

УДК 81'33+616.89

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

Литвинова Т.А., Литвинова О.А.

Проблема выявления особенностей речи (как устной, так и письменной) лиц, находящихся в различных психических состояниях, и построения математических моделей для классификации текстов в зависимости от психического состояния автора относится к числу актуальных и активно разрабатываемых проблем современной компьютерной и психиатрической лингвистики. Большая часть подобных исследований выполнена на материале английского языка. В работе представлен обзор исследований, посвященных анализу письменных текстов лиц, совершивших суицид, а также лиц с высоким риском аутоагрессивного поведения, при этом особое внимание уделяется работам, выполненным на материале русского языка. Обозначены перспективы дальнейших исследований.

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

**LINGUISTIC PARAMETERS OF RUSSIAN TEXTS
BY SUICIDES AND PERSONS WITH A HIGH RISK
OF SELF-DESTRUCTIVE BEHAVIOUR**

Litvinova T.A., Litvinova O.A.

The problem of identifying the features of speech (both oral and written) of persons in different mental states and constructing mathematical models for classifying texts depending on the author's mental state is one of the topical and actively developed problems of modern computer and psychiatric linguistics. Most of these studies are based on the material of the English language. The work presents a

review of the research devoted to the analysis of written texts of persons who commit suicide as well as persons with a high risk of self-destructive behaviour, with special attention paid to work done on the material of the Russian language. Prospects for further research are also outlined.

Keywords: suicidal behavior, corpus of texts, computer linguistics, personality recognition from text, linguistic methods of mental health assessment, Internet text, social networks.

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».

Проблема диагностирования характеристик личности автора письменного текста изучается исследователями на протяжении нескольких десятилетий, но в последнее время в мировой науке наблюдается особый интерес к данной проблеме вследствие стремительного развития интернет-коммуникации и возросшей потребности в методиках, позволяющих на основе количественного анализа анонимного и псевдоанонимного онлайн-текста воссоздать облик (пол, возраст, уровень образования, родной язык, психологические характеристики и др.) его автора.

Приоритет в исследовании проблемы диагностирования личности по тексту принадлежит психологам и лингвистам. Однако уже с 1990-х гг. к решению данной проблемы подключаются математики и специалисты по информационным технологиям, начинается активное использование методов математической статистики, компьютерной лингвистики, в частности средств автоматической обработки языка (NLP), что позволяет быстро обрабатывать большие массивы текстового материала. На основе найденных корреляций между численными значениями поддающихся квантификации лингвистических параметров текста и характеристиками авторов исследователями строятся

математические модели и разрабатываются программные средства для автоматизированного диагностирования характеристик личности по тексту. При этом подчеркивается особая значимость формально-грамматических параметров текста, которые не контролируются автором и, следовательно, наименее подвержены сознательному искажению (доли служебных слов, биграммы и триграммы частей речи и т.д.) (см. подробнее в работе Т.А. Литвиновой [2]). Заметим, однако, что подавляющее большинство подобных исследований выполнено на материале английского языка.

В рамках исследований в области моделирования личности по тексту предпринимаются и попытки диагностировать наличие/отсутствие у автора письменного текста того или иного психического заболевания (депрессии, шизофрении, биполярного расстройства и т.д.). При этом, как показывают исследования, анализ исключительно лексического уровня не позволяет сделать достоверные выводы о психологическом статусе индивида. Так, в работе J.L. Baddeley [6], выполненной на материале текстов электронных писем, было показано, что люди, страдающие депрессией, используют больше слов, обозначающих положительные эмоции, чем лица в контрольной группе, видимо, не желая показывать свое истинное психологическое состояние. Совершенно очевидна необходимость анализ текста на разных языковых уровнях, а не только на лексическом, который, кроме того, легко поддается имитации. В последние годы в мировой науке наблюдается повышенный интерес к решению задач подобного рода с привлечением специалистов по компьютерной лингвистике и интеллектуальному анализу данных (data mining). Результаты подобных междисциплинарных исследований ежегодно обсуждаются на специальном воркшопе *Computational Linguistics and Clinical Psychology – From Linguistic Signal to Clinical Reality* (<http://clpsych.org>), который проводится с 2014 года в рамках самой престижной международной конференции по компьютерной лингвистике *Conference of the Association for Computational Linguistics* (ACL).

