohesion of texts. We argue that the only way to obtain a greater insight into the language features typical of a particular group of individuals is to use a comprehensive psycholinguistic discourse analysis taking into account the current understanding of text generation and its underlying neural mechanisms and also data from research into the neurobiology of personality. Employing automatic language processing, statistical methods, and neurobiology data in investigating texts produced by individuals who committed suicide is seen as crucial to studying the language correlates of self-destructive behaviour and so designing prognostic models.

5. Conclusions

A worldwide and rapidly developing approach to detecting the personality of authors from their texts, involving the design of predictive models based on the correlations between quantifiable text parameters and individual psychological traits, is not without flaws, since the text parameters are normally selected intuitively and without reference to any theory. These results in the correlations found being provided with no adequate scientific explanation. The approach, presented here, of detecting individual psychological personality traits in texts via mathematical, statistical analysis of text parameters which have been selected and used with reference to the results of neuroscience research, suggests that corpus material can be employed to even greater effect in the future. Our research showed that a text, on its different levels, can be indicative of particular psychological traits of its author which constitute self-destructive thought/behaviour patterns. It is however clear that for a maximally correct solution of the problem at hand it is necessary that even more parameters are used and that a multi-level text analysis is employed – relating to current understandings of speech generation and the involvement of different parts of the brain. This complex approach will not only allow a greater insight into the connection between the personality and speech production but also more efficient practical methods of detecting personality traits using texts produced by individuals.

References

Angst, J. and P. Clayton. 1986. Premorbid Personality of Depressive, Bipolar, and Schizophrenic Patients with Special Reference to Suicidal Issues. *Comprehensive Psychiatry* 27(6). 511–532.

Argamon, S. et al. 2009. Automatically profiling the author of an anonymous text. *Communications of the ACM* 52(2). 119–123.

Baddeley, J. L., Daniel, G. R. and J. W. Pennebaker. 2011. How Henry Hellyer's Use of Language Foretold His Suicide. *Crisis* 32 (5). 288–292.

Bloom, L. R. et al. 1994. Hemispheric Responsibility and Discourse Production: Contrasting Patients with Unilateral Left and Right Hemisphere Damage. In L. R. Bloom, L. K. Obler, S. D. Santi and J. S. Ehrlich (eds.). *Discourse Analysis and Applications: Studies in Adult Clinical Populations*, 91-94. Lawrence Erlbaum Associates Publishers.

Chung, C. K. and J. W. Pennebaker. 2009. The psychological functions of function words. In K. Fiedler (ed.), *Social communication*, 343-359. New York: Psychology Press.

Demjen, Z. 2015. *Sylvia Plath and the Language of Affective States: Written Discourse and the Experience of Depression*. Bloomsbury.

Fernбndez-Cabana, M. et al. 2013. Suicidal Traits in Marilyn Monroe's Fragments: An LIWC Analysis. *Crisis: The Journal of Crisis Intervention and Suicide Prevention* 34(2). 124–130.

Fotekova, T. A. and T. V. Akhutina. 2002. *Diagnostika rechevikh narushenii shkol'nikov s ispol'zovaniem neiropsikhologicheskikh metodov* [Detecting Speech Impediments in School Children Using Neuropsychological Methods]. Moscow: ARKTI.

Handelman, L. D. and D. Lester. 2007. The Content of Suicide Notes from Attempters and Completers. *Crisis* 28, 102–104.

Joiner, T. E., Brown, J. S. and L. R. Jr. Wingate. 2005. The Psychology and Neurobiology of Suicidal Behaviour. *Annu Rev Psychol* 56. 287–314.

Jones, N. and C. Bennell. 2007. The Development and Validation of Statistical Prediction Rules for Discriminating Between Genuine and Simulated Suicide Notes. *Archives of Suicide Research: Official Journal of the International Academy for Suicide Research* 11(2). 219.

Koppel, M., Argamon, S. and A. Shimoni. 2003. Automatically Categorizing Written Texts by Author Gender. *Lit and Ling Computing* 17(4). 401–412.

Lester, D. 2014. *The* "I" *of the Storm: Understanding the Suicidal Mind*. De Gruyter Open Ltd.

Lightman, E. J. et al. 2007. Using Computational Text Analysis Tools to Compare the Lyrics of Suicidal and non-suicidal Songwriters. In D. S. McNamara & G. Trafton (eds.), *Proceedings of the 29th Annual Cognitive Science Society*, 1217-1222. Hillsdale, NJ: Erlbaum.

Litvinova, T. A. 2014. Profiling the Author of a Written Text in Russian. *Journal of Language and Literature* 5(4). 210–216.

Litvinova, T. A., Seredin, P. V. and O. A. Litvinova. 2015. Using Part-of-Speech Sequences Frequencies in a Text to Predict Author Personality: a Corpus Study. *Indian Journal of Science and Technology* 8(9). 93–97.

Long, D. L., et al. 2012. The Organization of Discourse in the Brain: Results from the Item-Priming-in-Recognition Paradigm. In M. Faust (ed.), *The Handbook of the Neuropsychology of Language*, 77–99. Wiley-Blackwell.

