

Ениколопов С.Н., Медведева Т.И., Воронцова О.Ю., Чудова Н.В., Кузнецова Ю.М., Пенкина М.Ю., Минин А.Н., Станкевич М.А., Смирнов И.В., Любавская А.А. Лингвистические характеристики текстов психически больных и здоровых людей



ЕНИКОЛОПОВ С.Н., МЕДВЕДЕВА Т.И., ВОРОНЦОВА О.Ю., ЧУДОВА Н.В., КУЗНЕЦОВА Ю.М., ПЕНКИНА М.Ю., МИНИН А.Н., СТАНКЕВИЧ М.А., СМИРНОВ И.В., ЛЮБАВСКАЯ А.А. ЛИНГВИСТИЧЕСКИЕ ХАРАКТЕРИСТИКИ ТЕКСТОВ ПСИХИЧЕСКИ БОЛЬНЫХ И ЗДОРОВЫХ ЛЮДЕЙ

English version: [Enikolopov S.N., Medvedeva T.I., Vorontsova O.Y., Chudova N.V., Kuznetsova J.M., Penkina M.Y., Minin A.N., Stankevich M.A., Smirnov I.V., Lubavskaya A.A. Linguistic characteristics of texts of mentally ill and healthy people](#)

Научный центр психического здоровья, Москва, Россия

Институт системного анализа Федерального государственного учреждения «Федеральный исследовательский центр "Информатика и управление" Российской академии наук», Москва, Россия

Московский государственный психолого-педагогический университет, Москва, Россия

Курганский государственный университет, Курган, Россия

[Сведения об авторах](#)

[Литература](#)

[Ссылка для цитирования](#)

Приведен обзор существующих методов интеллектуального анализа текстов, используемых в качестве исследовательских инструментов в клинической психологии и психиатрии. В качестве первого этапа разработки комплексного метода лингвистического анализа текстов, который помимо лексического учитывает также другие аспекты письменной речи, проведен лингвистический анализ текстов, написанных здоровыми людьми (142 человека) и больными с эндогенными психическими заболеваниями (22 человека). Для анализа семантико-синтаксической структуры высказываний применялся разработанный для систем искусственного интеллекта метод реляционно-ситуационного анализа, который опирается на синтаксемный анализ Г.А.Золотовой и на концепцию неоднородных семантических сетей Г.С.Осипова. Также был применен метод частотного лексического анализа с использованием анализатора PLATIn, разработанного на основе процессора Exactus Expert. Результатами работы анализатора являются частотные характеристики, позволяющие оценивать выраженность в конкретном тексте и коллекции текстов лексики, принадлежащей к тематическим группам слов с семантикой различных психологических состояний и личностных установок. Примененные методы позволили выявить различия в текстах по ряду параметров: частоте встречаемости определенных синтаксем, психолингвистических маркеров, лексико-тематических маркеров. Показано, что в текстах больных значимо реже встречаются синтаксеммы, связанные с активностью и с отношениями каузации и классификации. Тексты больных также отличаются от текстов здоровых по ряду психолингвистических показателей,

отражающих наличие дефицитарной психопатологической симптоматики, сосредоточенность больных на себе. Лексико-частотный тематический анализ показал актуальность темы «болезни и здравоохранение» и сниженную значимость темы «дети и образование» в текстах больных испытуемых.

Ключевые слова: автоматический анализ текстов, реляционно-ситуационный анализ, синтаксемы, психолингвистические показатели, эндогенные психические заболевания

В последнее десятилетие в области компьютерной лингвистики и искусственного интеллекта возникла новая предметная область – интеллектуальный анализ текстов, результаты исследований в которой уже доведены до уровня программных систем. Это позволяет перейти к использованию этих методов анализа текстов в качестве исследовательских инструментов в других предметных областях, в частности, в области клинической психологии и психиатрии. Задача выявления системы параметров письменной речи, отвечающей за создание «здорового» текста, не создающего впечатление слабости или искаженности мыслительного процесса и эмоционального неблагополучия автора, может быть поставлена в настоящее время как междисциплинарная.

Психологический подход к исследованию речи был заложен в работах классиков: Э.Блейлер, К.Клейст, Э.Крепелин, П.И.Карпова, В.Н.Образцова, В.П.Осипова. К характерным для патологических состояний были отнесены такие явления, как «разорванная речь», соскальзывание и бессвязность речи, атаксия (отсутствие смысловых связей между словами и снижение направленности высказываний) [Микиртумов, Завитаев, 2012]. В качестве средств формальной оценки тех особенностей, которые характеризуют речь психически больных, были предложены различные психолингвистические показатели: коэффициент Трейгера, коэффициент опредмеченности действия, показатель Готтшалка–Глезер, индекс прегнантности Эртеля, индекс Шлисмана, фог-индекс и др. Наряду с разработкой проблемы общей специфики патологической речи [Белянин, 2003] выделяются и развиваются отдельные направления, связанные с исследованиями и систематизацией нарушений при разных вариантах психической патологии, в частности мании, паранойи, нарциссизме [Додонова, 1988; Микиртумов, Ильичев, 2007; Ремесло, 2007]. Анализ литературы показывает, что исследования в основном посвящены устной речи испытуемых. Психопатологических исследований письменной речи заметно меньше [Алмаев, Малкова, 2006].

