

УДК 81

# ЭЛЕКТРОННАЯ БАЗА ДАННЫХ О ЯЗЫКОВОЙ НОРМЕ И ЕЕ ВАРИАНТНОСТИ КАК ОСНОВА НАУЧНЫХ ИССЛЕДОВАНИЙ ОРТОЛОГИЧЕСКОГО ПАРАМЕТРА ИДИОЛЕКТА<sup>1</sup>

**ЗАГОРОВСКАЯ Ольга Владимировна,**

доктор филологических наук, профессор  
кафедры русского языка, современной русской и зарубежной литературы;

**ЛИТВИНОВА Татьяна Александровна,**

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

**АННОТАЦИЯ.** Статья посвящена малоисследованным проблемам организации идиолекта носителя современного русского языка и вопросам его изучения с учетом современных научных достижений в области корпусной и компьютерной лингвистики. Предметом рассмотрения впервые является ортологический параметр идиолекта и пути его определения, в качестве которых авторы предлагают использовать специально созданную электронную базу данных стилистических вариантов русского слова (по данным языкового сознания носителей русского языка). Предлагается авторская концепция названной ЭБД, описываются ее особенности и перспективы применения как основы выделения ортологического параметра идиолекта.

**КЛЮЧЕВЫЕ СЛОВА:** идиолект, параметр идиолекта, электронная база данных, языковая норма, стилистическая норма, вариант нормы, ортология, корпусная лингвистика, языковое сознание, идентификация личности по тексту.

LANGUAGE NORM ELECTRONIC DATABASE AND ITS VARIABILITY AS BASIS FOR IDIOLECT ORTHOLOGICAL PARAMETER RESEARCH

**ZAGOROVSKAYA O.V.,**

Dr. Philolog. Sci., Professor of the Department of Russian Language, Modern Russian and Foreign Literature, Voronezh State Pedagogical University;

**LITVINOVA T.A.,**

Cand. Philolog. Sci., Research Fellow of the Regional Centre of the Russian Language, Voronezh State Pedagogical University

**ABSTRACT.** The article is devoted to problems of the modern Russian native speaker's idiolect organization and its study using modern scientific achievements in the field of corpus and computer linguistics. For the first time in research, the subject is the orthological parameter of the idiolect and ways of defining it with the help of a specially created electronic database of stylistic versions of the Russian word (according to the data of the linguistic consciousness of Russian speakers). The authors suggest a concept of the electronic database, describes its characteristics and application prospects as the basis of determining the idiolect orthological parameter.

**KEY WORDS:** idiolect, idiolect parameter, electronic database, language norm, stylistic norm, variant of the norm, orthology, corpus linguistics, language consciousness, text-based personality identification.

**В** настоящее время в мировой лингвистике наблюдается всплеск интереса к идиолекту как «индивидуальному языку личности», «индивидуальному варианту общенародного языка» [18, с. 119] или – в иных определениях – как к «совокупности особенностей, характеризующих речь индивида» [3, с. 143], «совокупности языковых привычек (паттернов), характерных для индивида», «индивидуализированных выборов языковых единиц» [21], то есть к той сущности, которая представляет собой результат реализации свойств языковой системы в условиях речи отдельного индивида.

Основными причинами значительной актуализации указанной проблематики, выявившей, помимо всего прочего, неоднозначность в понимании термина «идиолект» отечественными и зарубежными исследователями, различия в трактовке соотношения данного термина с номинацией «идиостиль», принятой прежде всего в русской и российской лингвистике (см., например: [1]), стали изменения в языковом общении, бурное развитие интернет-коммуникации, появление множества текстов «рядовых» носителей языка в цифровом виде и практическая необходимость анализа подобных текстов в идентификационном аспекте с целью установления личности пишущих. Появление в Сети большого количества анонимных текстов, в том числе текстов криминальной и экстремистской направленности, привело к необходимости разработки новых методов их анализа с целью установления авторства. В мировой науке сформировалась даже особая область исследований – digital text forensics, соче-

<sup>1</sup> Исследование выполнено при поддержке гранта Российского научного фонда № 18-78-10081 «Моделирование идиолекта носителя современного русского языка в аспекте идентификации автора текста».

