

RusNeuroPsych: Open Corpus for Study Relations between Author Demographic, Personality Traits, Lateral Preferences and Affect in Text

Tatiana A. Litvinova, Ekaterina S. Ryzhkova

Abstract—A text reflects a range of combinations of individual inter-acting characteristics of its author, both stable (gender, psychological traits, neuropsychological characteristics) and variable (feelings, emotions). It is obvious that it is not in isolation but in a combination that a variety of characteristics comes forth in a text. For example, according to some studies, men and women express their emotions in a text in different ways. It is obvious, though that there are other characteristics that influence the way one chooses to express his/her emotions. Studies of these ways are critical multidisciplinary problems that call for text corpora providing relevant metadata. The paper is devoted to the description of a manually collected corpus of texts (letters to a friend and narratives about pictures from Thematic apperception test, i.e. informal writing describing emotions and opinions) in the Russian language *RusNeuroPsych*, containing metalabelling in the form of information about their authors (gender, age, psychological testing scores, brain laterality preferences). To the best of our knowledge, this is a unique corpus in terms of breadth of metadata about the authors. The corpus is freely available on RusProfiling Lab webpage. The collection and processing of the material to design the corpus, its composition and structure are considered. The possibilities of the application of *RusNeuroPsych* corpus in different domains of knowledge are analyzed.

Keywords—corpus linguistics, personality prediction from text, Russian language, text corpus.

I. INTRODUCTION

Human speech, including writing, is capable of providing a myriad of information about a particular individual. It is by analyzing a coherent and cohesive statement (text) that one is able to get better insight into a variety of individual traits. It is indicative of demographic characteristics (gender, age), education level, personality traits, neuropsychological characteristics etc. of its author.

Manuscript received January 27, 2018. This work was supported by the grant of RFBR “Linguistic Parameters of a Written Text and Neuropsychological Characteristics of its Author: A Corpus Study”, project number 16-36-00036.

T. L. is with the Voronezh State Pedagogical University, Voronezh, 394071 Russia, (+7-980-342-00-73; e-mail: centr_rus_yaz@mail.ru).

E. R. is with the Voronezh State University of Engineering Technologies, Voronezh, 394036 Russia, (e-mail: ryzhkowa.katerina@yandex.ru).

It is obvious that it is not in isolation but in a combination that a variety of characteristics comes forth in a text. For example, Schler et al. [1] showed mutual influences of gender and age. They found out that writing style grows increasingly “male” with age: pronouns and assent/negation become scarcer, while prepositions and determiners become more frequent. Lately, there has been a lot of focus on these interactions (e.g., see the workshop “Computational Modeling of People’s Opinions, Personality, and Emotions in Social Media” co-located with CoLing 2016, <https://peoples2016.github.io/>) especially dedicated to the study of how different traits characterizing whole person are reflected in a text in their combination (stable – like gender – and contextually prompted – like emotions).

Such studies are of both theoretical and practical importance. For example, the authors of [2] showed that the use of text properties describing emotions improves significantly the task of identifying gender.

In order to examine how different personality characteristics are manifested in texts and how exactly individuals display their emotions in them, we need text corpora with relevant metadata about their authors. E.g., there are such corpora as essay dataset designed by J. Pennebaker, myPersonality3, Stylometry Investigation Corpus (CSI).

Essay dataset [3] is a large corpus of stream-of-consciousness texts in English (about 2400, one for each author), collected between 1997 and 2004 and labelled with personality trait scores (Big5 test).

MyPersonality3 is a sample of personality scores (Big5 test), Facebook profile data as well as status updates [4]. This corpus also contains English texts.

Stylometry Investigation Corpus (CSI) corpus [5] is a yearly expanded Dutch corpus of student texts in two genres: essays and reviews. There is a vast amount of metadata available, both on the author (gender, age, sexual orientation, region of origin, personality profile) and on the document (genre, veracity, sentiment, etc.).

There is currently a need to design similar corpora for different languages.

The first Russian text corpus to contain metadata providing information about their authors (gender, age, education level, personality test scores) is *RusPersonality* [6]. This paper looks at a new Russian corpus *RusNeuroPsych* with metadata providing information about

the authors which is suitable to use to investigate how emotions are expressed by individuals with a variety of demographic, psychological and neuropsychological characteristics.

II. RUSNEUROPSYCH CORPUS: COMPOSITION AND STRUCTURE

The data for *RusNeuroPsych* were collected during a psycholinguistic experiments where participants were instructed to answer some survey questions and write texts in the presence of the researcher. As the texts were converted into a digital format, any misprints were eliminated but the original punctuation was retained.