Один из наиболее исследуемых в мировой психиатрии и психиатрической лингвистике предметов – различные речевые (прежде всего, синтаксические и семантические) аномалии при шизофрении [Карякина и др., 2017; Condray et al., 2002; Ketteler et al., 2012; Moro et al., 2015]. В частности, П.Лидделом с соавторами [Liddle et al., 2002] была предложена диагностическая схема анализа речевой продукции Thought and Language Index (TLI), позволяющая оценивать выраженность специфических для шизофрении нарушений и включающая в себя такие показатели, как обеднение речи, ослабление целенаправленности, неустойчивость, нечеткий синтаксис, алогичность, персеверация, отвлекаемость, наличие неологизмов.

Наряду с изучением особенностей психотической речи проводятся психолингвистические исследования невротических расстройств. Среди характерных для речи невротиков по сравнению со здоровыми людьми выделены такие показатели, как уровень выраженности в речи состояний тревоги, депрессии, агрессии; частота употребления местоимений первого лица; показатели предметности и связности речи; средние коэффициенты пассивного залога, эмболии и словарного разнообразия; количество элементарных дискурсивных единиц, частотность некоторых дискурсивных отношений; многословие и снижение коэффициента словарного разнообразия; уменьшение по сравнению с группой здоровых в речи соотношения категорий «совершенные действия» и «несовершенные действия» [Додонова, 1988; Кибрик, 2009; Корабельникова и др., 1999]. Были выделены маркеры, свидетельствующие о том, что аффективность, выраженная в

тексте, превышает порог невротичности; показано, что невротическая форма аффекта коррелирует с семантикой агрессии, замкнутости, тревоги и с тенденцией избавления от эмоционального переживания; агрессивная форма аффекта – с конфигурацией пассивности и паратаксиста, тематикой силы, отрицательной оценки, ригидности; депрессивная форма – с доминированием единиц со значением пассивности, статальности, абстрактности, с выраженностью сочинительности связи в тексте и темами напряженности, пассивности, негативизма, вины, страдания и одиночества [Алмаев и др., 2009; Ремесло, 2007; Хандамова, 2002]. Установлено, что к речевым предикторам суицидального и аутоагрессивного поведения относятся такие параметры, как меньшее лексическое разнообразие, меньшее число предлогов, большее число местоимений, более высокий индекс логической связанности, более высокий индекс длины предложения [Загоровская и др., 2016].

Современный этап развития патопсихологического изучения речи в значительной степени определяется проблематикой разработки средств автоматического анализа и математического аппарата для него.

Одним из часто используемых средств автоматического анализа текстов для исследований в области психологии и психолингвистики является программа Linguistic Inquiry and Word Count (LIWC) [Pennebaker et al., 2015]. Основная идея, заложенная в ней, заключается в том, что психологические характеристики автора связаны с количественными параметрами текста: частота знаков препинания, слов определенной части речи (предлогов, союзов, местоимений, наречий), слов определенной лексико-семантической группы (передающих негативные или позитивные эмоции, описывающих когнитивные процессы). Например, с помощью LIWC исследовались различные лексические признаки психологического неблагополучия [Collison, 2016; Pennebaker, 1997] и многое другое. Математический подход к психотической «энтропийной» речи предложен в [Ribalov, Zislin, 2005]. При проведении текстовой диагностики шизофрении выявлена целесообразность применения таких статистических показателей, как коэффициент Ципфа, коэффициент лексической упорядоченности текста Хердана, спектральное распределение Хайтун–Тулдавы и др. [Пашковский и др., 2009], а также использования кластеризации [Elvevåg et al., 2002] и тематического моделирования (например, с помощью латентно-семантического анализа (LSA)) [Elvevag et al., 2007]. В работе [Resnik et al., 2015] лексический анализ текстов с помощью LIWC был использован совместно с методами тематического моделирования (LDA) для выявления выраженности нейротизма и депрессии у учащихся вузов. Группа исследователей из Института психиатрии Колумбийского университета показала применимость LSA для ранней диагностики риска развития шизофрении [Bedi et al., 2015]; в настоящее время осуществляется проект Automated analysis of semantics/syntax in psychosis, развивающий данное исследование [Corcoran, 2018], предложен метод для определения текстов, написанных больными шизофренией, на основе метода опорных векторов.

Большое значение для развития средств компьютерного лингвистического анализа имеют тематические shared tasks (или «открытые соревнования»). Одно из них, проведенное в 2011 г., было посвящено теме выявления лингвистических особенностей предсмертных записок суицидентов [Pestian et al., 2012]. Наилучшие результаты в этом соревновании показала система, представленная в работе [Yang et al., 2012]. В ней применен гибридный подход, в котором реализован поиск ключевых слов по словарям и тезаурусам, методы разметки последовательностей на основе случайных полей [Lafferty et al., 2001], а также другие методы машинного обучения. Результаты, полученные от разных анализаторов, комбинировались с помощью нескольких стратегий голосования.