© Загоровская О.В., Литвинова Т.А., 2018

<sup>2</sup> Информация для связи с авторами: olzagor@yandex. ru

тающая в себе методы математической лингвистики, анализа больших данных, автоматической обработки текстов, судебной лингвистики [20].

Несмотря на то, что проблемы идиолекта (идиостиля) не являются для лингвистики новыми, а задачи исследования реализации свойств языковой системы в речи (тексте) отдельной личности ставятся в языкознании на протяжении многих десятилетий (см., например, обзор: [16]), а в настоящее время весьма активно рассматриваются в плане решения прикладных задач атрибуции текста [19], в данной области остается много нерешенных вопросов, связанных с соотношением механизмов устойчивости и изменчивости идиолекта, характеристиками стабильных и вариативных параметров идиостиля, степени их стабильности/вариативности в пространстве всех речевых произведений носителя языка, а также выявлением устойчивых и вариативных квантифицируемых языковых признаков, обладающих неодинаковыми различительными способностями в аспекте идентификации личности, применительно к идиалектам носителей разных языков, в том числе русского [23].

Задачи, связанные с указанной проблематикой, находятся в центре внимания и специальной лаборатории Лаборатории корпусной социолингвистики и автороведческих исследований (<http://rusprofilinglab.ru>), созданной на базе Воронежского государственного педагогического университета и ведущей научные изыскания в тесном контакте с Региональным центром русского языка названного вуза.

Как показывают исследования, проведенные учеными, работающими в указанных научных структурах, к числу стабильных параметров идиолекта может быть отнесен и так называемый «ортологический» (от греческого *ortos* – 'правильный') параметр, связанный с отношением продуцента текста к языковой норме и выбору ее вариантов. Правомомерность включения ортологического параметра в структуру идиолекта носителя русского языка подтверждается, с одной стороны, реальной практикой специалистов в области судебного автороведения, всегда учитывающих однотипность ошибок, связанных с нарушением языковых норм, при атрибуции текстов (см. об этом, например: [2; 17]), а с другой – достижениями современной теоретической лингвистики, теории языковой нормы и русской ортологии, доказавших особую значимость нормативных/ненормативных выборов в организации языковой личности и языкового сознания носителя русского языка.

Как свидетельствуют наши исследования, проведение научных разысканий по определению ортологического параметра идиолекта современного носителя русского языка как совокупности его устойчивых ортологических (нормативных/ненормативных) выборов с необходимостью предполагает подготовку электронных баз данных (ЭБД), содержащие сведения о языковой норме, ее речевой реализации, вариативности и представленности в языковом сознании современных носителей русского языка. При этом принципиально важным является не только учет различных типов языковой нормы (о типологии языковых норм см.: [6; 12]) и особенностей их реализации в разных вариантах существования (модусах) и разных жанрах русской речи (см. об этом: [4; 9; 11]), но и опора на специально созданные корпуса текстов [5; 10; 13; 15; 22].

В настоящее время нами подготовлена экспериментальная версия ЭБД *Russtylnorm*, ориентированная на изучение стилистических норм современного русского языка и нормативно-стилистических вариантов русского слова, представленных в языковом сознании современных молодых носителей русского языка (школьников и студентов). Названная база создавалась на основе теоретических разысканий в области проблем вариативности нормы, лексикографической параметризации стилистической значимости русского слова [8] в компьютерных базах данных и лингвистического портретирования современной российской молодежи [7] на основе широких полевых исследований. Респондентами при этом выступали молодые люди в возрасте от 15 до 25 лет, учащиеся средних и высших учебных заведений, проживающие в различных регионах России, в том числе крупных столичных и небольших провинциальных городах. Полевые исследования проводились в Воронеже, Москве, Петербурге, Нижнем Новгороде, Тамбове, Липецке, Ростове, Пензе, а также в Воронежской, Тамбовской и Липецкой областях. Общая численность респондентов составила 4000 человек: 2000 школьников и 2000 студентов.