A. Characteristics of the Authors of the Texts

The corpus *RusNeuroPsych* that we have collected contains **644 texts by 455 authors**. The collection is divided into two parts: “Children” (texts written by school children aged from 12 to 17) and “Adult” (texts written by peoples from 18 to 35, mostly students).

Gender. The corpus includes texts written by **190 males** and **259 females**, 6 chose not to report their gender.

Age. Individuals from **12** to **35** years of age participated in designing the written text corpus.

Native language. Russian was a native language of all the participants.

Education. The corpus *RusNeuroPsych* contains texts by people who have not completed their high school education (246 individuals – 6th -10th graders of schools of Voronezh), with high school education (2 individuals), those who have not completed their university degree (199 individuals – 1st -4th year students in different fields at Voronezh State University of Engineering Technology), with university degree (8 individuals – a variety of professionals (teachers, doctors, engineers, etc.)).

Psychological characteristics. All the informants were tested for the identification of their psychological conditions and personality traits. For the school students Eysenck Personality Questionnaire “Self-Assessment of Personality Traits” adapted by N.V. Peresheyina and M.N. Zaoztrovseva was offered. This test is used for measuring levels of aggressiveness, anxiety, rigidity, frustration. This questionnaire includes a description of different psychological conditions that a participant is asked to confirm of (not) experiencing. The survey is commonly used for identifying suicidal tendencies in teenagers.

For “adult” respondents, i.e. students and professionals, The Hospital Anxiety and Depression Scale (HADS) and Five-Factor Personality Inventory by Costa and McCrae (Big5 Test) were offered.

Lateral preference tasks. The special feature of our corpus is that it has metalabelling of the lateral preference of the authors, i.e. data on dominant hand, foot, ear and so on. In scientific literature there is a plethora of data on the connection between lateral preferences and different personal characteristics (cognition, psychological traits, etc.) [7-10].

In scientific literature one finds descriptions of various methods of the identification of lateral preferences in children and adults. In addition, there are tests performed on

special equipment that definitely contribute to making findings more accurate. In this study we have employed the methods that required no special equipment and can thus be used in “field” settings on a large number of participants.

Hence in order to identify motor asymmetries (hand and foot dominance), the respondents were given the following series of tests:

- *hand preference test:* test on interlocking fingers, manual midline crossing, or Napoleon’s pose, clapping [11], filling two 2×2 cm squares with vertical lines (first by the right hand – the right square and then by the left hand – the left square) [10], tests to determine the dominant hand (catching objects) [11], picking up an object [11], test to draw a circle on one hand with the other and identify which one is drawing [10];

- *foot preference test:* crossing legs, a forward step, a backward step, sitting up and down, jumping on one foot [11].

In order to identify sensory asymmetries, the respondents were asked to perform the following tests:

- *dominant eye test:* “blinking with one eye”, “looking through a tube”, tests to identify features of the muscles of the non-dominant eye [11];

- *dominant ear test:* a respondent was asked to determine near which of their ears a hand clapping sound was heard (it is made behind their back equally far from both ears) [10], “clock ticking test” [11], “whisper” test [11], “A Phone Receiver” test (to see which ear a respondent holds the receiver to) [11].

In order to determine the type of cognitive laterality profile [12] the following tests were performed:

1. Test by I.P. Pavlov where respondents are asked to class the words such as “*carp*”, “*eagle*”, “*sheep*”, “*feathers*”, “*scales*”, “*fur*”, “*to fly*”, “*to swim*”, “*to run*” into three groups so that the words in each had something in common [10];

2. Test to class the words “*light*”, “*ear*”, “*vision*”, “*hearing*”, “*nose*”, “*sense of smell*”, “*eye*”, “*sound*”, “*smell*” into three groups based on a property they share [13];

3. Test to class the adjectives “*good*”, “*not intelligent*”, “*bad*”, “*intelligent*”, “*stupid*”, “*not bad*”, “*not stupid*” into two groups so that the words in each had something in common [13];

4. Class the numbers *I 2* and *I II* into two groups randomly [13];

5. Test to disqualify 8 sentences into two groups based on common properties (*Vanya beat up Petya*, *Petya beat up Vanya*, *Vanya was beaten up by Petya etc.*) [13].