Marcicczuk, M., Zaњko-Zielicska, M. and M. Piasecki. 2011. Structure Annotation in the Polish Corpus of Suicide Notes. In I. Habernal and V. Matoušek (ed.), *Text, Speech and Dialogue. 14th International Conference, TSD 2011, Pilsen, Czech Republic, September 1-5, 2011. Proceedings,* 419–426. Springer Berlin Heidelberg.

Nini, A. 2014. Authorship Profiling in a Forensic Context. *PhD thesis*. Aston University.

Noecker Jr, J. W., Ryan, M. and P. Juola. 2013. Psychological Profiling Through Textual Analysis. *Lit Linguist Computing* 28(3). 382–387.

Oborneva, I. V. 2005. Avtomatizatsiia otsenki kachestva vostriiatiya vospriiatiya teksta [Automatisation of the Assessment of Perception of a Text]. *Vestnik Moskovskogo gorodskogo*

pedagogicheskogo universiteta [Herald Journal of Moscow State Pedagogical University] 2(5). 86–92.

Pennebaker, J. W. 2011. *The Secret Life of Pronouns: What Our Words Say About Us*. New York: Bloomsbury Publishing.

Pennebaker, J. W., Mehl, M. R. and K. Niederhoffer. 2003. Psychological Aspects of Natural Language Use: Our Words, Our Selves. *Annual Review of Psychology* 54. 547–577.

Pennebaker, J. W. and L. D. Stone. 2004. What Was She Trying To Say? A Linguistic Analysis of Katie's Diaries. In D. Lester (ed.), *Katie's Diary: Unlocking the Mystery of a Suicide,* 55–80. New York: Brunner-Routledge.

Pestian, J. et al. 2010. Suicide Note Classification Using Natural Language Processing: A Content Analysis. *Biomed Inform Insights* 3. 19–28.

Pilyagina, G. Ya. 2003. Mekhanismi suitsidogeneza i otsenka suitsidal'nogo riska pri razlichnikh formah autoagressivnogo povedeniya [Mechanisms of Suicidogenesis and Assessments of Suicidal Risks in Different Forms of Self-destructive Behaviour]. *Arhiv psihiatrii* [Psychiatry Archives] 9(4). 18–26.

Rangel, F. et al. 2014. Overview of the 2nd Author Profiling Task at PAN 2014. In L. Cappellato, N. Ferro, M. Halvey and W. Kraaij (eds.), *CLEF 2014 Labs and Workshops, Notebook Papers*. CEUR-WS.org, vol. 1180 898–827.

Rangel, F. et al. 2015. Overview of the 3rd Author Profiling Task at PAN 2015. In CEUR Workshop Proceedings. [Online] Available from: http://www.sensei-conversation.eu/wp-content/uploads/2015/09/15-pan@clef.pdf [Accessed: 19.12.2016]

Rozanov, V. A. 2004. Neirobiologicheskie osnovi suitsidal'nogo povedeniya [Neurobiuological Foundations of Suicidal Behaviour]. *Vestnik biologicheskoj psihiatrii* [Herald Journal of Biological Psychiatry] 6. [Online]
Available from: http://scorcher.ru/neuro/science/data/mem102.php [Accessed: 19.12.2016]

Rude, S., Gortner, E. M. and J. Pennebaker. 2004. Language use of depressed and depression-vulnerable college students. *Cognition and Emotion* 18(8). 1121-1133.

Sakharniy, L. V. 1994. Chelovek i tekst: dve grammatiki teksta [Man and Text: Two Grammars of a Text]. *Chelovek – tekst – kul'tura* [Man – Text – Culture]. Yekaterinburg. 17–20.

Schler, J. et al. 2006. Effects of Age and Gender on Blogging. In *Proc. of AAAI Spring Symposium on Computational Approaches for Analyzing Weblogs*, 199-205. AAAI.

Sedov, K. F. 2007. *Neiropsikholingvistika* [Neurolinguistics]. Moscow: Labirint.

Stirman, S. W. and J. W. Pennebaker. 2001. Word Use in the Poetry of Suicidal and Non-Suicidal Poets. *Psychosom Med* 63(4). 517–522.

Tausczik, Y. R. and J. W. Pennebaker. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. Journal of Language And Social Psychology 29(1). 24–54.

Yegorov, A. Yu. 1999. Koordinatsiya dejatel'nosti polusharii mozga cheloveka pri osushestvlenii kognitivnikh funktsii [Coordination of the Activities of the Right Hemisphere of the Human Brain]: abstract of thesis for PhD in Medicine. Saint Petersburg.

Yegorov, A. Yu. and O. V. Ivanov. 2007. Osobennosti individual'nykh profilei funktsional'noi assimetrii u lits sovershivshikh suitsidal'nuiu popytku [Features of Individual Profiles of Functional AssymetryAsymmetry in Individuals Committed a Suicidal Attempt]. *Social and Clinical Psychiatry* 2. 20–24.