

УДК 81'33

# ИССЛЕДОВАНИЕ ДИНАМИКИ ИДИОСТИЛЯ СУИЦИДЕНТА<sup>1</sup>

**ЛИТВИНОВА Татьяна Александровна,**

кандидат филологических наук, научный сотрудник Регионального центра русского языка, Воронежский государственный педагогический университет;

**СЕРЕДИН Павел Владимирович,**

доктор физико-математических наук, доцент кафедры физики твердого тела и наноструктур, Воронежский государственный университет

**АННОТАЦИЯ.** Речь человека является окном в его внутренний мир. Анализ речи (как устной, так и письменной) является ценным психодиагностическим инструментом. В последние годы такой анализ проводится совместно лингвистами, в том числе компьютерными, и психологами. Очень важным, но мало разработанным направлением современных междисциплинарных исследований является анализ речи лиц, совершивших суицид. Такой анализ позволяет глубже понять особенности мышления суицидента, а в дальнейшем разработать методики диагностирования риска суицидального поведения на основе лингвистического анализа. В работе приводится анализ динамики ряда высокочастотных параметров идиостиля автора блога, совершившего суицид. Впервые подобный анализ проводится на материале русских интернет-текстов. Для обработки текстов используется программа LIWC. Выполнен корреляционный анализ для выявления связи между численными значениями параметров текста и датой создания текста (числом дней до суицида). Проведена визуализация данных, а также сравнение с результатами работ, выполненных на материале других языков. 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.

## THE STUDY OF THE SUICIDENT'S IDIOSTYLE DYNAMICS

**Litvinova T. A.,**

Cand. Philol. Sci., Researcher in Regional Centre for Russian Language, Voronezh State Pedagogical University;

**Seredin P.V.,**

Dr. Phys.-Math. Sci., Docent of the Department of Solid State Physics and Nanostructures, Voronezh State Pedagogical University

**ABSTRACT.** A person's speech is a window into his inner world. The analysis of speech (both oral and written) is a valuable psycho diagnostic tool. In recent years, such an analysis is conducted jointly by linguists, including computer linguists, and psychologists. A very important, but understudied area of modern interdisciplinary research is the analysis of suicide speech. Such an analysis can provide useful insight to the peculiarities of suicidal thinking which can be used to further development of methods for diagnosing the risk of suicidal behavior based on linguistic analysis. The paper analyzes the dynamics of a number of high-frequency parameters in the idiostyle a blogger who committed a suicide. For the first time such an analysis is conducted on the material of Russian Internet texts. For text processing, the LIWC program is used. A correlation analysis was performed to identify the relationship between the numerical values of the linguistic parameters and the date the texts were created. Data visualization, as well as comparison with the results of works performed on the material of other languages, was carried out.

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

**В**ведение. Анализ речи (как письменной, так и устной) представляет собой ценный диагностический инструмент, используемый в числе прочего в психологии, в том числе судебной, криминалистике, рекрутинге и т.д. Развитие интернет-коммуникации предоставило ученым богатый материал для исследования проблемы взаимоотношения психологических характеристик

и состояний и языка. Одной из характерных особенностей работ последних лет, направленных на изучение указанной проблемы, является использование инструментария компьютерной и корпусной лингвистики и консолидация усилий лингвистов и психологов (см., например, материалы воркшопа Computational Linguistics and Clinical Psychology, <http://clpsych.org>). Отдельной научной проблемой, над которой активно работают ученые, является проблема диагностики склонности к суициду на основе комплексного анализа речи, в том числе интернет-текстов. Большая часть подобных исследований была выполнена на материале английского языка (см. подробнее: [1]). Впервые русские он-

© Литвинова Т.А., Середин П.В., 2017

<sup>1</sup> 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».

лайн-тексты лиц, совершивших суицид, были рассмотрены в работах [2; 3]. В них было показано, что такие тексты обладают своей спецификой по сравнению с текстами лиц контрольной группы. Однако в указанных работах не исследовалась динамика языковых параметров по мере приближения даты смерти автора. Надо отметить, что данная проблема относится к числу малоисследованных в мировой науке, между тем как анализ динамики идиостиля позволит глубже понять психологическое состояние суицидента, а результаты такого анализа могут быть использованы при разработке диагностического инструментария. Цель нашей работы – исследование динамики языковых параметров интернет-дневника человека, совершившего законченный суицид.