ЭБД *Russtylnorm* содержит информацию двух типов: а) о нормативно-стилистических и собственно стилистических характеристиках основных типологических вариантов слов русского языка второго десятилетия XXI века, представленных в языковом сознании российской молодежи (школьников и студентов) с учетом их соответствия или несоответствия кодифицированным нормам русского литературного языка, отраженным в современных ортологических словарях; б) о количественных показателях анализируемых языковых явлений.

С точки зрения структуры названная ЭБД представляет собой функционирующий в компьютерной среде комплекс из четырех составляющих, которыми являются:

- 1) вариантный словник (ВС);
- 2) специализированная информационно-стилистическая база данных лексикографического типа (ИстБД);
- 3) иллюстративно-текстовая база данных, или автоматический конкорданс (ИТБД);
- 4) комплекс лингвистических алгоритмов и программ (КЛАП).

Вариантный словник представляет собой массив стилистических вариантов слов русского языка (собственно вариантов и стилистических вариантов-аналогов), который построен в виде дерева бинарных оппозиций, представленных теми или иными стилистически противопоставленными вариантами словесных знаков, и разделен на шесть подгрупп в соответствии с типом названных лингвистических феноменов. В ВС включены следующие типы вариантов слов современного русского языка: 1) произносительные (*компьют[е]р* – *компьют[э]р*, *двое[ш]ник* – *двое[ч]ник*); 2) акцентологические (*договОр* – *дОговор*, *украИнский* – *украАинский*); 3) морфологические (*договоРЫ* – *договоРА*, *грамОВ* – *грамМ*); 4) словообразовательные (*картофель* – *картошка*, *общежитие* – *общага*); 5) орфографические и графические (*Интернет* – *интернет*, *никнейм* – *ник-нейм*); 6) варианты-аналоги (*солдат* – *воин*, *отец* – *папа*). Общее количество словесных единиц в созданном вариантном словнике составляет 1200. Каждая группа вариантов представлена 200 единицами, объединенными в пары стилисти-

ческих оппозиций. Соответственно в каждой группе вариантного словника представлено по 100 оппозиций. Единицами ВС словника являются слова, формы слов или устойчивые словосочетания (*мобильный телефон – мобильник, зачетная книжка – зачетка, электронная почта – мыло* и т.п.).

Формирование ВС для ЭБД *Russtylnorm* проводилось в несколько этапов. Первоначально на основе анализа нормативных общих и школьных словарей русского языка различных типов было определено 840 различающихся стилистическими характеристиками парных словесных знаков, представляющих собой собственно варианты одного и того же слова разных типов. При этом основными критериями включения лексических единиц в формируемый словник были их отнесенность к актуальной лексике современного русского языка и вхождение в так называемые «зоны ошибок», анализируемые в современных школьных и вузовских учебных пособиях по культуре речи. В дальнейшем (на втором этапе выполнения работ) по результатам проведения полевых исследований и анализа полученного языкового материала, представляющего живую устную и письменную речь современных молодых носителей русского языка, или так называемую «естественную письменную речь» [14], исходный ВС был скорректирован и расширен: помимо собственно вариантов, в него были включены весьма частотные в современном молодежном дискурсе так называемые варианты-аналоги, представляющие собой, как правило, стилистические синонимы для обозначения тех или иных реалий (*есть – жрать, до свидания – пока*). С другой стороны, из формируемого ВС были исключены лексико-семантические варианты как менее значимые для целей проводимого исследования. В результате корректировки и дополнений объем вариантного словника достиг вполне репрезентативного показателя в 1200 единиц.

Специализированная информационно-стилистическая база данных лексикографического типа (ИстБД) является центральной составляющей сформированной ЭБД. Именно в ней содержатся сведения о тех нормативно-стилистических и собственно стилистических характеристиках основных типологических вариантов слов русского языка второго десятилетия XXI века, которые представлены в языковом сознании российской молодежи (школьников и студентов) и далеко не всегда совпадают с соответствующей информацией нормативных словарей современного русского языка. ИстБД представляет собой массив информационных данных, организованных по типу словарных статей в электронном словаре и созданных в диалоговом и автоматическом режимах на основе обработки информации, полученной в результате полевых исследований языкового сознания современных молодых носителей русского языка (целенаправленного интервьюирования информантов), изучения информации, представленной в современных нормативных словарях русского языка, в том числе школьных, а также в результате обработки полученной по итогам выполнения проекта текстовой базы данных.