В рамках соревнований CLPsych 2015 [Coppersmith et al., 2015] были проведены исследования, направленные на разработку методов идентификации признаков различных видов психического неблагополучия (депрессия, посттравматическое расстройство, сезонное аффективное расстройство и др.). В соревновании CLPsych 2016 [Milne et al., 2016] была поставлена задача оценки степени выраженности в письменной речи признаков психологического неблагополучия автора. Наилучшие результаты были показаны методами, представленными в работах [Mac Kim et

al., 2016; Malmasi et al., 2016]. Стоит также отметить, что в настоящий момент осуществляются соревнования CLEF eRisk 2017 и CLPsych 2017, задачей которых является развитие средств выявления признаков психологического неблагополучия по текстам из социальных сетей.

Как видно из представленного обзора, большинство методов автоматического анализа текстов опираются лишь на лексические признаки (категории LIWC, лексические кластеры, n-граммы и др.) и не учитывают структуру текста на более высоком уровне – синтаксическом и семантическом и, кроме того, ориентированы на работу с английским языком. В настоящей работе представлены данные исследования, ориентированного на создание комплексного метода лингвистического анализа текстов, который помимо лексического учитывает также другие аспекты письменной речи и предназначен для анализа текстов на русском языке.

Методы

Выборка

В настоящем исследовании приняли участие две группы испытуемых: 22 пациента клиники ФГБНУ НЦПЗ (средний возраст 24,1 года), отвечавших критериям F 20., F 21., F 31.3 F 32.1 по МКБ-10, лечившихся в отделе по изучению эндогенных психических расстройств и аффективных состояний. Общей чертой исследованных больных шизофренического спектра было отсутствие острой продуктивной симптоматики в период исследования; и 142 человека – группа здоровых (студенты гуманитарных и технических вузов Москвы и Кургана, взрослые жители этих городов в возрасте от 32 до 46 лет). Выборки были уравнены по социодемографическим показателям.

Методики

Для анализа семантико-синтаксической структуры высказываний в нашем исследовании применялся метод реляционно-ситуационного анализа [Osipov et al., 2013], который опирается на синтаксемный анализ Г.А.Золотовой [Золотова и др., 2004] и на концепцию неоднородных семантических сетей Г.С.Осипова [Осипов, 1997]. Синтаксемой называется минимальная синтактико-семантическая единица языка, несущая обобщенный категориальный смысл и характеризующаяся взаимодействием морфологических, семантических и функциональных признаков. Для анализа предложения важно не только знание значений встречающихся в нем синтаксем, но и их сочетаемость друг с другом. Эта сочетаемость определяется семейством бинарных отношений на множестве синтаксем. Реляционно-ситуационный анализ позволяет выявлять семантику текста, ставя в соответствие синтаксемную структуру предложения логической структуре действий, описанных в этом предложении. Главную роль здесь играют глаголы, имеющие, как правило, центральное положение в семантической структуре предложения и оказывающие решающее влияние на именные словосочетания и предложения, составляющее синтаксемные группы. Сведения о синтаксической сочетаемости каждого глагола с синтаксемами занесены лингвистами в Словарь предикатных слов (в настоящее время словарь содержит 2,8 тыс. статей, в которых описаны сочетания 75 семантических ролей для более 5 тыс. предикатных слов), так что автоматический анализ текстов ведется с опорой на знания о русской языковой картине мира.

Для выявления реляционно-ситуационной структуры высказываний выполнялся семантико-синтаксический анализ, в результате которого строится дерево полного синтаксического разбора, определяются предикаты и их аргументы, для аргументов устанавливаются семантические роли (синтаксические значения). В разработанном в ИСА ФИЦ ИУ РАН семантико-синтаксическом парсере используется комбинация словарного подхода (использование словаря предикатных слов) и подходов машинного обучения, при этом синтаксический и семантический анализ выполняются в одной процедуре, что все вместе обеспечивает высокое качество семантического анализа текста.

Таким образом, наше исследование опирается на реляционно-ситуационное представление текста, учитывающее все уровни языка и доказавшее свою эффективность в решении многих задач интеллектуального анализа текстов и информационного поиска.

Также был применен метод частотного лексического анализа с использованием анализатора PLATIn, разработанного на основе процессора Exactus Expert [Девяткин и др., 2014]. Результатами работы анализатора являются частотные характеристики, позволяющие оценивать выраженность в конкретном тексте и коллекции текстов лексики, принадлежащей к тематическим группам слов с семантикой различных психологических состояний и личностных установок. В лингвистике под тематической группой слов (ТГС) понимается объединение слов на основе классификации предметов и явлений по принципу сопряженности с определенной темой, то есть на экстралингвистическом параметре; входящие в ТГС единицы принадлежат к различным частям речи [Алефиренко, 1981; Зеленецкий, Новожилова, 2003; Филин, 1982]. В ТГС наблюдаются различные (как парадигматические, так и синтагматические) виды связи; для них не обязательна идентифицирующая (ядерная) тема [Васильев, 1990]. Описание ТГС различного содержания является одним из предметов лингвистического анализа, результаты которого отражаются в многочисленных диссертационных работах, а также в специальных тематических и идеографических словарях (например, [Богусловский, 2004; Саяхова и др., 2000; Фридман, 2003]).