**Методы исследования.** Материалом исследования послужили тексты блога на платформе LiveJournal («Живой журнал», ЖЖ) пользователя light\_medelis (Лайт) <http://light-medelis.livejournal.com/> Пользователь также вел блог под ником lm\_diary <http://lm-diary.livejournal.com/> Указанные аккаунты принадлежат Сергею Макарову, создателю сайта mysuicide.ru, одного из крупнейших сайтов рунета, посвященных вопросам суицида. Сергей покончил с собой в 12.12.2005, о чем доподлинно известно из комментариев его друзей в ЖЖ и публикаций в прессе.

Мы отдельно рассматривали два блога С. Макарова, поскольку они имеют разную направленность: блог light\_medelis посвящен обсуждению вопросов суицида, депрессии и т.д. (далее для краткости будет именоваться его CS, от англ. case study); блог lm\_diary более личный, по жанру близкий к личному дневнику (далее для краткости будет именоваться его ЛД, «личный дневник»), т.к. в нем автор описывает свои чувства и переживания. Оба блога велись практически до дня смерти автора, однако CS велся с 13.06.2003 по 11.12.2005, а ЛД – с 28.07.2004 до 11.12.2005. Для корректного сравнения полученных данных мы брали тексты, написанные за один и тот же период, т.е. анализировали записи из CS начиная с 28.07.2004. Все тексты (посты) автора, написанные за один день, заносились в отдельный файл, именуемый по дате записи. Также в этот файл заносились комментарии автора к своему посту. Такие операции проводились отдельно

для каждого блога. Весь неавторский текст (цитаты, ссылки и т.д.) удалялся.

Тексты обрабатывались при помощи программы Linguistic Inquiry and Word Count [4], которая позволяет вычислять в тексте число слов тех или иных грамматических (число местоимений, союзов, предлогов), лексических (число слов, обозначающих когнитивные процессы, чувства, эмоции и т.д.) категорий, слов длиннее 6 букв, знаков препинания и т.д., в процентах от общего числа слов в тексте. Помимо стандартного словаря, мы разработали набор собственных словарей программы (подробнее см. [3]). Всего были извлечены значения 142 параметров текста. Далее мы проводили отбор частотных параметров, т.е. тех, значения которых отличаются от нуля более чем в 50 % текстов (в обоих блогах). На данной стадии анализа число параметров текста сократилось до 65. Далее мы провели корреляционный анализ для выявления взаимосвязи между численными значениями параметров текста и числом днем до смерти, причем мы проводили расчеты как с использованием коэффициента Пирсона, так и с использованием коэффициента ранговой корреляции Спирмена (поскольку значения параметров текста обладали разным типом распределения). Мы составили общий список параметров текста для обоих блогов, которые коррелировали с числом днем до смерти хотя по одному из критериев.

**Результаты и обсуждение.** Как показало наше исследование, из отобранных параметров текста 17 коррелировало с числом дней до смерти как в одном, так и в другом блоге: процент слов, указывающих на пишущего («я», «мой» и их формы; наречие «по-моему» и т.п.); глаголов; слов, обозначающих действие; союза «и»; личных местоимений (словарь программы); личных местоимений (пользовательский словарь); местоимений всего; слов, обозначающих социальные взаимодействия; слов, обозначающих положительные эмоции; предлогов; предлогов «с»; союзов; 100 самых частотных слов русского языка; числительных; процент местоимений «я».

Для детального анализа поведения отобранных параметров текста мы визуализировали наши данные. Мы построили на одном графике зависимости интенсивности постинга для двух блогов от числа дней до смерти (рис. 1).

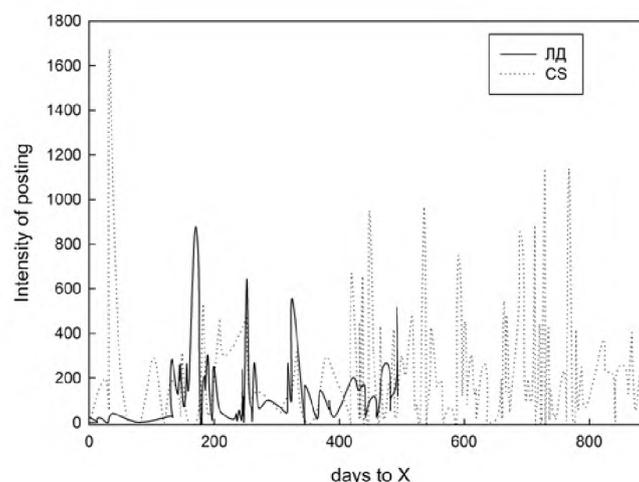
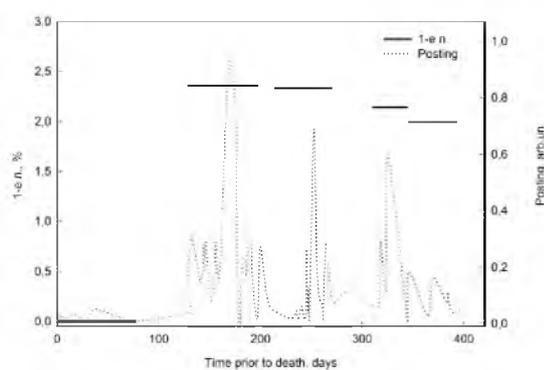