Структура ИстБД представлена тремя типами составляющих (информационных зон): 1) зона информации о стилистических характеристиках того или иного варианта словесного знака, представленного в языковом сознании современных молодых носителей русского языка; 2) зона количественных показателей названных характеристик с учетом

двух основных групп респондентов: школьников и студентов; 3) зона соответствия или несоответствия названных характеристик, представленных в языковом сознании современной российской молодежи, кодифицированным нормам русского литературного языка, закрепленным в современных ортологических словарях, в том числе школьных.

Первая из названных зон включает в себя данные о 1) функционально-стилистических, 2) экспрессивно-стилистических, 3) квантитативно-стилистических и 4) дискурсивно-стилистических характеристиках того или иного варианта словесного знака. Указанная информация передается с помощью системы специальных обозначений, некоторые из которых соответствуют традиционным стилистическим пометам русских словарей, а другие являются авторскими: *литер.*, *общелитер.* (литературное, общелитературное, т.е. общенормативное, допустимо во всех сферах общения и всех стилевых разновидностях русского литературного языка); *литер., разг.* (литературное, разговорное, т.е. нормативное, но допустимо только в бытовой сфере общения и разговорно-бытовом стиле русского литературного языка); *жарг.* (жаргонное); *прост.* (просторечное, в том числе нарушающее собственно языковые нормы разных типов или нормы морально-этические); *высок.* (высокое, торжественное), *сниж.* (сниженное), *нейтр.* (нейтральное), *част.* (частотное), *редк.* (редкое), *допуст. в любой сит.* (допустимо в любой ситуации), *недопуст. в какой-л. сит.* (недопустимо в какой-либо ситуации, например, в общении со старшими). Например, для грамматического варианта *договора* в зоне информации о стилистических характеристиках словесного знака ИстБД в соответствии с результатами проведенных полевых исследований языкового сознания современных молодых носителей русского языка указано: *литер., общелитер.; част.; допуст. в любой сит.* (Очевидно, что приведенные стилистические характеристики названного грамматического варианта не соответствуют общепринятым представлениям о содержании стилистического компонента значения указанного словесного знака, отраженным в современных нормативных словарях и справочниках по культуре русской речи.) Содержание сведений, указанных в зоне информации о стилистических характеристиках того или иного варианта словесного знака, представленных в языковом сознании современных молодых носителей русского языка, было определено по результатам статистического анализа материалов, полученных в ходе проведенных полевых исследований.

Вторая зона ИстБД (зона количественных показателей стилистических характеристик того или иного варианта словесного знака) содержит выраженные в процентах (применительно к двум обследуемым категориям испытуемых: школьникам и студентам) данные о количестве респондентов, предложивших приведенные стилистические характеристики. Третья зона (зона соответствия или несоответствия стилистических характеристик словесного знака, представленных в языковом сознании современной российской молодежи, кодифицированным нормам русского литературного языка, закрепленным в современных ортологических словарях) включает в себя данные, выделенные на основе сравнения информации, полученной в результате полевых исследований, с информацией ортологических словарей. Названная зона включает в себя две подзоны: а) подзону непосредственной инфор-

мадии о соответствии или несоответствии стилистических характеристик словесного знака, представленных в языковом сознании современной российской молодежи, кодифицированным нормам русского литературного языка, закрепленным в современных ортологических словарях; б) подзону справочно-библиографических сведений, содержащую отсылки к материалам нормативных словарей русского языка. Например, для варианта *договора* (по данным исследования языкового сознания для 65% современных студентов и школьников такой вариант является общелитературным и допустимым во всех сферах общения) в первой из названных подзон указывается: «Не соответствует нормам современного русского литературного языка, закрепленным в нормативных словарях»; во второй подзоне приводятся ссылки на нормативные словари русского языка: БАС-2; Словарь правильной русской речи Л.И. Скворцова, Большой орфоэпический словарь русского языка под ред. Л.Л. Касаткина и др.