For each type of lateral preference test there is a number of “right”, “left” and “mixed” answers (e.g., a respondent came up with two variants of the classification of sentences during the cognitive laterality profile test). Therefore, handedness, footedness, etc. can be regarded as a continuous as well as a categorical variable.

The most challenging were the cognitive laterality profile tests. Some of the respondents failed to do them. Others failed to do some of the sections of the motor and sensory profile test due to previously suffered injuries (as they reported in the questionnaire) as well as to limited time. If

that was the case, extra information was provided in the “Comments” section.

Additionally, some of the participants also failed to do the personality tests.

Below you can find the example of corpus metadata (Fig. 1, 2).

ID	Motor			Sensory			Cognitive			Notes						
	Right	Left	Mixed	Right	Left	Mixed	Right	Left	Mixed							
1	5	2	0	3	2	0	3	0	1	2	0	3	3	2	0	
2	5	2	0	5	0	0	4	0	0	2	1	2	4	1	0	
3	4	2	1	4	1	0	3	0	1	3	1	1	4	1	0	
4	5	0	2	5	0	0	1	1	2	3	0	2	1	4	0	
5	7	0	0	4	1	0	0	2	1	3	2	0	3	1	0	failed tasks K1,C1
6	3	3	1	4	1	0	1	1	2	4	0	1	5	0	0	
7	3	3	1	5	0	0	2	0	2	2	2	0	2	3	0	failed task C1
8	4	2	1	4	1	0	1	1	2	3	2	0	2	3	0	
9	5	1	1	3	2	0	1	1	2	2	1	2	3	2	0	
10	6	1	0	5	0	0	2	0	2	3	0	2	3	2	0	
11	1	6	0	1	4	0	1	2	1	0	5	0	4	1	0	
12	5	2	0	5	0	0	1	1	2	4	0	0	2	2	0	failed tasks K1,C1
13	6	1	0	4	1	0	2	1	1	5	0	0	3	2	0	
14	4	2	1	2	3	0	1	3	0	2	2	1	3	1	0	failed task K1
15	5	1	1	2	3	0	1	1	2	1	1	3	3	1	0	failed task K1
16	6	0	1	5	0	0	1	0	3	1	2	2	4	1	0	
17	3	3	1	2	2	1	2	0	2	1	2	4	0	0	0	failed task K1
18	4	2	1	3	2	0	2	2	0	3	1	1	2	2	0	failed task K1
19	6	0	0	3	2	0	1	3	0	3	0	2	2	0	0	failed tasks H3,C1

Fig. 1. Corpus metadata (lateral preferences)

ID	Gender	Year of birth	Education	Extraversion	Agreeableness	Conscientiousness	Neuroticism	Openness to experience	Notes
1	M	1991	higher	35	44	38	65	55	
2	M	1996	student	39	41	47	41	43	
3	F	1988	higher	58	70	58	51	65	
4	F	1982	no degree	60	50	67	19	64	
5	F	1986	higher	46	58	72	48	50	
6	F	1992	higher	57	69	55	48	54	
7	F	1993	higher	64	56	55	45	61	
8	F	1993	higher	60	45	63	63	62	
9	F	1993	student	65	52	61	31	62	
10	F	1993	student	62	44	58	61	58	
11	F	1993	student	59	57	64	33	63	
12	M	1992	student	48	46	52	58	62	

Fig. 2. Corpus metadata (demographics, education, Big5 scores)

B. Characteristics of the Texts

The average text in the corpus *RusNeuroPsych* is 165 words. The maximum text length is 731 words and the minimum one is 5 words. Before writing the texts, the respondents were instructed to write whatever first comes to their mind without thinking and planning as they would do while speaking, i.e. with no fear of mistakes.

Through the course of the psycholinguistic experiment, the respondents were instructed to write a letter to a friend and a picture description that they could see in the survey (the same picture for all the respondents). It was a provocative yet ambiguous picture included in the Thematic apperception test (TAT) (Fig. 3).

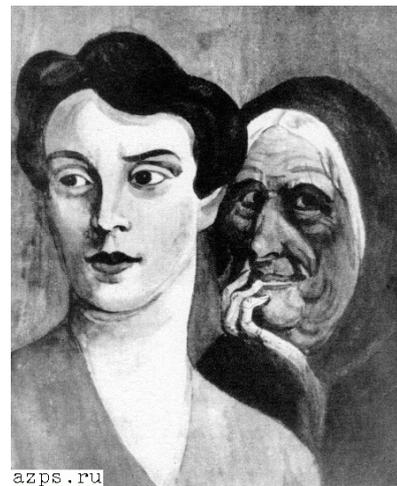


Fig. 3. Picture used for writing task

The subject was asked to tell as dramatic a story as they can for the picture presented, including the following:

- what has led up to the event shown;
- what is happening at the moment;
- what the characters are feeling and thinking;
- what the outcome of the story was.

