

УДК 81'33+ 616.89

**ВЫЯВЛЕНИЕ СКЛОННОСТИ ЛИЧНОСТИ  
К СУИЦИДАЛЬНОМУ ПОВЕДЕНИЮ НА ОСНОВЕ  
КОЛИЧЕСТВЕННОГО АНАЛИЗА ЕЕ РЕЧЕВОЙ ПРОДУКЦИИ**

**Загоровская О.В., Литвинова О.А., Литвинова Т.А.**

Суицид является одной из ведущих причин смертности в большинстве стран мира, причем большинство совершивших суицид при жизни не сообщали о своих намерениях. В связи с этим особую актуальность приобретает проблема разработки методик выявления лиц, склонных к суицидальному поведению. Одним из направлений таких исследований является поиск типологических особенностей речи суицидентов с применением методов математической лингвистики и автоматической обработки текстов. Большинство работ, посвященных этой проблеме, выполнены на материале английского языка. В статье представлен анализ указанных работ, и намечены перспективы решения данной проблемы на материале русского языка.

**Ключевые слова:** диагностирование личности по тексту, суицид, суицидальное поведение, психиатрическая лингвистика, количественные методы в языкознании, корпусная лингвистика, автоматическая обработка текстов.

**IDENTIFICATION OF SUICIDAL TENDENCIES  
IN INDIVIDUALS USING A QUANTITATIVE ANALYSIS  
OF THEIR SPEECH PRODUCTION**

**Zagorovskaya O.V., Litvinova O.A., Litvinova T.A.**

Suicide is one of the top factors contributing to deaths around the globe with individuals not making other people aware of their suicidal plans. Therefore it is of increasing importance to develop the methods of identification of these individuals. One of the directions of the ongoing research is identification of typological features of their speech patterns using the methods of mathematical linguistics and automatic text processing. Most studies addressing the problem use materials written in English.

The article presents the analysis of the above studies and points out the ways of dealing with the issue employing materials written in Russian.

**Keywords:** author profiling using texts, suicide, suicidal behavior, psychiatric linguistics, quantitative methods in language studies, corpus linguistics, automatic text processing.

*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».*

По данным, приводимым Всемирной организацией здравоохранения, более восьмисот тысяч человек умирают ежегодно вследствие суицида, т.е. каждые 40 секунд совершается самоубийство, при этом, по имеющимся данным, только 30 % совершивших суицид ранее сообщали о своих намерениях [4]. Следовательно, существует объективная потребность в разработке методик, направленных на выявление лиц, склонных к суицидальному поведению, и предотвращение суицида. Ценнейшим диагностическим инструментом, позволяющим выявить особенности психики личности, в том числе ее склонность к суицидальному поведению, является анализ ее речевой продукции, в том числе на формально-грамматическом уровне, не поддающемся контролю сознанию.

Проблема диагностирования характеристик личности по ее речевой продукции изучается исследователями на протяжении нескольких десятилетий. В последнее время в мировой науке наблюдается особый интерес к данной проблеме вследствие стремительного развития интернет-коммуникаций и в связи с возросшей потребностью в методиках, позволяющих на основе количественного анализа анонимного и псевдоанонимного текста воссоздать облик (пол, возраст, уровень образования, родной язык, психологические характеристики и др.) его автора [2].

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.) [2]. Заметим, однако, что подавляющее большинство подобных исследований выполнено на материале английского языка.

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 показывает анализ научной литературы, контент-анализ речевой продукции не позволяет сделать достоверные выводы о психологическом статусе индивида. Так, в работе J.L. Baddeley [5], выполненной на материале текстов электронных писем, было показано, что страдающие депрессией люди использовали больше слов, обозначающих положительные эмоции, чем в контрольной группе, видимо, не желая показывать свое истинное состояние. Совершенно очевидно, что для выявления психологического состояния автора текста необходим анализ речевой продукции на разных уровнях ее порождения, а не только на лексическом, который, кроме того, легко поддается имитации.

Очевидно, что для выявления психологических характеристик личности, в том числе ее склонности к суицидальному поведению, необходим комплексный психолингвистический анализ ее речевой продукции с опорой на данные нейролингвистики, нейропсихологии индивидуальных различий, нейробиологии суицидального поведения, с применением современных методов автоматической обработки языка и с использованием корпусных технологий. Изучение разножанровой письменной речевой продукции суицидентов, созданной ими в разные периоды жизни, с целью выявления языковых предикторов суицидального поведения – изменений, происходящих на различных уровнях текста как продукта речемыслительной деятельности, по мере прогрессирования суицидальных тенденций, в сравнении с речевой продукцией контрольной группы – лиц, максимально схожих по уровню образования, социальному статусу и другим характеристикам, но не совершавших суицид, позволит в дальнейшем разработать диагностические инструменты, направленные на оценку вероятности суицидального поведения личности на основе количественных параметров ее речевой продукции.