Следует подчеркнуть, что в ИстБД получены отражение прежде всего результаты тех полевых исследований, которые проводились методом целенаправленного интервьюирования респондентов, проводившегося на основе 6 типов специально разработанных анкет (в соответствии с шестью типами исследуемых вариантов слова, входящих в вариантный словник). В каждой из анкет были перечислены 200 вариантов определенного типа (произносительные, акцентологические, словообразовательные, морфологические, орфографические и графические, варианты-аналоги) и указаны параметры, по которым требовалось получить информацию от респондентов. Включенные в анкеты параметры соответствовали основным системным и дискурсивным стилистическим характеристикам слова в современном русском языке. Предлагаемые в анкетах формулировки были построены с учетом уровня лингвистической компетенции среднего носителя русского языка и максимально адаптированы для восприятия современными школьниками и студентами (к каждому параметру в скобках давались разъяснения).

Использование для подготовки Специализированной информационно-стилистической базы данных описанных выше методов сбора и обработки материала обеспечило достаточно высокий уровень достоверности полученных сведений, что было доказано специально проведенным экспериментальным тестированием ИстБД.

Иллюстративно-текстовая база данных (ИТВД), или автоматический конкорданс, как третья составляющая ЭБД *Russtylnorm* представляет собой корпус текстов, включающий три подкорпуса: 1) подкорпус текстов естественной письменной речи, продуцированной респондентами в условиях эксперимента; 2) подкорпус текстов естественной устной речи молодых носителей русского языка, переведенной в письменную форму в упрощенной фонетической транскрипции, которая позволяет передать особенности произношения звуков и постановки ударений; 3) подкорпус «дополненной» чужой речи (тексты, написанные респондентами по предложенной основе с заданиями вставить те или иные буквы, морфемы, слова, формы слов).

В основу первого из названных корпусов были положены материалы имеющихся у разработчиков проекта «молодежных» корпусов письменной речи «RusPersonality» и «Россия и мир глазами студентов», дополненные материалами, полученными в результате полевых исследований и сбора новых

образцов «естественной» письменной речи по заданным темам («Как я провожу свободное время», «Интересный случай в моей жизни», «Что я делал вчера (на каникулах)», «Что мне нравится и что не нравится в школе (в институте, университете)». Общий объем данного подкорпуса – 800 000 словоупотреблений.

Второй подкорпус создавался по результатам полевого сбора материала по специально разработанной программе, предполагающей предъявление респондентам тем для устных монологических высказываний, записи звучащей речи на диктофон и последующего ее перевода в письменную форму в упрощенной фонетической транскрипции. Темы, включенные в программу формирования данного подкорпуса, соответствовали темам, используемым для работ по созданию подкорпуса естественной письменной речи. Объем названного подкорпуса – 200 000 словоупотреблений.

Третий подкорпус представлен текстами, написанными респондентами в условиях направленного эксперимента по заданиям на восстановление тех или иных пропущенных в предложенном тексте (произведении «чужой» речи) элементов, касающихся выбора написания слова, выбора его грамматической формы или словообразовательного варианта, выбора варианта-аналога (стилистического синонима) в зависимости от контекста или указанной ситуации общения. Основу для составления заданий и подготовки исходных текстов, предлагаемых респондентам для восстановления, составили речевые произведения, извлеченные из реальной коммуникации молодых носителей русского языка в социальных сетях. Объем данного подкорпуса – 300 000 словоупотреблений. Все тексты, представленные в созданной ИТВД, снабжены метатаржеткой, включающей информацию о социальных (школьник или студент), возрастных и гендерных характеристиках информантов, а также о регионе записи данного текста и времени сбора материала.

