DOI: https://dx.doi.org/10.18454/RULB.12.19

Litvinova T.A.1, Zagorovskaya O.V.2

1Kandidat filologicheskikh nauk, 2Professor, doktor filologicheskih nauk, Voronezhskij gosudarstvennyj pedagogicheskij universitet

Исследование выполнено при поддержке гранта Президента РФ для молодых российских ученых – кандидатов наук, проект № МК-4633.2016.6 «Диагностирование склонности личности к суицидальному поведению на основе анализа ее речевой продукции».

ЯЗЫКОВЫЕ ОСОБЕННОСТИ РУССКОЯЗЫЧНЫХ ТЕКСТОВ ЛИЦ, СОВЕРШИВШИХ СУИЦИД, И ЛИЦ С ВЫСОКИМ РИСКОМ АУТОАГРЕССИВНОГО ПОВЕДЕНИЯ

Один из перспективных направлений современных исследований является анализ речи с целью выявления психического состояния и оценки психического здоровья говорящего/пишущего. В последние годы наблюдается повышенный интерес к решению подобной задачи с применением методов и средств компьютерной лингвистики и интеллектуального анализа данных (data mining). Отдельной научной проблемой, далекой от своего решения и, несмотря на это, требующей консолидации усилий психологов, лингвистов и специалистов по интеллектуальному анализу данных, является проблема диагностирования склонности личности к аутоагрессивному поведению и (суициду как к крайней его форме) на основе лингвистического анализа речи. Эта проблема имеет не только теоретическую, но и очевидную практическую значимость и в последние годы активно изучается лингвистами и психологами. С применением методов natural language processing ученье анализируют тексты (преимущественно англоязычные) суицидентов и строят модели, позволяющие классифицировать текст как принадлежащий или не принадлежащий суициденту, а также выявляют особенности таких текстов. При этом если ранее анализировались преимущественно художественные тексты суицидентов, то в новейших работах ученье исследуют интернет-тексты (blogs, tweets, посты в соцсетях) лиц, совершавших суицид либо выражающих намерение его совершить. Русский язык долгое время оставался на периферии подобных исследований. В статье представлены результаты исследований, направленных на выявление языковых особенностей русских текстов лиц, совершивших законченный суицид, а также лиц, склонных к аутоагрессивному поведению. В указанных исследованиях применялись методы и приёмы корпусной лингвистики, компьютерной лингвистики, статистического анализа. Обозначены перспективы дальнейших исследований.

Ключевые слова: суицидальное поведение, корпус текстов, компьютерная лингвистика, диагностирование личности по тексту, лингвистические методы оценки психического здоровья.

Introduction
The problem of personality profiling of authors of written texts has been dealt with for over several decades, but recently there has been a growing interest to it due to a rapid growth of Internet communication and an increasing need for the methods allowing one based on the quantitative analysis of anonymous and pseudoanonymous online texts to trace the personality (gender, age, education level, native language, psychological traits, etc.) of their authors. Psychologists and linguists have been leading the way in text-based personality profiling (see [2] for more detail).

As part of studies of text-based personality profiling, attempts have been made to diagnose certain mental diseases (depression, schizophrenia, bipolar disorder, etc.) in authors of written texts [6; 8; 9]. Another problem waiting to be addressed as part of the joint effort by psychologists, linguists and data mining experts is that of detecting suicidal tendencies in individuals based on their speech. That certainly is of both theoretical and practical significance.
Method

Over 800000 people are reported to die of suicide annually [20] with only 30% having previously stated their suicidal intentions [20]. Therefore there is a pressing need to develop methods to identify individuals displaying suicidal tendencies and thus to prevent them from committing suicides. Linguistic analysis appears to be one of the ways to tackle that [8]. Lately there has been a lot of focus on analyzing Internet texts related to suicide (blogs, tweets, etc.) [7, 17, 18]. However, these are mostly texts regarding suicide that are analyzed rather than Internet texts by individuals who committed suicide (those few that are analyzed are written by one person, i.e. case study, see [14] for example). It should also be noted that most studies dealing with linguistic analysis of suicidal individuals have been performed on English-language texts. However, as it is rightly pointed out in [8], in order to address the problem, it is crucial to make use of other languages to be able to identify culturally (and linguistically) universal suicide predictors. It is also essential that related studies are interdisciplinary.

Another way to go about addressing the issue is to identify individuals with high risks of autoaggressive behavior that turns into suicidal one in its extreme form.

Russian texts have long not been analyzed as part of the problem except fiction texts [10]; [12]. Russian Internet texts by suicidal individuals were first analyzed in [16]. Linguistic features of Russian texts by individuals with high risks of autoaggressive behavior were investigated in [3]. This paper summarizes the previous findings and outlines directions for future research.