Способом сплошной выборки из Русского орфографического словаря Российской академии наук были составлены тематические группы лексических единиц, семантика которых соответствует описанным в социологии социально-экономическим причинам социального стресса. В соответствии с темами, используемыми социологическими службами (ВЦИОМ, Левада-центр) при оценке уровня социальной напряженности, выделены лексические группы следующего содержания: «Законность и правопорядок» (ок. 400 ед.), «Дети и образование» (ок. 350 ед.), «Экономика» (ок. 600 ед.), «Демография и экология» (ок. 400 ед.), «Преступность» (ок. 100 ед.), «Власти» (ок. 200 ед.), «Силовые структуры» (ок. 300 ед.), «ЖКХ» (ок. 550 ед.), «Социальное неравенство и несправедливость» (ок. 330 ед.), «Катастрофы» (ок. 100 ед.), «Здравоохранение и бесплатная медицина» (ок. 100 ед.). Эти ТГС также использовались в работе анализатора PLATIn.

В арсенал психолингвистических средств, позволяющих оценить наличие эмоционального напряжения автора на момент написания текста, в нашем исследовании вошли следующие показатели лексико-морфологического уровня: коэффициент Трейгера (КТ) – отношение количества глаголов к количеству прилагательных в единице текста; коэффициент опредмеченности действия (КОД) – соотношение количества глаголов к количеству существительных в единице текста; количество существительных и глаголов по сравнению с прилагательными и наречиями; отношение количества инфинитивов к общему количеству глаголов; количество глаголов в страдательном залоге; количество безличных глаголов; количество глаголов прошедшего времени; количество глаголов будущего времени; количество причастий и деепричастий; длина слов; количество местоимений, в т.ч. первого лица множественного числа, первого лица единственного числа, третьего лица множественного числа.

В ИСА ФИЦ ИУ РАН на базе метода сравнения ситуационно-реляционных моделей документов создана машина реляционно-ситуационного анализа (РСА), позволяющая проводить глубокий лингвистический анализ текста и выделять в нем около 150 параметров, представляющих собой характеристики синтаксической структуры предложений и семантических ролей предикатов, психолингвистические и лексические маркеры эмоциональности, характеристики обсуждаемых в тексте тем.

Всем испытуемым было предложено написать эссе на тему «Я, другие, мир» (объем – 1 стр.). Автоматический анализ текста, проведенный по 57 параметрам (синтаксемы, психолингвистические показатели, ТГС социальной напряженности), дал возможность сравнить частоту встречаемости этих параметров в группах больных и здоровых. Для статистического анализа использовалась

программа SPSS, для сравнения групп применялся критерий различий Манна–Уитни.

Результаты

Результаты, полученные при анализе различий, представлены в табл. 1. Приведены только те параметры, по которым были выявлены различия.

Таблица 1

Сравнение средних значений лингвистических параметров для больных и здоровых испытуемых

Лингвистические параметры	Здоровые	Больные	Значимость различий
<i>Синтаксемы</i>			
Агенты	0,0012±0,0023	0,0004±0,0012	*
Каузативы	0,0048±0,0041	0,0040±0,0055	*
Генеративы	0,0007±0,0015	0,0002±0,0009	*
Инструментативы	0,0079±0,0064	0,0058±0,0079	*
Предикаты	0,0081±0,0053	0,0054±0,0051	*
Субъекты содействия	0,0010±0,0019	0,0002±0,0012	*
Субъекты	0,0410±0,0124	0,0494±0,0191	*
<i>Психолингвистические показатели</i>			
Отношение количества инфинитивов к общему количеству глаголов	0,3118±0,0874	0,2363±0,1351	**
Местоимения первого лица множественного числа	0,0109±0,0112	0,0051±0,0102	**
Местоимения первого лица единственного числа	0,0277±0,0222	0,0598±0,0288	**
Местоимения третьего лица множественного числа	0,0053±0,0058	0,0043±0,0066	*
Коэффициент Трейгера (отношение количества глаголов к количеству прилагательных)	0,8056±0,2200	0,6909±0,2697	*
Отношение числа существительных и глаголов к количеству прилагательных и наречий	2,3669±0,7435	3,8636±4,7238	*
Длина слов	4,4506±0,3359	4,2794±0,3456	*
<i>ТГС социальной проблематики</i>			
Тема: Дети и образование	0,0026±0,0043	0,0005±0,0016	**
Тема: Здоровоохранение и бесплатная медицина	0,0002±0,0009	0,0018±0,0049	**

Примечания. уровень статистической значимости различий: * $p < 0,05$; ** $p < 0,001$.

Анализ показал, что тексты больных значимо отличаются от текстов здоровых испытуемых по частоте встречаемости определенных синтаксем.