Отдельной научной проблемой, далекой от своего решения и, несомненно, требующей консолидации усилий психологов, лингвистов и специалистов по интеллектуальному анализу данных, является проблема диагностирования склонности личности к суициду на основе анализа речевой продукции. Эта проблема имеет не только теоретическую, но и очевидную практическую значимость. По данным, приведенным в работе J. Pestian и др. [20], более 800000 человек умирают ежегодно вследствие суицида, при этом только 30% совершивших суицид ранее сообщали о своих намерениях. Следовательно, существует объективная потребность в разработке методик, направленных на выявление лиц, склонных к суицидальному поведению, и предотвращение суицида. Одним из направлений работ в русле указанной проблемы может стать анализ речи (см. обстоятельный обзор исследований по данной теме в статье R. Calvo и др. [8]).

В последние годы исследователями уделяется большое внимание анализу интернет-текстов, посвященных суициду (блоги, твиты и др.) [7; 17; 18]. Однако в основном анализируются тексты, связанные с тематикой суицида, и очень ограниченное число работ посвящено непосредственно анализу интернет-текстов лиц, совершивших законченный суицид. Преимущественно проводится анализ текстов одного человека (case study) (см., например, работу T.M. Li и др. [14]). В то же время изучение такого рода текстов с применением современных методов *natural language processing* и *data mining* позволит получить данные, на основе которых в дальнейшем может быть разработан новый инструментарий для выявления лиц со склонностью к суицидальному поведению, который поможет практикующим специалистам-психологам в их повседневной работе.

Следует отметить также, что большинство работ, посвященных анализу речи суицидентов, и постов, связанных с тематикой суицида, выполнено на англоязычном материале. Однако, как справедливо отмечается в работе R. Calvo и др. [8], для решения указанной проблемы крайне необходимы исследования на материале других языков для поиска культурно (и

лингвистически) обусловленных и универсальных предикторов суицида. Отмечается также крайняя важность междисциплинарного подхода к решению указанной задачи.

Также с позиции профилактики суицида чрезвычайно важно выявлять лиц с высоким риском аутоагрессивного поведения, крайней формой которого является суицидальное поведение.

На протяжении длительного времени тексты на русском языке не анализировались в русле описанной выше проблематики, за исключением художественных текстов (см. работы Ch. Davidson [10]; S. Ermakov и L. Ermakova [12]). Впервые интернет-тексты суицидентов на русском языке были проанализированы в статье Т.А. Литвиновой и др. [16]. Языковые особенности русскоязычных текстов лиц с высоким риском аутоагрессивного поведения были исследованы в работе «Диагностирование склонности автора письменного текста к аутоагрессивному поведению» [3]. В настоящей работе нами обобщены полученные ранее результаты, а также намечены пути дальнейших исследований.

1. Языковые особенности художественных русскоязычных текстов лиц, совершивших суицид

Тексты русских поэтов (в переводе на английский), наряду с текстами поэтов других национальностей, исследовались в работе S. Stirman, J. Pennebaker [22], в которой ставилась задача выявления отличий текстов поэтов-суицидентов от текстов поэтов, не совершавших суицид, без учета национальности и родного языка автора. Было выявлено, что в целом суициденты употребляли больше местоимений «я» и меньше слов, обозначающих взаимодействие с обществом.