Ученые, занимающиеся проблемой выявления особенностей текстов суицидентов, анализируют преимущественно тексты предсмертных записок. Создаются корпуса текстов этого жанра на разных языках, и с применением современных методов компьютерной лингвистики и математической статистики анализируются их формально-грамматические характеристики. Это – средняя длина предложений, доля слов разных частей речи и т.п., и содержательный уровень (доля слов, обозначающих положительные и отрицательные эмоции; слов, указывающих на время, место и т.д.). Предпринимаются попытки создать математические модели, разграничивающие настоящие и поддельные тексты этого жанра на основе количественных параметров текста.

Несмотря на бесспорную важность изучения предсмертных записок, они, вследствие малого объема, не дают возможности изучить в полной мере особенности речевой продукции суицидентов. Таким образом, специалистами

была осознана необходимость анализа текстов разных жанров, созданных лицами, впоследствии совершившими суицид, в сравнении с текстами лиц, не совершавших суицид (с учетом жанра текста, демографических и др. характеристик автора), в том числе в динамике с целью выявления изменений в идиостиле по мере приближения трагической развязки. Такого рода работы немногочисленны, что во многом объясняется трудностями, связанными с подготовкой исследовательских корпусов текстов.

В проанализированных нами работах отмечается, в частности, что качественные методы анализа текстов суицидентов, используемые психологами и психиатрами, с необходимостью должны быть дополнены количественными методами анализа, основанными на использовании компьютерных программ. Так в работе с использованием программы LIWC [14], позволяющей вычислять в тексте доли в тексте слов разных частей речи, тех или тематических групп лексики и т.д., было выявлено, что в поэтических текстах суицидентов, созданных в разные периоды, чаще по сравнению с текстами контрольной группой используется местоимений «я». С течением времени в текстах суицидентов наблюдалось уменьшение доли местоимений «мы», доли глаголов, связанных с разными видами взаимодействия (*e.g., talk, share, listen*), но не наблюдалось, вопреки ожиданиям, увеличения доли слов, обозначающих отрицательные эмоции (по этому параметру статистически значимых различий между суицидентами и контрольной группой выявлено не было). Авторы считают, что их результаты подтверждают одну из психологических теорий суицидогенеза, связывающих суицидальное поведение с возрастающей разобщенностью с другими людьми.

Исследование, материалом которого послужило около 300 поэтических текстов 18 американских и русских поэтов – суицидентов и не совершавших суицид, получило широкое освещение в зарубежной прессе, хотя его методология вызывает ряд вопросов. В частности, анализировались не оригинальные русские тексты, а их переводы на английский язык. Кроме того, поэтические тексты зачастую подвергаются их авторами тщательному

редактированию, причем работа над ними может вестись на протяжении нескольких лет или даже десятилетий, что в контексте указанной проблемы снижает их ценность как материала исследования.

Оригинальные русскоязычные поэтические тексты 6 русских поэтов (трех покончивших жизнь самоубийством и трех не совершавших суицид) стали объектом специального исследования Ch. Davidson [6]. В качестве параметров исследования (о методах разметки текста в статье не сказано) были взяты параметры, которые, по данным S.W. Stirman, J.W. Pennebaker [14], различают тексты суицидентов и контрольной группы. Было также выявлено, что в текстах суицидентов меньше доля слов, обозначающих разные виды взаимодействия людей. Однако, как показало данное исследование, доля местоимений «я» (и косвенных форм) в текстах суицидентов увеличивается со временем, а не является стабильно высокой, а в текстах контрольной группы – уменьшается. Были выявлены и другие отличия от результатов, полученных S.W. Stirman, J.W. Pennebaker, что говорит о необходимости анализа оригинальных, а не переводных текстов в свете указанной проблемы. Кроме того, автором, в отличие от S.W. Stirman, J.W. Pennebaker, было проанализировано число отрицаний (*не, нет*), и было выявлено, что их доля растет в текстах суицидентов и уменьшается в текстах контрольной группы с течением времени. Автор также делает вывод о необходимости сравнения результатов, полученных на материале текстов суицидентов разных национальностей.