Рисунок 1 – График зависимости интенсивности постинга (в словах) для двух блогов от числа дней до смерти

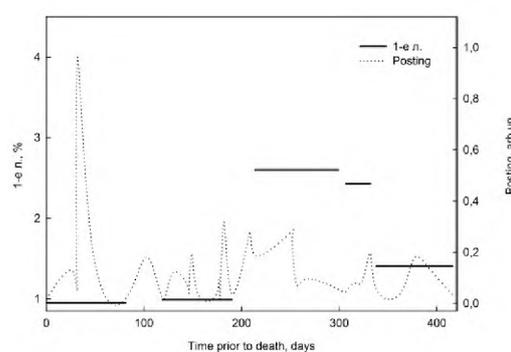
Как хорошо видно из экспериментальных данных, представленных на рис. 1, для обоих блогов характерно наличие нескольких периодов всплеска и спада интенсивности постинга. В определенные периоды характер интенсивности постинга совпадает в обоих блогах. Для дальнейшего анализа мы выбрали пять периодов, в которые одновременно наблюдалось усиление активности постинга в обоих блогах (см. графики ниже). Далее мы рассчитали для всех приведенных выше 17 параметров текста их средние значения в указанные периоды активности постинга. Полученные результаты мы представили графически, приведя усредненные значения выбранного параметра текста в анализируемые периоды одновременно с нормированной зависимостью интенсивности постинга в конкретном блоге.

Остановимся на некоторых параметрах, которые активно исследовались в работах на материале других языков, посвященных исследованию динамики параметров идиостиля суицидента с использованием программы LIWC. В указанных работах (см. обзор

полученных результатов в работе [5]) исследователи опираются на существующие концепции суицидогенеза, согласно которым суицид связан с усиливающейся социальной изоляцией (социологическая концепция), чувством безнадежности, грусти, тоски (психологические концепции суицида). В связи с этим особое внимание уделяется анализу частности местоимений «я» и «мы», слов, обозначающих социальные взаимодействия; числу слов, обозначающих положительные и отрицательные эмоции. В ряде работ было показано, что по мере приближения даты суицида частотность местоимений «я» увеличивается, число местоимений «мы», напротив, уменьшается (см., например, [6]); падает число слов, обозначающих социальные взаимодействия, а также положительные эмоции, увеличивается число слов, обозначающих негативные эмоции (см. обзор: [5]). Однако в некоторых работах были получены противоположные результаты (см. подробнее там же: [5]).

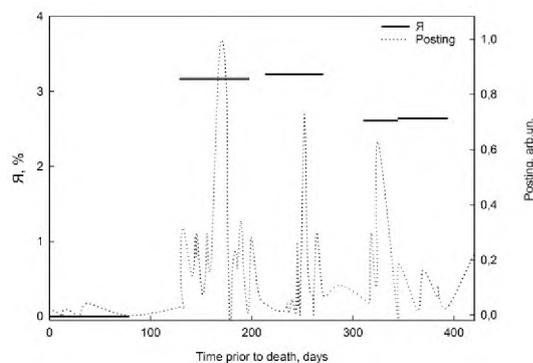


а)

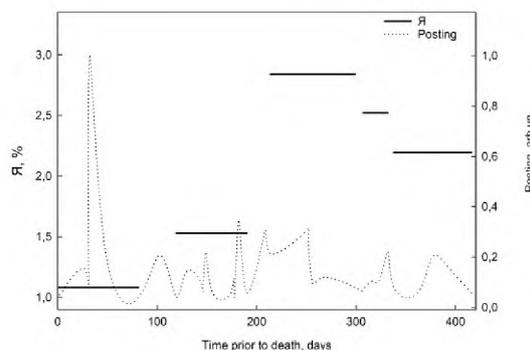


б)

Рисунок 2 – Графики изменения параметра «Процент слов, указывающих на пишущего»: а – ЛД, б – CS



