Authorship Profiling in Russian-Language Texts

Tatiana Litvinova[1], Olga Litvinova[1]

[1]Voronezh State Pedagogical University + Voronezh – Russian Federation

Abstract

Authorship profiling, which is a process of the extraction of information about the unknown author of a text (demographics, psychological traits, et al.) using the methods of mathematical statistics based on the analysis of quantitative linguistic parameters, is a problem of great importance. The English language materials have been mainly focused on, which is mainly due to no special research text corpora in place. A similar study of the Russian language has not been performed as of now. The article looks at the results of experimental studies conducted using a corpus of Russian-language texts "RusPersonality" including the authors' information (gender, age, psychological testing results, lateral profile, etc.) to identify self-destructive tendencies in authors A mathematical model which predicts the probability of self-destructive behaviour has been obtained.

Key words: authorship profiling, authorship attribution, corpus linguistics, Russian language, self-destructive behavior.

1. Introduction

Studies in language psychology show that there is a direct correlation between personality and language: personality is projected linguistically and can be perceived through language (see, for example, Pennebaker et al., 2003). However, these studies have long been performed not in a prediction context, but in a descriptive statistics context. Since 2005 as a paper by Argamon et al. was published, a diagnostic approach has been commonly used. Using mathematical statistics, researchers design mathematical models for profiling certain personality traits (gender, age, psychological characteristics, education level, native language, etc.) based on numerical values of different text parameters (frequencies of n-grams of symbols, words, parts of speech (POS), certain function words (FW), etc.). This task which is often referred to as authorship profiling task (AP) is of growing importance – for national security, criminal investigations, and market research. Yet, many of the basic research questions of this field are not studied systematically or even not at all. Therefore, in spite a large number of studies into the problem, viable methods of personality profiling using texts to be applied in forensic context have not been developed yet (Nini, 2014). There is no question that «it is the feature selection (or feature construction) problem which is most important in this field rather than the choice of machine learning method» (Daelemans, 2013, p. 3), and it has to be well-grounded.

One of the major issues facing AP apart from selecting text parameters for analysis is that of a real reference corpora that have to include meta-data providing details about the authors and

to be carefully balanced according to genre, topic, age, and gender; "only then can real progress be made in solving the fundamental problems of computational stylometry" (Daelemans, 2013, p. 8). As Verhoeven and Walter rightfully point out, authorial profile can be hard to get, whereas having large amounts of data remains the key to reliable results in computational stylometry (Verhoeven and Walter, 2014). There are few text corpora specially designed to study AP containing information about the authors used in linguistics worldwide. One of the largest (about 2400, one for each author/user, 1.9 million words) is that of students' essays (often referred to as essay dataset) with annotations as to their BIG5 personality type (Pennebaker and King 1999). One of the few examples of such corpora of non-English texts is CLiPS Stylometry Investigation (CSI), a freely available Dutch corpus which provides textual data in two genres (reviews and essays written by university students, 305 000 tokens spread over 749 documents) with large amounts of meta-data (age, gender, region of origin, personality (Big5, MBTI), sexual orientation) and expanded on a yearly basis (Verhoeven and Daelemans, 2014). Therefore according to the analysis, there are no sufficient research text corpora with meta-data providing details about their authors designed considering the genre and topic and used for research of AA and AP particularly for languages other than English. So far scientists seeking to monitor a few characteristics of text authors (e.g., gender, age, native language, and neuroticism level) have had to make use of a variety of text corpora instead of one: «Ideally, we would prefer to use a single corpus for all these problems but, unfortunately, there is no single corpus in which the documents are labeled for all four issues we consider» (Argamon et al., 2009). In this paper we are introducing *RusPersonality,* a Russian text corpus containing meta-data providing details about the authors and used in stylometry studies. According to our knowledge, there has been no such a corpora for Slavic languages. We are also presenting the results of a study using this corpus to identify self-destructive tendencies applying the method in selecting language variables to be analyzed according to neuroscience data.