Агент – производитель действия (пример: «кафе посещалось преимущественно *философами*»), значимо реже встречается у больных.

Субъект – компонент, которому приписывается предикативный признак (пример: «движения мои учтивы, решения неторопливы, и помыслы мои чисты»), чаще встречается в текстах больных, но

при этом у больных значимо реже встречается *Субъект содействия* – лицо, оказывающее помощь (пример: «Скворцова оказала первую помощь пассажиру на борту самолета»).

Предикат – второй из двух организующих модель предложения компонентов, признак, приписываемый субъекту (пример: «главная трагедия в жизни – прекращение борьбы» (Н. Островский)). Реже встречается в текстах больных.

Каузатив – причина действия или появления признака, свойства (пример: «Я не из гордости – из горести так прямо голову держу» (Б.Ахмадулина)) реже встречается в текстах больных, различие на уровне статистической тенденции.

На уровне статистической тенденции тексты различаются по параметру *Генеративы*, который также снижен у больных. Роль генератива появляется в предложении при установлении родовидовых отношений, при классификации: генератив – это класс, более крупная общность, в которую включается объект (пример: Изберу я себе из бедных, говорит, повиднее. Ей моего благодеяния всю жизнь не забыть (Островский)).

Инструментатив – орудие действия (пример: «что написано пером, того не вырубишь топором»).

Также в текстах больных слабее выражены такие психолингвистические показатели, как коэффициент *Трейгера*, количество местоимений 3-го лица множественного числа, длина слов, отношение количества инфинитивов к общему количеству глаголов. Также снижено количество местоимений первого лица множественного числа при увеличении местоимений первого лица единственного числа и общего количества местоимений. Кроме того, для текстов группы больных оказался выше показатель отношение числа существительных и глаголов к количеству прилагательных и наречий.

Лексика, связанная с социальной проблематикой, когда речь идет о детях, учебе, получении образовании, о воспитании (ТГС *Дети и образование*), в текстах больных встречается реже, а лексика, связанная с проблемами здоровья (ТГС *Здравоохранение и бесплатная медицина*), – чаще.

Обсуждение результатов

Некоторые данные хорошо интерпретируемы. Например, в [Самохвалов, 2002] указано, что снижение коэффициента *Трейгера* наблюдается при наличии дефицитарной психопатологической симптоматики, диссоциативных расстройствах, соматизированной патологии, астении; указывает на такие личностные характеристики, как нерешительность, зависимость, тревожность. Поскольку в группе больных преобладают люди с диагностированной депрессией, сниженность этого коэффициента ожидаема и согласуется с клинической симптоматикой депрессивных расстройств. С этим же, вероятно, связано и снижение количества такой синтаксемы, как *агенс* – производитель действия, однако только в том случае, если в этой роли стоит местоимение «я» (отметим, что местоимение «я» у больных встречается значимо чаще). Последнее требует отдельной проверки, которую можно осуществить с помощью, разработанной нами машины РСА; в задачу настоящего исследования выполнение подобных работ не входило.

Снижение частоты встречаемости синтаксемы *каузатив* может говорить об ограничениях во внимании к причинности, каузальной связности событий. Снижение количества *генеративов* как показателя категоризации и обобщения опыта требует осмысления. Можно выдвинуть гипотезу о том, что снижение количества таких синтаксем, как *агенс* (активность), *каузатив* (причинность) и *генератив* (обобщенность), может быть обусловлено снижением способности к регуляции как интегративного аспекта психического опыта, из которого происходит способность чувствовать себя хозяином собственных компетентных действий, обретение уверенности из опыта эффективности. Это снижение приводит к тому, что в представлении больных действия происходят

стихийно и произвольно. Выдвинутая гипотеза требует проверки, что не входило в цели данного исследования.

Снижение частоты встречаемости синтаксемы *субъект содействия*, а это лицо, оказывающее помощь, может отражать как то, что больные ощущают дефицит поддержки, социальную изоляцию, так и то, что больные не видят себя в роли помогающих. О первом косвенно может свидетельствовать снижение количества *местоимений первого лица множественного числа*, которое может говорить о снижении чувства единения с другими, ощущении отделенности от других. Проверка этих предположений потребует применения машины РСА, позволяющей установить категорию слов, занимающую интересующую исследователя семантическую роль.

Увеличение *местоимений первого лица единственного числа* показывает, что больные сосредоточены на себе, образы других отсутствуют, или присутствуют только для того, чтобы выразить их отношение к автору. Увеличение этого показателя показано и в других исследованиях [Al-Mosaiwi, Johnstone, 2018]. Собственные потребности и аффекты затрудняют восприятие других как целостных отдельных личностей, со своей историей, сильными и слабыми сторонами, о чем свидетельствует снижение *количества местоимений третьего лица множественного числа*.