Тексты русских поэтов-суицидентов стали объектом специального исследования Ch. Davidson «Comparative Psychological Analysis of Six Russian Poets» [10]. В качестве параметров исследования (о методах разметки текста в статье не сказано) были взяты параметры, по которым были обнаружены различия в работе S.W. Stirman, J. W. Pennebaker [22]. Было также выявлено,

что в текстах суицидентов меньше слов, обозначающих разные виды взаимодействия людей. Однако Ch. Davidson было установлено, что доля местоимений «я» (и косвенных форм) в текстах суицидентов увеличивается со временем, а не является стабильно высокой, а в текстах контрольной группы – уменьшается. Кроме того, было проанализировано число отрицаний (не, нет) и выявлено, что их доля растет в текстах суицидентов и уменьшается в текстах контрольной группы с течением времени.

Авторы работы «Linguistic Approach to Suicide Detection» [12] создавали классификаторы для разграничения художественных текстов русских поэтов-суицидентов и поэтов из контрольной группы. Наилучшую эффективность ($F\text{-measure} = 0.825$) продемонстрировал классификатор на основе полного набора параметров (n-граммы слов, относительные частоты частей речи, знаков препинания, длина слов и т.д.). К сожалению, в работе не проводился анализ различий текстов суицидентов и лиц из контрольной группы.

2. Языковые особенности онлайн-текстов лиц, совершивших суицид

В работе Т.А. Литвиновой и др. «Identification of Suicidal Tendencies of Individuals Based on the Quantitative Analysis of Their Internet Texts» [16] описаны результаты пилотного исследования интернет-текстов – онлайн-дневников (на платформе *Live Journal*) лиц, совершивших законченный суицид (далее корпус СУИЦ). Путем ручного поиска и последующей проверки было обнаружено 45 таких дневников. Общий объем текстов корпуса СУИЦ составил 196037 слов.

В качестве материала для сравнения (т.е. текстов лиц контрольной группы) были использованы тексты – образцы естественной речи студентов российских университетов, входящие в корпус *RusPersonality* [15], общий объем текстов составил 198045 слов (далее корпус НСУИЦ).

Все тексты были размечены при помощи программы LIWC [19] с созданными авторами пользовательскими словарями для русского языка, всего 104 параметра. Были выявлены статистически значимые различия между параметрами текстов суицидентов и лиц контрольной группы, произведен ряд

операций по отбору признаков и построен классификатор, точность которого составила 71,5%. Предложенный авторами подход показал достаточно высокую точность классификации текстов, учитывая, что были использованы языковые параметры, максимально не зависящие от содержания текста (доля запятых, неполнозначных слов и т.д.), что показывает перспективность применений методов *natural language processing* и *data mining* для решения задач диагностирования по тексту склонности авторов к суицидальному поведению.

Как показал анализ, в русскоязычных текстах суицидентов больше неполнозначных слов, глаголов, союзов, слов, описывающих когнитивные процессы, больше запятых, меньше предлогов, больше слов, указывающих на сравнение, больше местоимений. Тексты суицидентов более абстрактны, в них меньше отсылок к пространственным отношениям.

Также было установлено, что в текстах суицидентов больше слов, обозначающих отрицательные эмоции, меньше слов, обозначающих социальные отношения и процессы восприятия (особенно визуального), что также указывает на сосредоточенность суицидентов на своих мыслях и отрешенность от окружающего мира (см. подробнее [16]).

Заметим, что существующие исследования, посвященные изучению речи суицидентов, как правило, опираются на социологическую концепцию суицида (см. известную работу Е. Durkheim [11]), согласно которой он не может интегрироваться в общество и исключен из социальной жизни, либо же психологической концепцией суицида, согласно которой суицид провоцируют чувства безнадежности, тоски и беспомощности и вызываемые ими отрицательные эмоции (см. подробнее работу К. Petrie и R. Brook [21]). Следовательно, в текстах суицидентов можно ожидать более частое использование слов, обозначающих автора, и слов, номинирующих отрицательные эмоции.