Results and Discussion

1. Linguistic features of Russian fiction texts by individuals who committed suicide

Texts by Russian poets (translated into English) as well as those by poets of other nationalities were researched by S. Stirman, J. Pennebaker [22] to compare them with those by poets who did not end up committing suicide regardless of the author’s nationality and native language. Suicidal poets were generally found to use more pronouns “I” and fewer words describing social interaction.

Texts by Russian suicidal poets were investigated as part of a special study [10]. Linguistic parameters (labeling methods are not mentioned in the article) were those used by S. W. Stirman, J. W. Pennebaker [22]. Davidson found, however, that the proportion of pronouns “I” and corresponding object pronouns were steadily on the rise in texts by suicidal poets and is not stably high, while in texts by the control group it goes down. Additionally, the number of negations (no, not) was analyzed and their proportion was found to increase in texts by suicidal individuals and to decrease over time in those by the control group.

The authors of [12] designed the classifiers to distinguish the fiction texts by Russian suicidal poets and those of the control group. The classifier based on a full set of parameters (word n-grams, relative frequencies of parts of speech, punctuation marks, word length, etc.) was shown to be most effective (F-measure = 0.825). Unfortunately, no analysis of the differences of texts by suicidal individuals and those of the control group was carried out.

2. Linguistic features of Russian online texts by individuals who committed suicide

In [16] the results of a pilot study of Internet texts – online diaries (on the LiveJournal platform) — by individuals who committed suicide (SUI corpus) are described. 45 such diaries were found by means of manual search and further checking. There is a total of 196037 words in the SUI corpus.

Texts for comparison (i.e. those by the control group) were samples of writing of students of Russian universities that make up RusPersonality corpus [15] with the total of 198045 words (NON-SUI corpus).

All the texts were labeled using the LIWC software [19] with the users’ dictionaries compiled by the authors with the total of 104 parameters. Statistically significant differences were found between the parameters of texts by suicidal individuals and those of the control group. A series of operations was performed in order to select the properties and as a result, the classifier was designed with the accuracy of 71.5%. The approach set forth by the authors was shown to be highly accurate for text classification given that linguistic parameters that are maximum content-independent (proportion of commas, function words, etc.), which is indicative of how effective methods of natural language processing and data mining can be in identifying suicidal tendencies of individuals based on their texts.

The analysis suggests that in Russian texts by suicidal individuals there are more function words, verbs, conjunctions, cognitive words, commas, fewer prepositions, more comparison words and pronouns. 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 (see [16] for more detail).

Note that existing studies addressing linguistic analysis of suicides commonly rely on the sociological concept of suicide [11] according to which a suicidal individual is not capable of social integration and is excluded from society. According to the psychological concept of suicide a suicidal individual provokes the sense of hopelessness, despair and helplessness and a range of associated negative emotions [21]. Therefore one can expect there to be more words to describe the author and negative emotions in these texts.

Through the course of existing studies varying results (sometimes contradictory ones) have been obtained regarding the linguistic features of texts by suicides (see [16] for more detail), but the above theories were mostly shown to be correct. The analysis that we have conducted for Russian texts showed these theories to generally hold as well.

In [13] there is a hypothesis that suicidal poets see the world as unstable, undetermined, hostile, which is expressed with ontological and epistemological categories that are reflective of one’s inner world. Suicidal poets were found to use fewer words to describe motion, space, bodily state (world’s general characteristics); more words for negation and exclusion (relationship with the world); more words expressing uncertainty but fewer words for vision and perception overall. The authors [18] assume that it is the perception of the outside world as being hostile and incomprehensible that causes these individuals to shut themselves down from it and as a result, to become increasingly self-centered and isolated. These results are largely in agreement with those obtained for Russian blogs [16].

3. Linguistic features of texts by individuals with autoaggressive tendencies

The authors [3] have conducted a study to find out how possible it would be to identify personality traits of authors of written texts that might be personal determinants of autoaggressive behavior (suicidal behavior is one of its forms) using data on the neurobiological nature of individual
The scientific literature suggests that in individuals with suicidal tendencies the right-hemispheric modus of solution predominates for both verbal and visual-spatial problems, which is associated with the left prefrontal dysfunction (see [1]; [3] for more detail). At the same time based on studies of temporary inactivation of cerebral hemispheres it is known what parts of the brain are responsible for producing certain discourse units (e.g., abstract nouns, function words, complex syntactic structures) as well as what “language functions” can be performed by the right and left cerebral hemispheres (see review in [4]; [5]). The authors [3] assumed that in texts by individuals with high risks of autoaggressive behavior there are more language structures that the right hemisphere is responsible for than in those by people with no autoaggressive tendencies and that there are increasingly fewer structures that the left hemisphere is responsible for, particularly the left part of the prefrontal cortex.