Преобладание в текстах больных такой темы, как «Здравоохранение», также легко интерпретируется – если для здоровых эта тема больше связана с социальными вопросами и может характеризовать уровень напряжения при обсуждении общественно значимых тем, то для людей, находящихся на момент обследования в стационаре, вопросы здравоохранения, доступности медицины – это вопросы уровня их актуального благополучия. Для больных одной из самых актуальных тем является тема болезни. Что может отражать как актуальность текущей ситуации нахождения в лечебном учреждении, так и то, что тема болезни может занимать ведущее положение в личности больного, через болезнь он может определять свою идентичность [Николаева, 1970]. В то же время другие социальные вопросы, такие как «*дети, образование*», для больных не являются актуальными, они не чувствуют вовлеченность в данную социальную проблематику.

Заключение

Проведен лингвистический анализ текстов, написанных здоровыми людьми и больными с эндогенными психическими заболеваниями. Анализ выделенных параметров, представляющих собой характеристики синтаксемной структуры предложений, психолингвистические маркеры эмоциональности и характеристики обсуждаемых в тексте тем, позволил выявить различия в текстах здоровых и больных людей. Так, у больных значимо снижена частота встречаемости таких синтаксем, связанных с активностью, как агенсы, субъекты содействия, инструментативы, предикаты, каузативы, генеративы. Отличия по психолингвистическим показателям, таким как местоимения первого лица множественного числа, местоимения первого лица единственного числа, коэффициент Трейгера, отношение количества инфинитивов к общему количеству глаголов, отражают наличие дефицитарной психопатологической симптоматики, сосредоточенность больных на себе. Характеристики обсуждаемых тем показали актуальность темы болезни и здравоохранения для больных людей при снижении интереса к другой социальной проблематике.

В целом проведенное исследование мы рассматриваем как пилотажное – группа больных по объему меньше группы здоровых, что может приводить к существенным сдвигам в статистике. К тому же сама группа больных неоднородна по диагнозу, что может сказываться на «средних» при проведении анализа различий. Полученные различия позволяют нам зафиксировать важность исследования определенных параметров текста и планировать как увеличение выборки больных, так и применение более сложной схемы работ, когда полученные в статистическом исследовании данные используются для формирования новых гипотез и проверки их уже с помощью нового инструментария, опирающегося на методы искусственного интеллекта – машины РСА.

Финансирование

Исследование выполнено при поддержке Российского фонда фундаментальных исследований, проект 17-29-02305 «Разработка методов анализа текстов, порождаемых в ситуации психологического неблагополучия и психического нездоровья».

Литература

Алефиренко Н. Некоторые аспекты сопоставительно-семантического анализа фразеологизмов русского и украинского языков. В кн.: З.Д. Попова (Ред.), Семантические категории сопоставительного изучения русского языка. Воронеж: Воронеж. гос. университет, 1981.

Алмаев Н.А., Дороднев А.Б., Малкова Г.Ю. Проявление психологической травмы в автобиографических рассказах. Экспериментальная психология, 2009, 2(2), 104–115.

Алмаев Н.А., Малкова Г.Ю. Контент-аналитическое исследование личности. Психология. Журнал высшей школы экономики, 2006, 3(1), 19–42.

Белянин В.П. Психолингвистика. М.: Моск. психол.-соц. ин-т, 2003.

Богусловский В.М. Оценка внешности человека. М.: Торсинг, 2004.

Васильев Л.М. Современная лингвистическая семантика. М.: Высшая школа, 1990.

Девяткин Д.А., Кузнецова Ю.М., Чудова Н.В., Швец А.В. Интеллектуальный анализ проявлений вербальной агрессивности в текстах сетевых сообществ. Искусственный интеллект и принятие решений, 2014, No. 2, 27.

Додонова Н. Диагностическое значение лексико-семантических и морфологических особенностей речи больных неврозами: дис. ... канд. психол. наук. Ленинград, 1988.

Загоровская О.В., Литвинова О.А., Литвинова Т.А. Выявление склонности личности к суицидальному поведению на основе количественного анализа ее речевой продукции. Studia Humanitatis, 2016, No. 1. http://st-hum.ru/sites/st-hum.ru/files/pdf/zagorovskaya_litvinova_litvinova.pdf

Зеленецкий А.Л., Новожилова О.В. Теория немецкого языкознания. М.: Академия, 2003.

Золотова Г.А., Онипенко Н.К., Сидорова М.Ю. Коммуникативная грамматика русского языка. М.: Институт русского языка РАН, 2004.

Карякина М.В., Сидорова М.Ю., Шмуклер А.Б. Нарушения речи у больных шизофренией. Социальная и клиническая психиатрия, 2017, 27(4), 93–100.

Кибрик А.А. Рассказы о сновидениях: корпусное исследование устного русского дискурса. М.: Языки славянских культур, 2009.

Корабельникова Е., Вейн А., Голубев В., Крейнс М. Психолингвистическое исследование сновидений детей и подростков с невротическими расстройствами. Журнал неврологии и психиатрии, 1999, No. 1, 18–21.

Микиртумов Б.Е., Завитаев П.Ю. Аутизм: история вопроса и современный взгляд. СПб.: НЛ, 2012.