В существующих работах были получены различные (иногда противоречивые) результаты, касающиеся языковых особенностей текстов суицидентов, но следует отметить, что в большинстве из них указанные выше

теории находили подтверждение. Проведенный анализ показал, что на материале русскоязычных текстов суицидентов указанные теории в целом находят подтверждение.

В работе «Escaping the World: Linguistic Indicators of Suicide Attempts in Poets» [13] была выдвинута гипотеза о том, что поэты-суициденты воспринимают мир как нестабильный, неопределенный, недружелюбный, что может выражаться в использовании слов онтологических и эпистемиологических категорий, т.е. категорий, отражающих мироощущение индивида. Было выявлено, что суициденты использовали меньше слов, описывающих движение, пространство, телесность (общие характеристики мира); больше слов, обозначающих отрицание и исключение (отношения с миром); больше слов, обозначающих неуверенность, но меньше слов, обозначающих видение и процессы восприятия в целом. Авторы предполагают, что именно восприятие внешнего мира как недружелюбного, непонятного ведет к избеганию к общению с ним, сосредоточенности на себе и социальной изоляции. Во многом эти результаты согласуются с полученными на материале русскоязычных блогов [16].

3. Языковые особенности текстов лиц, склонных к аутоагрессивному поведению

В статье «Диагностирование склонности автора письменного текста к аутоагрессивному поведению» [3] представлены результаты исследования, посвященного оценке возможностей диагностирования психологических характеристик автора письменного текста, которые могут являться личностными детерминантами аутоагрессивного поведения (одной из форм которого является суицидальное поведение), с привлечением данных о нейробиологической природе индивидуальных характеристик, с одной стороны, и мозговых механизмах порождения дискурса – с другой.

Известно, что у лиц, склонных к суицидальному поведению, преобладает правополушарный модус решения задач – как вербальных, так и зрительно-пространственных, что связывают с левой префронтальной дисфункцией (см.

подробнее работу А.Ю. Егорова, О.В. Иванова [1]). В то же время из исследований по временной инактивации полушарий мозга известно, какие отделы мозга ответственны за продуцирование тех или иных единиц дискурса (например, абстрактных существительных, служебных слов, сложных синтаксических конструкций), а также о том, какие «языковые функции» способны выполнять правое и левое полушария мозга (см. обзор в диссертации Т.Е. Петровой [4], пособия Т.А. Фотековой, Т.В. Ахутиной [5]).

В статье «Диагностирование склонности автора письменного текста к аутоагрессивному поведению» [3] мы предположили, что в текстах лиц с высоким риском аутоагрессивного поведения будет наблюдаться большее число языковых структур, за продуцирование которых ответственно правое полушарие, чем в текстах лиц, не склонных к аутоагрессивному поведению. Соответственно, будет наблюдаться снижение числа структур, в продуцировании которых принимает участие левое полушарие, в частности левая часть префронтальной коры.

Материалом для исследования послужил корпус текстов *RusPersonality*, снабженный метаразметкой в виде данных об их авторах [15]. Из корпуса были сделаны две выборки. В первую выборку вошли тексты лиц, склонных к аутоагрессивному поведению, т.е. имеющих высокие оценки (7-9) по шкалам «Спонтанная агрессивность», «Депрессивность», «Эмоциональная лабильность» и низкие оценки (1-3) по шкале «Уравновешенность» Фрайбургского личностного опросника (FPI), N = 33. Во вторую группу вошли тексты лиц, имеющих соответственно низкие оценки по шкалам «Спонтанная агрессивность», «Депрессивность», «Эмоциональная лабильность» и высокие по шкале «Уравновешенность», N = 27. Каждый респондент писал по два текста. Для анализа был выбран ограниченный набор параметров текста, легко поддающихся квантификации. Для того чтобы уйти от зависимости от длины текста, мы использовали только относительные величины, т.е. различные соотношения параметров. Все тексты были размечены при помощи морфологического парсера и других программных средств, далее было

произведено извлечение числовых значений выбранных параметров текста. Данные для расчетов были занесены в Excel, экспортированы в программу *SPSS Statistics*, где был произведен корреляционный анализ между числовыми значениями выбранных параметров текста и баллами по шкалам теста (отдельно для каждой шкалы), $p < 0,05$.