The study material was a text corpus “RusPersonality” [15]. As was shown in [3], overall for texts by individuals with high risks of autoaggressive behavior lower lexical diversity, fewer prepositions, more pronouns overall particularly personal ones with a higher index of logical cohesion (created due to more conjunctions and deictic units) and a larger average word length are typical. The data on a lower index of lexical diversity are in agreement with those on a decreasing vocabulary level for hyperactivity of the right hemisphere, a lower proportion of prepositions is due to insufficient activation of the zones of the left hemisphere that is known to be responsible for most abstract vocabulary; a higher pronominisation index is commonly observed when paradigmatic language connections that rely on the back hemisphere get weakened. It was also shown that it is insufficient activation of the back hemisphere that is associated with aggressive and suicidal behavior [1].

Conclusion

According to the World Health Organization, suicide is one of the most common causes of death in young individuals (15-19 years old). Unfortunately, in Russia this problem is also relevant. The problem of suicide in teenagers has drawn a significant amount of attention from the state and public. In several countries systems of monitoring social networks and redirecting users to a psychological counseling website as well as mobile applications are developed for timely detection of individuals with high risks of suicidal behavior and counseling [9]. Methods of natural language processing are essential in designing such systems. It was not until recently that Russian texts started being investigated as part of this problem.

As the analysis suggests, texts by individuals who committed suicide and those with suicidal and, more broadly, autoaggressive tendencies have typical linguistic features. Certainly, the obtained data require further clarification first of all due to expanding language materials to be studied. One of the directions for future research is to analyze the dynamics of idiolect of individuals who committed suicide as well as indices of linguistic complexity of corresponding texts.

Список литературы / References

4. Петрова Т.Е. Особенности построения текста в аспекте функциональной асимметрии мозга: дис. ... канд. филол. наук / Т.Е. Петрова. – СПб., 2000.
6. Baddeley J.L. Email Communications Among People with and Without Major Depressive Disorder: Doctoral diss / J.L. Baddeley. – Austin, TX, 2011.
Introduction

In modern Russian literary language there is a special functional kind of speech, which is used in the professional and scientific sphere of communication. It is called by researchers in different ways: scientific style, language of science, language for special purposes, professional (special) language, academic discourse, public academic speech, etc.

The study of the genres of public academic speech was previously conducted solely on the basis of book-written sources, and the description of the features of the oral form of the academic discourse was carried out in isolation from the structure of the lingual personality and the cognitive-semantic aspect of the individual's speech activity. Studies of the linguistic and structural features of the genre forms of public academic speech and professional communication in its connection with the phenomena of extralinguistic reality and sociolinguistic factors have practically not been carried out.

The research hypothesis is that the description of the genre-forming features of public academic speech can serve as a basis for constructing models of speech genres and revealing the cognitive-semantic mechanisms of their text generation. Based on the preliminary study of the material, we assume that the speech models of the genre are systemic and ordered. For this reason we can identify their basic elements and structural modifications, depending on the influence of sociolinguistic and extralinguistic factors.

To this day in Russian and foreign studies of similar textual material on the basis of academic discourse, no parallels have been drawn between linguistic and non-linguistic factors. Obviously, therefore, it was difficult to establish the relationship between the intentional and structural-semantic plan for generating an utterance in the context of a certain genre of academic speech. The novelty of the problem also manifests itself in the anthropocentric and interdisciplinary approach that we propose simultaneously to the study of oral academic texts within the framework of socio- and pragma-linguistics.

Methods and principles of the study

The study is planned to be performed on the basis of audio and video records of Russian public academic speeches, not completely studied from the point of view of the structural and semantic features and genre typology.

The methodology of this study is determined by its main purpose — the study of the genre repertoire of public academic speeches based on the materials of their text transcripts. We take into account the specific functioning and interrelation of their structural-semantic, lexical and prosodic properties in the discourse of the lingual personality. The research is based on the results of a linguistic description of the phonetic features of oral academic speech [2], including its intonational division and the accentually prosodic aspect of the utterance [5, 6].

Such anthropocentric interdisciplinary approach assumes the use of the principle of linguistic expansionism, that is, the attraction in purely linguistic research of information from such fields of knowledge as socio- and pragma-linguistics, corpus linguistics, cognitive linguistics and functional stylistics. The study of the structural-semantic and composition-stylistic features of oral academic texts explains the use of the functional method of investigation. The need to compare the variants of oral texts of different genres with written analogs in the discourse of the lingual personality requires extensive use of the comparative method and principles of typological analysis. It is obvious that the appeal in the study to cultural-linguistic and social factors will result in an explanatory description.

Discussion

The written form of the language of science is sufficiently well studied and described. Thus, the linguistic