Комплекс лингвистических алгоритмов и программ (КЛАП) как четвертая составляющая раскрываемой ЭБД представляет собой специальный программный комплекс, включающий в себя программы лемматизации и автоматизированного поиска необходимой информации в любой составляющей созданной ЭБД. С помощью данного комплекса в диалоговом режиме возможно получение информации как в направлении от вариантного словника (с целью получения информации, касающейся стилистических характеристик того или иного представленного в словнике языкового знака, количественных показателей его стилистических оценок в языковом сознании современных школьников и студентов, соответствии или несоответствии данных оценок кодифицированным нормам русского литературного языка), так и в направлении от иллюстративно-текстовой базы к вариантному словнику и специализированной информационно-стилистической базе данных (с целью получения информации об использованном в тексте том или ином варианте или уровне соблюдения норм русского литературного словоупотребления в тексте в целом).

Анализ материалов ЭБД *Russtylnorm* позволяет делать существенные выводы, касающиеся не только речевой культуры современных молодых носителей русского языка, различий между «реальной» и кодифицированной нормой, динамических процессов в изменении стилистических норм русского языка, но и выводы, имеющие отношение к ор-

ганизации и особенностям идиолектов представитель современной лингвокультуры и существенных характеристиках самого понятия «ортологический параметр идиолекта». Представляется, что ЭБД *Russtylnorm* может рассматриваться как готовая основа для проведения специальных исследований по выявлению устойчивых выборов современных носителей русского языка не только применительно к тем или иным видам стилистических норм, но и ортологического компонента идиолекта носителя русского языка в целом. Очевидно, что для решения названной задачи должен быть создан специализированный корпус текстов, представляющих не только речевые произведения разных авторов, но и

тексты одних и тех же авторов в различных языковых модусах и жанрах, что позволит провести статистический анализ, направленный на определение уровня стабильности и вариативности нормативных выборов автора текста и впоследствии – статистический анализ, направленный на оценку идентификационной способности ортологического параметра идиолекта в целом, и будет способствовать построению модели идиолекта, структура которого включает в себя зоны квантифицируемых признаков, обладающих неодинаковой степенью различительной способности для идентификации личности по тексту.