Как показало наше исследование [3], в целом для текстов лиц с высокой склонностью к аутоагрессивному поведению характерно меньшее лексическое разнообразие, меньшее число предлогов, большее число местоимений в целом и личных в частности при более высоком индексе логической связанности (создаваемом за счет большего количества союзов и дейктических частиц) и большей средней длине предложения. Данные о более низком индексе лексического разнообразия согласуются с данными об уменьшении лексического запаса при гиперактивности правого полушария, более низкая доля предлогов объясняется недостаточной активацией зон левого полушария, ответственного, как известно, за продуцирование наиболее абстрактных лексических единиц. Более высокий индекс прономинализации, как правило, наблюдается при ослаблении парадигматических связей языка, опирающихся на работу задних отделов мозга (см. [5]). Между тем в ряде исследований установлено, что именно недостаточная активность задних отделов левого полушария связана с агрессивным и суицидальным поведением [9].

Подводя итоги, следует отметить, что по данным ВОЗ, суицид является одной из наиболее частых причин смерти среди лиц молодого возраста (15-19 лет). В России данная проблема, к сожалению, также актуальна. В последнее время проблема подросткового суицида привлекла повышенное внимание власти и общественности. В ряде стран разрабатываются системы мониторинга соцсетей и перенаправления пользователя на сайт психологической помощи, а также разрабатываются мобильные приложения для раннего выявления лиц с высоким риском суицидального поведения и оказания им помощи [9]. При этом важную роль в построении таких систем играют методы *natural language*

processing. До недавнего времени тексты на русском языке не исследовались в указанном аспекте.

Как показал проведенный нами анализ, тексты лиц, совершивших суицид, и лиц, склонных к суицидальному, и шире – аутоагрессивному поведению, имеют характерные языковые особенности. Безусловно, полученные данные нуждаются в уточнении, прежде всего за счет расширения объема языкового материала. Одним из направлений дальнейших исследований нам представляется анализ динамики идиостиля лиц, совершивших суицид, в том числе анализ показателей языковой сложности соответствующих речевых произведений.

Список литературы:

1. Егоров А.Ю., Иванов О.В. Особенности индивидуальных профилей функциональной асимметрии у лиц, совершивших суицидальную попытку // Социальная и клиническая психиатрия. 2007. № 2 (17). С. 20-24.
2. Литвинова Т.А. Установление характеристик (профилирование) автора письменного текста // Филологические науки. Вопросы теории и практики. 2012. № 2 (13). С. 90-94.
3. Литвинова Т.А., Середин П.В., Литвинова О.А., Загоровская О.В., Сердюк М.Е. Диагностирование склонности автора письменного текста к аутоагрессивному поведению // Вестник Воронежского гос. ун-та. Серия: Лингвистика и межкультурная коммуникация. 2015. № 3. С. 98-104.
4. Петрова Т.Е. Особенности построения текста в аспекте функциональной асимметрии мозга: дис. ... канд. филол. наук: спец. 10.02.19 – теория языка. СПб., 2000.
5. Фотекова Т.А., Ахутина Т.В. Диагностика речевых нарушений школьников с использованием нейропсихологических методов: пос. для логопедов и психологов. М.: АРКТИ, 2002.

6. Baddeley J.L. Email communications among people with and without major depressive disorder. Unpublished doctoral dissertation. Austin, TX: University of Texas at Austin, 2011.

7. Barak A., Miron O. Writing characteristics of suicidal people on the Internet: a psychological investigation of emerging social environments // *Suicide and Life-Threatening Behavior*. 2005. № 35 (5). P. 507-524.