а)



б)

Рисунок 3 – Графики изменения параметра «Процент местоимений «я»»: а – ЛД, б – CS

Поскольку эти параметры тесно связаны друг с другом, мы рассматриваем их вместе. В предыдущих работах рассматривался только параметр «Процент местоимений «я»». Поведение этих двух параметров практически идентично. В личном дневнике процент слов указанной категории является стабильно высоким в периоды интенсивного постинга, но в последний период число таких слов резко снижается, как и интенсивность постинга. Однако в текстах CS, несмотря на всплеск активности в последний период, также наблюдается снижение частотности языковых единиц, указывающих

на автора текста, что находится в противоречии с результатами, показанными в ряде работ на материале художественных текстов, однако согласуется с результатами, полученными в работе [5], наиболее близкой к нашей работе по методологии и материалу (исследовались тексты блога в течение года до смерти автора). Заметим, однако, что в русском языке, как известно, местоимение «я» может опускаться. Следовательно, в дальнейшем необходимо также исследовать в динамике частотность форм 1-го лица ед. ч. глаголов.

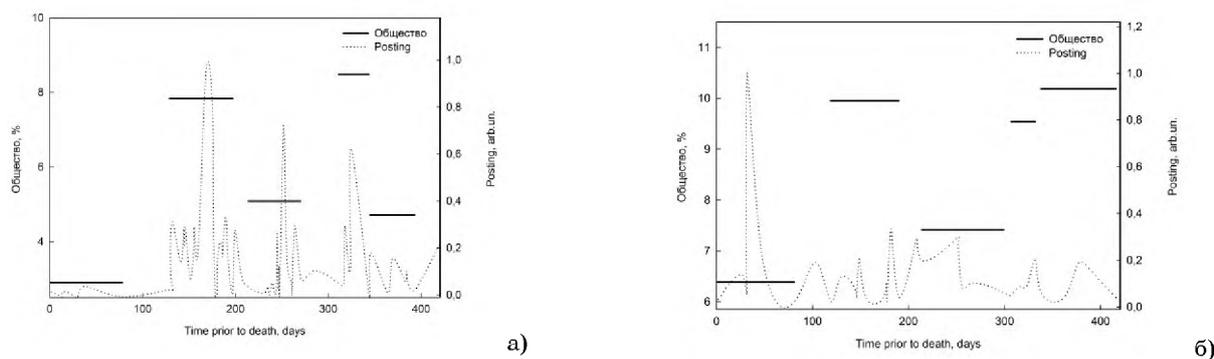


Рисунок 4 – Графики изменения параметра «Процент слов, обозначающих социальные взаимодействия»: а – ЛД, б – СС

Результаты анализа поведения параметра «Процент слов, обозначающих социальные взаимодействия» в проанализированном нами материале согласуются с полученными в большинстве работ: неза-

долго до смерти доля таких слов в текстах снижается, что находится в соответствии с социологической концепцией суицидогенеза.

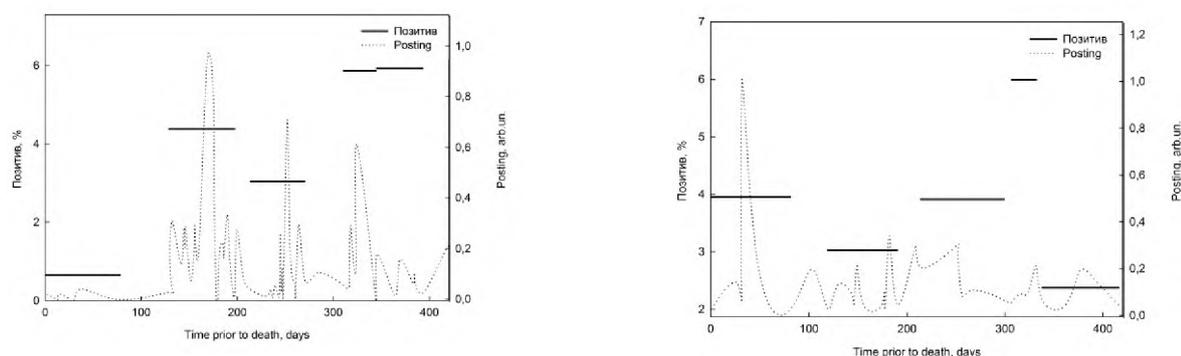


Рисунок 5 – Графики изменения параметра «Процент слов, обозначающих положительные эмоции»: а – ЛД, б – СС