Let us give an example of the text “A letter to a friend”:
Привет, Данил! Как дела? Я надеюсь, что все хорошо. Последний месяц был очень напряженным и интересным. Я получил права на вождение автомобиля. Я целых полгода ходил на занятия. Это было утомительно, но это того стоило. Теперь я могу управлять авто. Это помогает делать много дел в один день. Побывать в сотнях новых мест, узнать много новых людей. Я понял смысл поговорки: «автомобиль не роскошь, а средство передвижения». И это правда! Учеба дается мне легко. Наша группа очень веселая и сильная. Мы сдаем завтра зачет по информатике. Через месяц у меня сессия. Немного волнуюсь, но да ладно! Расскажи о себе, мне все интересно. Жду ответа! Я знаю, что ты не любишь писать, но надеюсь на ответ. Может, позвонишь, мне будет приятно услышать твой голос. Мы давно не разговаривали по телефону. Твой номер не изменился? Или ты пользуешься только мобильным? Не пропадай. Пока!”

Picture description: *На картине двое: бабушка и внук. Молодость и старость. Былое и будущее. молодой парень устремлен в будущее. Его взгляд открыт и дерзок, немного хитроват. Он уверен: впереди все лучшее. бабушка смотрит на повзрослевшего внука и вспоминает: еще совсем недавно это был ребенок. Как быстро пронеслось время. Он уже совсем взрослый. Скоро уйдет из родного дома. У него будет своя жизнь. Как она сложится? Кто будет рядом с ним? Для бабушки самое главное, чтобы внук был здоров, успешен, счастлив. А внуку хочется новых ощущений. Покорять новые вершины. Узнавать новые места и людей. Чтобы каждый день был не похож на прежний. Хочется веселья и беззаботности и совсем не хочется думать о плохом и грустном. Молодость и старость: два мира, два взгляда на жизнь. Их разделяет целая пропасть, а вернее целая жизнь.*

We did not make it our purpose to interpret the resulting narratives. What we did want was to urge the respondents to express their emotions, feelings, attitudes through their texts.

The tasks varied depending on the group of the respondents: the school students were asked to write one text of choice, “adults” were instructed to write two texts (however, 13 adult respondents wrote one text).

C. Corpus Access Terms

The corpus is freely available for research purposes on the RusProfiling Lab webpage

<http://en.rusprofilinglab.ru/korpus-tekstov/rusneuropsych-corpus/>

III. RESEARCH CONDUCTED USING *RUSNEUROPSYCH*

A. Linguistic characteristics of text by peoples with different lateral preference profiles

One of the most critical neuropsychological characteristics to indicate individual differences when both human cerebral hemispheres function is the lateral preference profile. It has been experimentally proved that asymmetry of functions is characteristic of all levels of signal processing: from the sensory level to the level of the most intricate cognitive tasks [12]. It is regarded as a foundation for the typology of individual differences as part of neuropsychology of individual differences of healthy individuals [7]. As the studies [7] suggest, the classification of people according to the types of interhemispheric interaction corresponds to the features of motor, cognitive, emotional spheres, which means that is the right foundation for the typology (see also [8, 9, 14]). However, features of texts by individuals with different lateral preference patterns have not yet been identified.

We have been performed a study to identify the correlations between text parameters and lateral preference patterns of their authors (only “adults” texts were used in this experiment). For that, the texts were linguistically labeled using a morphological analyzer *pymorphy2* and online service *istio.com* as well as LIWC software [15] supplemented by developed dictionaries (see [16-17] for details about LIWC).

Therefore, we used part-of-speech frequencies, lexical diversity indices and LIWC parameters as features.

The choice of the parameters is firstly due to the fact that they are inherent to any text. Secondly, they are not much dependent on the topic and cannot be consciously imitated.

The lateral preference index has been calculated as the difference between the number of the “right”, “left” and “mixed” answers in all of the tests divided into the number of the tests: $(\text{right} - \text{left} - \text{mixed}) / \text{the total of tests}$.

E.g., in order to determine the dominant hand, a respondent was asked to do the total of 7 tests, 5 of which they did with their right hand, 1 with the left one, in one of the tests there was no dominance of the right/left hand respectively, the index “dominant hand” for this respondent is $(7-1-1)/(7)=0,7$.