8. Calvo R., Milne D., Hussain M., Christensen H. Natural language processing in mental health applications using non-clinical texts // *Natural Language Engineering*. 2017. № 23 (5). P. 649-685.

9. Christensen H. E-Health Interventions for Suicide Prevention // *International Journal of Environmental Research and Public Health*. 2014. Vol. 11 (8). P. 8193-8212.

10. Davidson Ch. Comparative Psychological Analysis of Six Russian Poets // *US-China Foreign Language*. 2013. Vol. 11 (1). P. 40-45.

11. Durkheim E. *Suicide*. New York: Free Press, 1951.

12. Ermakov S., Ermakova L. Linguistic Approach to Suicide Detection // *Труды ИСП РАН*. 2014. Т. 26. Вып. 4. С. 113-121.

13. Katarzyna P., Trzebiński J. Escaping the World: Linguistic Indicators of Suicide Attempts in Poets // *Journal of Loss and Trauma: International Perspectives on Stress & Coping*. 2014. Vol. 19 (5). P. 389-402.

14. Li T.M., Chau M. Yip P.S., Wong P.W. Temporal and computerized psycholinguistic analysis of the blog // *Crisis*. 2014. Vol. 35 (3). P. 168-75.

15. Litvinova T., Litvinlova O., Zagorovskaya O., Seredin P., Sboev A., Romanchenko O. "RusPersonality": A Russian corpus for authorship profiling and deception detection // *Proceedings of International FRUCT Conference on Intelligence, Social Media and Web (ISMW FRUCT)*. St. Petersburg, 2016. P. 1-7.

16. Litvinova T.A., Seredin P.V., Litvinova O.A., Zagorovskaya O.V. Identification of Suicidal Tendencies of Individuals Based on the Quantitative Analysis of Their Internet Texts // *Computación y Sistemas*. 2017. Vol. 21(2). P. 243-252.

17. Masuda N., Kurahashi I., Onari H. Suicide Ideation of Individuals in Online Social Networks [Электронный ресурс] // PloS One. 2013. Vol. 8. Issue 4. <https://goo.gl/3DxoYM> (дата обращения 15.12.2017).
18. O'Dea B., Wan S., Batterham P.J., Caelear A.L., Paris C., Christensen H. Detecting suicidality on Twitter // Internet Interventions. 2015. Vol. 2 (2). P. 183-188.
19. Pennebaker J.W. The development and psychometric properties of LIWC2007. Austin, TX: LIWC.net, 2007.
20. Pestian J. et al. Suicide Note Classification Using Natural Language Processing: A Content Analysis // Biomed Inform Insights. 2010. Vol. 3. P. 19-28.
21. Petrie K., Brook R. Sense of coherence, self-esteem, depression and hopelessness as correlates of reattempting suicide // British Journal of Clinical Psychology. 1992. Vol. 31. P. 293-300.
22. Stirman S., Pennebaker J. Word use in the poetry of suicidal and nonsuicidal poets // Psychosomatic Medicine. 2001. Vol. 63. P. 517-522.

Сведения об авторах:

Литвинова Татьяна Александровна – кандидат филологических наук, научный сотрудник Регионального центра русского языка, руководитель лаборатории корпусной социолингвистики и автороведения Воронежского государственного педагогического университета (Воронеж, Россия).

Литвинова Ольга Александровна – аспирант Воронежского государственного педагогического университета (Воронеж, Россия).

Data about the authors:

Litvinova Tatiana Aleksandrovna – Candidate of Philological Sciences, Staff Scientist of the Russian Language Regional Centre, Head of the Laboratory for Corpus Sociolinguistics and Authorship Profiling, Voronezh State Pedagogical University (Voronezh, Russia).

Litvinova Olga Aleksandrovna – graduate student of Voronezh State Pedagogical University (Voronezh, Russia).

E-mail: centr_rus_yaz@mail.ru.

E-mail: olga_litvinova_teacher@mail.ru.