- Микиртумов Б.Е., Ильичев А.Б. Клиническая семантика психопатологии. СПб.: СПбГПМА, 2007.
- Николаева В.В. Внутренняя картина болезни при некоторых психических заболеваниях: автореф. дис. ... канд. психол. наук. Москва, 1970.
- Осипов Г.С. Приобретение знаний интеллектуальными системами. Основы теории и технологии. М.: Наука, 1997.
- Пашковский В.Э., Пиотровская В.Р., Пиотровский Р.Г. Психиатрическая лингвистика. М.: URSS, 2009.
- Ремесло М.Б. Клинико-лингвистические характеристики больных невротами и их динамика в процессе психотерапии: автореферат дис. ... канд. мед. наук. Санкт-Петербург, 2007.
- Самохвалов В.П. Психиатрия. Ростов-на-Дону: Феникс, 2002.
- Саяхова Л.Г., Хасанова Д.М., Морковкин В.В. Тематический словарь русского языка. М.: Русский язык, 2000.
- Филин Ф.П. Очерки по теории языкознания. М.: Наука, 1982.
- Фридман С.А. Человек. Характер и поведение. Толковый словарь. М.: Мнемозина, 2003.
- Хандамова Э.Ф. Вербализация психоэмоциональных состояний в речевой деятельности: автореф. дис. ... канд. филолог. наук. Краснодар, 2002.
- Al-Mosaiwi M., Johnstone T. In an Absolute State: Elevated Use of Absolutist Words Is a Marker Specific to Anxiety, Depression, and Suicidal Ideation. *Clinical Psychological Science*, 2018, 6(4), 529–542.
- Bedi G., Carrillo F., Cecchi G.A., Slezak D.F., Sigman M., Mota N.B., Ribeiro S., Javitt D.C., Copelli M., Corcoran C.M. Automated analysis of free speech predicts psychosis onset in high-risk youths. *Nature Partner Journals Schizophrenia*, 2015, 1(15030). <https://www.ncbi.nlm.nih.gov/pubmed/27336038>
- Collison E.A. Evaluating the Pennebaker Paradigm with Bereaved Emerging Adults: Applications of Text Analysis. Virginia: Commonwealth University, 2016.
- Condray R., Steinhauer S.R., van Kammen D.P., Kasperek A. The language system in schizophrenia: effects of capacity and linguistic structure. *Schizophrenia Bulletin*, 2002, 28(3), 475–490.
- Coppersmith G., Dredze M., Harman C., Hollingshead K., Mitchell M. CLPsych 2015 shared task: Depression and PTSD on Twitter. *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 2015, 31–39.
- Corcoran C. 26. Novel approaches to psychosis risk: movement, stress modulation, reward and language. *Schizophrenia Bulletin*, 2018, 44(1), 42.
- Elvevåg B., Fisher J., Gurd J., Goldberg T. Semantic clustering in verbal fluency: schizophrenic patients versus control participants. *Psychological Medicine*, 2002, 32(5), 909–917.
- Elvevag B., Foltz P.W., Weinberger D.R., Goldberg T.E. Quantifying incoherence in speech: an automated methodology and novel application to schizophrenia. *Schizophrenia Research*, 2007, 93(1–3), 304–316.
- Ketteler D., Theodoridou A., Ketteler S., Jager M. High order linguistic features such as ambiguity processing as relevant diagnostic markers for schizophrenia. *Schizophrenia Research and Treatment*,

Lafferty J., McCallum A., Pereira F.C. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. Proceedings of the International Conference on Machine Learning (ICML-2001), 2001, 282–289.

Liddle P.F., Ngan E.T., Caissie S.L., Anderson C.M., Bates A.T., Quedstedt D.J., White R., Weg R. Thought and Language Index: an instrument for assessing thought and language in schizophrenia. The British Journal of Psychiatry, 2002, 181(4), 326–330.

Mac Kim S., Wang Y., Wan S., Paris C. Data61-csiro systems at the clpsych 2016 shared task. Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology, San Diego: ACL, 2016. pp. 128–132.

Malmasi S., Zampieri M., Dras M. Predicting post severity in mental health forums. Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology, San Diego: ACL, 2016. pp. 133–137.

Milne D.N., Pink G., Hachey B., Calvo R.A. Clpsych 2016 shared task: Triaging content in online peer-support forums. Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology, San Diego: ACL, 2016. pp. 118–127.

Moro A., Bambini V., Bosia M., Anselmetti S., Riccaboni R., Cappa S.F., Smeraldi E., Cavallaro R. Detecting syntactic and semantic anomalies in schizophrenia. Neuropsychologia, 2015, Vol. 79, 147–157.

Osipov G., Smirnov I., Tikhomirov I., Shelmanov A. Relational-situational method for intelligent search and analysis of scientific publications. Proceedings of the Integrating IR Technologies for Professional Search Workshop, Moscow, 2013. pp. 57–64.

Pennebaker J.W. Writing about emotional experiences as a therapeutic process. Psychological science, 1997, 8(3), 162–166.

Pennebaker J.W., Boyd R.L., Jordan K., Blackburn K. The development and psychometric properties of LIWC2015. Austin, TX: University of Texas at Austin, 2015.