В работе [9] с использованием программ LIWC и Coh-Metrix анализировались тексты песен, написанные лицами, впоследствии совершившими суицид, в сравнении с текстами лиц контрольной группы, и было установлено, что суициденты использовали больше абстрактных слов, меньше слов в целом, больше глаголов будущего времени и меньше слов тематической группы «Смерть». Указанное исследование имеет, однако, свои ограничения: тексты песен часто являются плодом коллективного творчества, и

их анализ для изучения идиостиля конкретной личности не всегда является релевантным.

Таким образом, была осознана необходимость исследования не подвергавшихся редактированию текстов, представляющих собой образцы естественной письменной речи лиц, совершивших суицид.

В исследованиях последних лет с применением указанных программных средств анализируются тексты дневников, писем (в том числе, опубликованные в Интернет) и прижизненных интервью лиц, покончивших жизнь самоубийством [12], обзор работ такого рода представлен у D. Lester [8].

Отметим, что в указанных работах не ставится задача создания методик диагностирования склонности к суицидальному поведению на основе количественного анализа речевой продукции, а лишь констатируется наличие статистически значимых отличий в текстах суицидентов и лиц, не совершавших суицид. Основной инструмент анализа текстов, применяющийся в указанных работах, – программа LIWC.

Задача разработки математической модели, которая бы предсказывала вероятность суицидального поведения на основе анализа речевой продукции, ставится в финансируемом Министерством обороны США Durkheim Project [15], материалом которого служат тексты, в том числе интернет-коммуникации, ветеранов войны, часть которых впоследствии совершили суицид [16]. На основе разработанных моделей исследователи планируют в дальнейшем в онлайн-режиме выявлять лиц с повышенным риском суицидального поведения на основе анализа их языковых особенностей их текстов. В настоящее время в разработанной авторами проекта математической модели, основанной на методах машинного обучения, учитывается только частотность отдельных полнозначных слов и фраз (точность классификации автора текста как суицидента/несуицидента составила 65 %) [13]. Формально-грамматические параметры текста, неподконтрольные сознанию автора, не анализировались, между тем как информативность такого рода параметров для диагностирования

различных параметров личности (пол, возраст, психологические особенности и др.) была доказана в многочисленных исследованиях (см. обзор: [2; 3; 11]).

В работе Mulholland M., Quinn J. [10], выполненной на материале текстов песен, также ставится задача разработки математической модели, классифицирующей тексты как принадлежащие суицидентам либо лицам из контрольной группы. Однако в исследовании был значительно расширен список параметров текста (TTR, доля слов некоторых частей речи, доля глаголов в пассивном залоге, доля слов некоторых семантических классов, n-граммы слов). Тексты были размечены с использованием современных средств автоматической обработки языка. С использованием методов машинного обучения был построен классификатор, точность которого составила 70.6%. Безусловно, точность классификатора далека от желаемой, однако полученные результаты показывают принципиальную возможность решения проблемы диагностирования риска суицидального поведения на основе анализа количественного анализа текстов с применением методов NLP и математической статистики. Как справедливо отмечают авторы исследования, для улучшения точности модели необходимо расширение корпусов текстов и набора параметров для анализа.

Таким образом, как показал анализ научной литературы, большинство исследований, направленных на выявление типологических особенностей речевых произведений суицидентов с применением статистических методов и средств автоматической обработки текстов, выполнены на материале английского. Очевидна необходимость таких исследований и на материале других языков, и в частности русского. Для изучения особенностей текстов суицидентов на русском языке необходимо, прежде всего, создать корпуса текстов лиц, совершивших законченный суицид. В настоящее время такие корпуса текстов на русском языке отсутствуют. Также необходимо выработать принципы отбора параметров текстов для исследований.

Заметим, что при выборе параметров зарубежные исследователи ориентируются, прежде всего, на возможность автоматизированного



извлечения их численных значений при помощи существующих программных средств и, в ряде работ, на существующие в психологии теории, объясняющие суицидальное поведение. При этом игнорируются полученные как отечественными, так и зарубежными учеными данные о нейробиологических механизмах суицидального поведения [1; 7]. Логично ожидать, что эти особенности найдут свое отражение и в языке как продукте деятельности мозга, и в частности в тексте как продукте индивидуальной речемыслительной деятельности [3]:

- выбрать из существующих либо разработать программы для анализа исследовательского корпуса текстов;
- сформулировать математическую постановку задачи.