Анализ слов, обозначающих эмоции, является важной частью исследований текстов суицидентов. Поскольку наш анализ не показал корреляций между процентом слов, обозначающих отрицательные эмоции в тексте, и числом дней до смерти, мы визуализировали поведение параметра «Процент слов, обозначающих положительные эмоции». В личном дневнике процент слов, обозначающих положительные эмоции, снижается вместе с интенсивностью постинга. В текстах СС, однако, в последний период процент слов указанной группы повышается вместе с интенсивностью постинга. Повышение доли слов, обозначающих положительные эмоции, в период накануне суицида было отмечено в 4 из 9 работ, посвященных анализу речи суицидентов с использованием LIWC (см. [5]), что связывают с улучшением психологического состояния автора после принятия решения о суициде.

**Выводы.** Из полученных результатов визуализации данных следует, что изменение отобранных параметров текста зачастую происходит нелинейно. Поэтому абсолютно недостаточно при анализе динамики идиостиля суицидента производить отбор параметров текста с использованием только корреляционного анализа данных, предполагающего поиск линейных взаимосвязей, без визуализации поведения параметра текста в различные периоды времени. Полученные в существующих работах противоречивые результаты, касающиеся характера динамики языковых параметров текстов суицидентов, могут объясняться, среди других причин, также недостаточным вниманием ученых к поведению

каждого параметра в разные периоды времени, что возможно только с использованием методов визуализации. Кроме того, указанные противоречия могут объясняться тем, что в существующих работах анализируются тексты разных жанров, причем в основном художественные тексты. Как показало наше исследование, различия в поведении параметров текста могут наблюдаться даже в рамках жанра интернет-блог, который может быть представлен, видимо, разными субжанрами. Помимо прочего, указанные различия в результатах исследования могут объясняться тем, что в большинстве работ используются художественные тексты, написанные в течение длительного периода времени, и на характер изменений параметров текста могли влиять возрастные изменения. Следовательно, в дальнейшем необходимо исследовать поведение параметров текстов разных авторов, созданных ими в течение одного и того же времени, например, за год до смерти. В дальнейшем нам представляется также перспективным провести исследование, направленное на поиск корреляций между параметрами текста и порядковым номером текста (поста), а не только числом дней до смерти, как в настоящем исследовании, поскольку изменения в поведении языковых параметров могут быть связаны не только с изменениями в состоянии автора, происходящими со временем, но и с какими-либо событиями в жизни автора текста, которые влияют на интенсивность постинга.

Несмотря на указанные сложности, исследование доказывает, что именно поиск трендов и анализ

поведения параметров текстов в динамике позволит глубже понять когнитивные особенности суицидентов и в дальнейшем разработать прогностические

модели для оценки риска суицидального поведения на основе лингвистического анализа, применимые в том числе и для интернет-текстов.

#### **СПИСОК ЛИТЕРАТУРЫ:**

1. Литвинова, Т.А., Выявление склонности личности к суицидальному поведению на основе количественного анализа ее речевой продукции [Электронный ресурс] / Т.А. Литвинова, О.В. Загоровская, О.А. Литвинова // Studia Humanitatis. 2016. N 1. ([http://st-hum.ru/sites/st-hum.ru/files/pdf/zagorovskaya\\_litvinova\\_litvinova.pdf](http://st-hum.ru/sites/st-hum.ru/files/pdf/zagorovskaya_litvinova_litvinova.pdf))
2. Литвинова, Т.А. Лингвистическая модель диагностики суицидального поведения [Текст] / Т.А. Литвинова, П.В. Середин, О.А. Литвинова, О.В. Загоровская // Филологические науки. Научные доклады высшей школы. – 2017. – № 5. – С. 49-54.
3. Litvinova, T.A. Identification of Suicidal Tendencies of Individuals Based on the Quantitative Analysis of Their Internet Texts / T.A. Litvinova, P.V. Seredin, O.A. Litvinova, O.V. Zagorovskaya // Computaciyn y Sistemas. – 2017. – № 21(2). – P. 243-252.
4. Pennebaker, J. W. The development and psychometric properties of LIWC2007 / J. W. Pennebaker. – Austin, TX: LIWC.net, 2007.
5. Li, T.M. Temporal and computerized psycholinguistic analysis of the blog / T.M. Li, M. Chau, P.S. Yip, P.W. Wong // Crisis. – 2014. – № 35(3). – P. 168-75.
6. Stirman, S. Word use in the poetry of suicidal and nonsuicidal poets / S. Stirman, J. Pennebaker // Psychosomatic Medicine. – 2001. – № 63. – P. 517-522.