Pestian J.P., Matykiewicz P., Linn-Gust M., South B., Uzuner O., Wiebe J., Cohen K.B., Hurdle J., Brew C. Sentiment Analysis of Suicide Notes: A Shared Task. Biomed Inform Insights, 2012, 5(1), 3–16.

Resnik P., Armstrong W., Claudino L., Nguyen T. The University of Maryland CLPsych 2015 shared task system. Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, Denver, Colorado: Association for Computational Linguistics, 2015. pp. 54–60.

Ribalov A., Zislin J. Entropy of psychotic speech. Mathematical analysis. Ninth Annual Meeting of the Israel society for biological psychiatry. Hagoshrim, 2005, 25–30.

Yang H., Willis A., de Roeck A., Nuseibeh B. A hybrid model for automatic emotion recognition in suicide notes. Biomed Inform Insights, 2012, 5(1), 17–30.

Поступила в редакцию 26 августа 2018 г. Дата публикации: 31 октября 2018 г.

[Сведения об авторах](#)

Ениколопов Сергей Николаевич. Кандидат психологических наук, доцент, заведующий отделом

медицинской психологии, Научный центр психического здоровья, Каширское шоссе, д. 34, 115230 Москва, Россия.

E-mail: enikolopov@mail.ru

Медведева Татьяна Игоревна. Младший научный сотрудник, отдел медицинской психологии, Научный центр психического здоровья, Каширское шоссе, д. 34, 115230 Москва, Россия.

E-mail: medvedeva.ti@gmail.com

Воронцова Оксана Юрьевна. Научный сотрудник, Научный центр психического здоровья, Каширское шоссе, д. 34, 115230 Москва, Россия.

E-mail: okvorontsova@inbox.ru

Чудова Наталья Владимировна. Кандидат психологических наук, старший научный сотрудник, Институт проблем искусственного интеллекта, Федеральный исследовательский центр «Информатика и управление» Российской академии наук», пр-т 60-летия Октября, д. 9, 117312 Москва, Россия.

E-mail: nchudova@gmail.com

Кузнецова Юлия Михайловна. Кандидат психологических наук, старший научный сотрудник, Институт проблем искусственного интеллекта, Федеральный исследовательский центр «Информатика и управление» Российской академии наук», пр-т. 60-летия Октября, д. 9, 117312 Москва, Россия.

E-mail: kuzjum@yandex.ru

Пенкина Марина Юрьевна. Старший преподаватель, кафедра общей психологии, Институт экспериментальной психологии, Московский государственный психолого-педагогического университет, ул. Сретенка, д. 29, 127051 Москва, Россия.

E-mail: mpenkina@mail.ru

Минин Алексей Николаевич. Старший преподаватель, кафедра социологии и социальной работы, Курганский государственный университет, ул. Томина, д. 40, 640000 Курган, Россия.

E-mail: Aminlex@yandex.ru

Станкевич Максим Алексеевич. Инженер, Институт проблем искусственного интеллекта, Федеральный исследовательский центр «Информатика и управление» Российской академии наук», пр-т. 60-летия Октября, д. 9, 117312 Москва, Россия.

E-mail: stankevich@isa.ru

Смирнов Иван Валентинович. Кандидат физико-математических наук, доцент, заведующий отделом “Интеллектуальный анализ информации”, Институт проблем искусственного интеллекта, Федеральный исследовательский центр «Информатика и управление» Российской академии наук», пр-т. 60-летия Октября, д. 9, 117312 Москва, Россия.

E-mail: ivs@isa.ru

Любавская Анастасия Александровна. Младший научный сотрудник, отдел медицинской психологии, Научный центр психического здоровья, Каширское шоссе, д. 34, 115230 Москва, Россия.

E-mail: stesi-94@list.ru

[Ссылка для цитирования](#)

Стиль psystudy.ru

Ениколопов С.Н., Медведева Т.И., Воронцова О.Ю., Чудова Н.В., Кузнецова Ю.М., Пенкина М.Ю.,

Минин А.Н., Станкевич М.А., Смирнов И.В., Любавская А.А. Лингвистические характеристики текстов психически больных и здоровых людей. Психологические исследования, 2018, 11(61), 1. <http://psystudy.ru>

Стиль ГОСТ

Ениколопов С.Н., Медведева Т.И., Воронцова О.Ю., Чудова Н.В., Кузнецова Ю.М., Пенкина М.Ю., Минин А.Н., Станкевич М.А., Смирнов И.В., Любавская А.А. Лингвистические характеристики текстов психически больных и здоровых людей // Психологические исследования. 2018. Т. 11, № 61. С. 1. URL: <http://psystudy.ru> (дата обращения: чч.мм.гггг).

[Описание соответствует ГОСТ Р 7.0.5-2008 "Библиографическая ссылка". Дата обращения в формате "число-месяц-год = чч.мм.гггг" – дата, когда читатель обращался к документу и он был доступен.]

Адрес статьи: <http://psystudy.ru/index.php/num/2018v11n61/1622-enikolopov61.html>

[К началу страницы >>](#)