2. "RusPesonality": corpus description

The text corpora *RusPesonality* was launched in 2012 and the work on it is still ongoing. The participants of the experiment are Russians residing in different cities of the Russian Federation (Moscow, Voronezh, Rostov-on-Don, etc.) with most of them being students of Russia's universities (in 2014 a text corpora from non-Russian citizens studying at Russian universities with a sufficient proficiency in Russian to produce a coherent text was started). *RusPesonality* contains texts, i.e. examples of Russian spontaneous written speech (description of pictures, what you did yesterday, a letter to a friend, etc.). All the texts were written exclusively for the corpus in the presence of the experimentator (so that there could be no borrowing, editing, etc.). All the participants were instructed to write the first thing that comes to mind on 1 or 2 (depending on the experimental conditions) topics within 30-40 minutes. As of now, the corpus contains texts from **1 145** respondents with the total of **1 867 texts.** The average text length ranged from 56 (a subcorpus of texts by respondents with mental disorders) to 230 words (a subcorpus with essays on "What I see as the meaning of life"). All the authors have explicitly given us permission to include their submissions and profile information in the corpus for research purposes. A unique aspect of our corpus is the breadth of the metadata. There is meta-data available on both the authors and the documents included in the corpus. For each author, we have information on gender, age, mother tongue, education, personality scores on psychological testing (for details see below), for some authors – data on neuropsychological assessment, medical diagnosis, profession, qualification.

3. Authorship Profiling Using Neuroscience Data

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.

As was noted above, 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 (Rozanov, 2004).

In order to test this hypothesis, we selected a corpus of texts by individuals with high (those scoring high (7-9) on 3 of 12 scales of Freiburger Personality Inventory: "Spontaneous Aggressiveness" (individuals scoring high on this, display high psychotisation levels resulting in growing impulsive behaviour risks), "Depressiveness" (individuals scoring high on psychopathological 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 low risks of self-destructive behaviour (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) and labelled it according to the list of parameters chosen based on the neurolinguistics data on speech production in the brain. Each participant 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 they (the respondents) were 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 correlation-regression analysis (p<0.05) show, texts produced by individuals with a greater likelihood of self-destructive behaviour 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 higher average sentence lengths as compared to texts produced by people with less likelihood of self-destructive behaviour. The data are overall consistent with our hypothesis. 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 a characteristics of written speech of people with a 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 (Rozanov, 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 the neurolinguistics data on the brain mechanisms of discourse production on one hand and neurobiology of personality on the other.

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. 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: $A \ x \ B = C,$

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

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 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%.

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%.

5. Conclusions and Further Research

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 behavioral data from neurolinguistics and neurobiology in particular; 2) it suggests that it is not a particular personality trait that needs to be analyzed 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 et al., 2015a; Litvinova, 2014); 5) a model which predicts the risk of self-destructive behaviour based on formal text parameters is proposed.

6. Acknowledgements

References

Argamon S., Dhawle S., Koppel M. and Pennebaker J. (2005). Lexical predictors of personality type. In Shannon W., editor, *Proceedings of Joint Annual Meeting of the Interface and the Classification Society of North America*.

Argamon S., Koppel M., Pennebaker J. and Schler J. (2009). Automatically profiling the author of an anonymous text, *Communications of the ACM*, vol. 52(2):119–123.

Daelemans W. (2013). Explanation in computational stylometry. In Gelbukh A., editor, *Proc. of 14th International Conference on Intelligent Text Processing and Computational Linguistics (CIC-Ling 2013)*, pages 451-462.

Fotekova T. A. and Akhutina T. V. (2002). *Diagnostika rechevikh narushenii shkol'nikov s ispol'zovaniem neiropsikhologicheskikh metodov* [Detecting speech impediments in school children using neuropsychological methods]. ARKTI.

Nini A. (2014). Authorship profiling in a forensic context. PhD thesis. Aston University.

Pennebaker J. W. and King L. (1999). Linguistic styles: language use as an individual difference. *Journal of Personality and Social Psychology*, vol.77(6):1296-1312.

Pennebaker J.W., Mehl M.R. and Niederhoffer K. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, vol.54:547-577.

Rozanov V. A. (2004). Neirobiologicheskie osnovi suitsidal'nogo povedeniya [Neurobiuological foundations of suicidal behaviour]. *Vestnik biologicheskoj psihiatrii* [Herald Journal of Biological Psychiatry]: vol.6.

Verhoeven B. and Daelemans W. (2014). CLiPS Stylometry Investigation (CSI) corpus: a Dutch corpus for the detection of age, gender, personality, sentiment and deception in text. In Calzolari N., editor, *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*.