For more objectivity an analysis of the same linguistic material has been carried out in two series of the experiment. Hence in the first series of the experiment both texts by the same author (a letter to a friend and a picture description)

was merged and considered as one text (“the total corpus”) and in the second one both texts were analyzed individually (“the individual corpus”). During the processing of the collected linguistic material only those text parameters that were shown to correlate (we used Pearson's correlation method) with the characteristics of motor and sensory laterality profiles of their authors in two series of the experiments have been taken into consideration.

The largest number of correlations ($p < 0.05$) was found between the text parameters and motor asymmetry indices (8), dominant hand index (8), integral profile of the lateral organization (7) (correlation coefficient ranged from 0.27 to 0.41).

A considerably lower number of correlations were found between the indices of sensory asymmetry except the parameter “dominant eye” (5).

Hence a positive correlation was found between the index “dominant hand”, “motor asymmetries” and integral profile of the lateral organization and TTR100 (the number of different words in the first 100 words in a text), i.e. the more “right” answers an individual has given, the higher the lexical diversity index of their text is.

A negative correlation was found between the lateral preference indices and proportions of function words; proportion of function words without pronouns; proportion of words describing cognitive processes and relations; proportion of punctuation marks; proportion of 100 most frequent Russian words, i.e. the more “right” scores there were in an individual's profile of the lateral organization, the lower these indices were.

Therefore the correlation between the linguistic parameters of the texts and lateral preferences indices of the authors has been shown.

B. Connection between gender and lateral preferences and its reflection in text production

In [18] it was shown that texts by authors of different genders but with an identical type of handedness are more similar linguistically than those by individuals of the same gender but with a different type of manual preference.

Using methodology described in detail in [18] the authors have found that texts by male and female with different degree of handedness differ with respect to the following text parameters.

- 1) *right-handed females and left-handed females*: The proportion of function words (FW) in the text; TTR100 (type/token ratio in the first 100 words of the text); proportion of words from the list of 100 most frequent Russian words; proportion of the particle “not” (“не”); proportion of deictic words; number of FW/number of punctuation marks;
- 2) *left-handed males and right-handed females*: proportion of FW in the text; proportion of quantitative words (numerals + pronominal adverbs); proportion of words describing perception;
- 3) *right-handed males and left-handed females*: proportion of words from the list of 100 most frequent Russian words; proportion of the preposition “on” (“на”); proportion of the preposition “by” (“y”); proportion of words describing emotions; number of FW/number of

- commas; number of FW/number of punctuation marks; proportion of the total number of punctuation;
- 4) *right-handed males and left-handed males*: proportion of function words including pronouns in the text; proportion of function words (without pronouns) in the text; percentage of 5 most frequent words excluding function words; proportion of function words in 5 most frequent words in text; proportion of quantitative words (numerals + pronominal adverbs); proportion of perception words; number of FW / number of commas;
 - 5) *right-handed males and right-handed females*: TTR100; proportion of 5 most frequent words including FW in text; proportion of all punctuation marks;
 - 6) *left-handed males and left-handed females*: proportion of words describing perception.

As was shown in [18], the distance measure between texts by right-handed males and right-handed females are the lowest, whereas the highest value of distance measure was found for the texts by right-handed females and left-handed females.

IV. CONCLUSION

We are planning to continue working on expanding the digital corpus of written Russian texts *RusNeuroPsych*. From our perspective, the corpus that contains samples of natural written speech on emotionally charged topics to provide an outlet for the author's feelings and emotions as well as an extensive metalabelling with the information on the authors (including lateral preferences) is going to contribute to the development of studies of the way emotions are described in a text depending on the author's various characteristics.