#### СПИСОК ЛИТЕРАТУРЫ:

1. Богданова, Е.В. О некоторых аспектах изучения термина «идиолект» в отечественной и западной лингвистике [Текст] / Е.В. Богданова // Вестник Ленинградского государственного университета им. А.С. Пушкина. Серия: Филология. – 2009. – Т. 1. – №4. – С. 100–108.
2. Галяшина, Е.И. Автороведческая экспертиза в российском судопроизводстве [Текст] / Е.И. Галяшина, Е.В. Приводнова // Lex Russica. – 2006. – №4. – С. 55–61.
3. Жеребило, Т.В. Словарь лингвистических терминов и понятий [Текст] / Т.В. Жеребило. – Изд. 6-6, испр. и доп. – Назрань : Пилигрим, 2016. – 610 с.
4. Загоровская, О.В. Вариантность нормы в русском языке начала XXI века и задачи создания современного учебного нормативно-стилистического словаря [Текст] / О.В. Загоровская // Известия Воронежского государственного педагогического университета. – 2015. – №1(266). – С. 185–191.
5. Загоровская, О.В. Диагностирование склонности автора письменного текста к аутоагрессивному поведению [Текст] / О.В. Загоровская, Т.А. Литвинова, О.А. Литвинова, П.В. Середин, М.Е. Сердюк // Вестник Воронежского государственного университета. Серия: Лингвистика и межкультурная коммуникация. – 2015. – №3. – С. 98–104.
6. Загоровская, О.В. Нормы русского литературного языка: Типология и основания для классификации [Текст] / О.В. Загоровская // Известия Воронежского государственного педагогического университета. – 2016. – №3(272). – С. 129–134.
7. Загоровская, О.В. Речевое портретирование социальной группы как задача современной лексикографии [Текст] / О.В. Загоровская // Вестник Воронежского государственного университета. Серия: Лингвистика и межкультурная коммуникация. – 2013. – №2. – С. 174–177.
8. Загоровская, О.В. Стилистическая значимость русского слова как явление национальной культуры [Текст] / О.В. Загоровская // Русское национальное сознание в его языковом воплощении: прошлое, настоящее, будущее. XXX Распоповские чтения : материалы Международной конференции, Воронеж, 2–4 марта 2012 г. / под ред. Л.М. Кольцовой. – Воронеж : Издательско-полиграфический центр Воронежского государственного университета, 2012.
9. Загоровская, О.В. Типологические разновидности национального русского языка и формы его существования в начале XXI века [Текст] / О.В. Загоровская // Известия Воронежского государственного педагогического университета. – 2015. – №3(268). – С. 96–101.
10. Загоровская, О.В. Электронный корпус студенческих эссе на русском языке и его возможности для современных гуманитарных исследований [Текст] / О.В. Загоровская, Т.А. Литвинова, О.А. Литвинова // Мир науки, культуры, образования. – 2012. – №3(34). – С. 387–389.
11. Загоровская, О.В. Языковая норма в современной русской визуально-письменной речи, функционирующей в интернет-коммуникации: к постановке проблемы [Текст] / О.В. Загоровская // Известия Воронежского государственного педагогического университета. – 2017. – №4(277). – С. 168–172.
12. Загоровская, О.В. Языковая норма и норма литературного языка как лингвистические понятия [Текст] / О.В. Загоровская // Известия Воронежского государственного педагогического университета. – 2016. – №2(271). – С. 161–165.
13. Загоровская, О.В. Корпус текстов RusPersonality как основа исследований «реальной» языковой нормы в современной русской письменной речи [Текст] / О.В. Загоровская, Т.А. Литвинова // Современные проблемы лингвистики и методики преподавания русского языка в вузе и школе / под ред. О.В. Загоровской. – Воронеж : Научная книга, 2018. – Вып. 28. – С. 51–57.
14. Лебедева, Н.Б. Естественная письменная русская речь: проблемы изучения [Текст] / Н.Б. Лебедева // Русский язык: исторические судьбы и современность: Международный конгресс исследователей русского языка : труды и материалы. – М., 2001. – С. 260–261.
15. Литвинова, Т.А. Возможности компьютерной лингвистики для решения задач диагностирования личности по тексту (на материале корпуса текстов Personality) [Текст] / Т.А. Литвинова // Вестник Воронежского государственного университета. Серия: Лингвистика и межкультурная коммуникация. – 2015. – №3. – С. 37–41.
16. Литвинова, Т.А. Идентификация и диагностирование автора письменного текста : монография [Текст] / Т.А. Литвинова, О.А. Литвинова. – Воронеж : Воронежский государственный педагогический университет, 2015. – 332 с.

17. Маркова, Г.Д. Идентификационные признаки письма в советской криминалистической экспертизе : автореф. дис. ... канд. юрид. наук [Текст] / Г.Д. Маркова. – Харьков, 1956. – 24 с.
18. Матвеева, Т.В. Полный словарь лингвистических терминов [Текст] / Т.В. Матвеева. – Ростов н/Д : Феникс, 2010. – 562 с.
19. Coulthard, M. Author identification, idiolect, and linguistic uniqueness / M. Coulthard // Applied Linguistics. – 2004. – Т. 24. – №4. – С. 431–447.
20. Gollub, T. Recent Trends in Digital Text Forensics and Its Evaluation / T. Gollub [et. al. ] // Lecture Notes in Computer Science. – 2013. – Vol. 8138. – С. 282–302.
21. Grant, T. Resources and constraints in linguistic identity performance: a theory of authorship / T. Grant, N. Macleod // Language and Law/Linguagem e Direito. – 2018. – Т. 5. – №1. – С. 80–96.
22. Litvinova, T. "RusPersonality": A Russian corpus for authorship profiling and deception detection / T. Litvinova, O. Litvinova, O. Zagorovskaya, P. Seredin, A. Sboev, O. Romanchenko // Proceedings of the International FRUCT Conference on Intelligence, Social Media and Web (ISMW FRUCT 2016). – IEEE. – С. 1–7.
23. Litvinova, T. On the Stability of Some Idiolectal Features / T. Litvinova, T. Dankova, O. Zagorovskaya // Lecture Notes in Computer Science. – 2018. – Vol. 11096. – С. 331–336.