Исследования, направленные на выявление типологических особенностей речи суицидентов, позволят в дальнейшем разработать диагностические инструменты для оценки склонности личности к суицидальному поведению на основе лингвистического анализа ее речевой продукции.

### **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. Профилирование автора письменного текста // Язык и культура. 2013. № 3 (23). С. 64-72.
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. Предотвращение самоубийства: глобальный императив [Электронный ресурс] // Российское Общество Психиатров [сайт]. 2014. URL: <http://goo.gl/XNLQdn> (дата обращения: 13.03.2016).

5. Baddeley J.L. Email communications among people with and without major depressive disorder. Unpublished doctoral dissertation. Austin, TX: University of Texas at Austin, 2011.
6. Davidson Ch. Comparative Psychological Analysis of Six Russian Poets // *US-China Foreign Language*. 2013. Vol. 11. №. 1. P. 40-45.
7. Joiner T.E., Brown J.S., Wingate L.R.Jr. The psychology and neurobiology of suicidal behavior // *Annu Rev Psychol*. 2005. Vol. 56. P. 287-314.
8. Lester D., *The «I» of the Storm: Understanding the Suicidal Mind*. De Gruyter Open Ltd, 2014. 170 p.
9. Lightman E.J., McCarthy P.M., Dufty D.F., McNamara D.S., Using Computational Text Analysis Tools to Compare the Lyrics of Suicidal and Non-Suicidal Songwriters // D.S. McNamara & G. Trafton (Eds.). *Proceedings of the 29th Annual Cognitive Science Society*. Hillsdale, NJ: Erlbaum, 2007.
10. Mulholland M., Quinn J. Suicidal Tendencies: The Automatic Classification of Suicidal and Non-Suicidal Lyricists Using NLP // *International Joint Conference on Natural Language Processing*. Nagoya, Japan, 14-18 October 2013. P. 680-684.
11. Nini A., *Authorship Profiling in a Forensic Context*. PhD thesis. Aston University, 2014.
12. Pennebaker J.W., Stone L.D. What Was She Trying to Say? A Linguistic Analysis of Katie's Diaries // D. Lester (Ed.), *Katie's Diary: Unlocking the Mystery of a Suicide*. New York: Brunner-Routledge, 2004. P. 55-80.
13. Poulin Ch., Shiner B., Thompson P. et al. Predicting the Risk of Suicide by Analyzing the Text of Clinical Notes [Электронный ресурс] // *PLoS ONE*. 2014. № 9 (1). URL: <http://goo.gl/iBzTIR> (дата обращения: 13.03.2016).
14. Stirman S.W., Pennebaker J.W., Word Use in the Poetry of Suicidal and Non-Suicidal Poets // *Psychosom Med*. 2001. № 63 (4). P. 517-522.
15. The Durkheim Project [Электронный ресурс] // [Durkheimproject.org](http://Durkheimproject.org) [сайт]. 2016. URL: <http://goo.gl/koczD0> (дата обращения: 13.03.2016).

16. Thompson P. et al. Predicting military and veteran suicide risk: Cultural aspects // Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2014. P. 1-6.

**Сведения об авторах:**

Загоровская Ольга Владимировна – доктор филологических наук, профессор, заведующий кафедрой русского языка, современной русской и зарубежной литературы Воронежского государственного педагогического университета (Воронеж, Россия).

Литвинова Ольга Александровна – ассистент кафедры английского языка Воронежского государственного педагогического университета (Воронеж, Россия).

Литвинова Татьяна Александровна – кандидат филологических наук, научный сотрудник Регионального центра русского языка при Воронежском государственном педагогическом университете (Воронеж, Россия).

**Data about the authors:**

Zagorovskaya Olga Vladimirovna – Doctor of Philological Sciences, Professor, Head of Russian Language, Modern Russian and Foreign Literature Department, Voronezh State Pedagogical University (Voronezh, Russia).

Litvinova Olga Aleksandrovna – Assistant Lecturer of English Language Department, Voronezh State Pedagogical University (Voronezh, Russia).

Litvinova Tatiana Aleksandrovna – Candidate of Philological Sciences, Staff Scientist of the Russian Language Regional Centre, Voronezh State Pedagogical University (Voronezh, Russia).

**E-mail:** olzagor@yandex.ru

**E-mail:** olga\_litvinova\_teacher@mail.ru

**E-mail:** centr\_rus\_yaz@mail.ru