REFERENCES

- [1] J. Schler, M. Koppel, S. Argamon, and J. Pennebaker, "Effects of age and gender on blogging", in *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, vol. 6, pp. 199–205, 2006.
- [2] F.M. Rangel and P. Rosso, "On the impact of emotions on author profiling", *Information Processing and Management*, vol. 52, no. 1, pp. 73–92, 2016.
- [3] J. W. Pennebaker and L.A. King, "Linguistic styles: language use as an individual difference", *Journal of Personality and Social Psychology*, vol. 77, pp. 1296–1312, 1999.
- [4] M. Kosinski, D.J. Stillwell, and T. Graepel, "Private traits and Attributes are predictable from digital records of human behavior", *Proc. of the National Academy of Sciences*, vol. 110, no. 15, pp. 5802–5805, 2013.
- [5] B. Verhoeven and W. Daelemans, "CLiPS Stylometry Investigation (CSI) corpus: a Dutch corpus for the detection of age, gender, personality, sentiment and deception in text", in *Proceedings of the 9th International Conference on Language Resources and Evaluation*, Reykjavik, Iceland, ACL, 2014, pp. 3081–3085.
- [6] T. Litvinova, O. Litvinova, O. Zagorovskaya, P. Seredin, A. Sboev, and O. Romanchenko, "RusPersonality": a Russian corpus for authorship profiling and deception detection", in *Proceedings of International FRUCT conference on Intelligence, Social Media and Web (FRUCT 2016)*, 2016, pp. 1–7.
- [7] E. D. Khomskaya, "The Neuropsychology of individual differences", *Journal of Russian & East European Psychology*, vol. 35, no. 5, pp. 22–34, 1997.
- [8] A. Gérard-Desplanches, C. Deruelle, S. Stefanini, C. Ayoun, V. Volterra, S. Vicari, G. Fisch, M. Carlier, "Laterality in persons with intellectual disability II. Hand, foot, ear, and eye laterality in persons with Trisomy 21 and Williams-Beuren syndrome", *Dev Psychobiol.*, vol. 48, no. 6, pp. 482–91, 2006.
- [9] S.M. Scharoun and P.J. Bryden, "Hand preference, performance abilities, and hand selection in children", *Front Psychol.*, vol. 5, p. 82, 2014.
- [10] A.L. Sirotyuk, *Neuropsychological and Psychophysiological Learning Component*. Moscow: Sfera, 2003.
- [11] N. N. Bragina and T. A. Dobrohotova, "Functional Asymmetries of the Person". 2nd ed. Moscow: Medicine, 1988.
- [12] T. V. Chernigovskaya, T. A. Gavrilova, A. V. Voinov, and K. N. Strel'nikov, "Sensorimotor and cognitive laterality profiles", *Human Physiology*, vol. 31, no. 2, pp. 142–149, 2005.
- [13] L. Ya. Balonov, V. L. Deglin, and T. V. Chernigovskaya, "Functional cerebral asymmetry in speech production", in *Sensory Systems. Sensory Processes and Hemispheric Asymmetry*. Leningrad: Nauka, 1985.
- [14] M. Papadatou-Pastou and D. Tomprou, "Intelligence and handedness: Meta-analyses of studies on intellectually disabled, typically developing, and gifted individuals", *Neurosci Biobehav Rev.*, vol. 56, pp. 151–65, 2015.
- [15] A. Kailer and C.K. Chung, "The Russian LIWC2007 dictionary", Austin, TX: LIWC.net, 2011.
- [16] C. K. Chung and J. W. Pennebaker, "The psychological function of function words", in *Social communication: Frontiers of social psychology*, K. Fiedler, Ed. New York: Psychology Press, 2007, pp. 343–359.
- [17] J. Pennebaker, R. Booth, R. Boyd, and M. Francis, "Linguistic Inquiry and Word Count: LIWC2015", Austin, TX: Pennebaker Conglomerates, 2015.
- [18] T. Litvinova, P. Seredin, O. Litvinova, and E. Ryzhkova, "Estimating the similarities between texts of right-handed and left-handed males and females", in *Jones G. et al. (eds.) Experimental IR Meets Multilinguality, Multimodality, and Interaction. CLEF 2017. Lecture Notes in Computer Science*, vol. 10456. Springer, Cham, pp. 119–124, 2017.



Tatiana A. Litvinova received her PhD from Voronezh State University, Voronezh, Russia. She is a founder and head of RusProfiling Lab, Voronezh State University, Voronezh, Russia. She is also a researcher in Kurchatov Institute, Moscow, Russia. She and her lab team are involved in the study of author profiling in Russian texts, text-based deception detection, author gender imitation, text-based suicide behavior prediction, etc. Tatiana Litvinova is in charge of collecting "RusPersonality" which is the largest Russian text corpus with rich metadata about their authors (gender, age, education, psychological traits, etc.). She is a member of the Russian Cognitive Linguists Association.



Ekaterina S. Ryzhkova is a PhD student in Voronezh State Pedagogical University, Voronezh. She is a member of RusProfiling research group. Her dissertation is related to the description of typical linguistic features of the texts written by people with different lateral preferences. She is also an assistant lecturer in Voronezh State University of Engineering Technologies.