

Литвинова Татьяна Александровна, Литвинова Ольга Александровна,
Диброва Елена Владимировна, Рыжкова Екатерина Сергеевна

КОРПУСНЫЕ ИССЛЕДОВАНИЯ ПИСЬМЕННОЙ РЕЧИ В РЕШЕНИИ ЗАДАЧ СУДЕБНОГО АВТОРОВЕДЕНИЯ

Обосновывается необходимость использования специальным образом составленных корпусов текстов при разработке методик проведения идентификационной и диагностической автороведческой экспертизы. Приводится краткий анализ существующих корпусов текстов, используемых для построения методик идентификации и диагностирования личности автора текста. Описывается структура корпуса текстов на русском языке Personality, содержащего метаразметку в виде информации об их авторах, и анализируются возможности его применения для решения задач судебного автороведения.

Адрес статьи: www.gramota.net/materials/2/2015/8-1/29.html

Источник

Филологические науки. Вопросы теории и практики

Тамбов: Грамота, 2015. № 8 (50): в 3-х ч. Ч. I. С. 107-113. ISSN 1997-2911.

Адрес журнала: www.gramota.net/editions/2.html

Содержание данного номера журнала: www.gramota.net/materials/2/2015/8-1/

© Издательство "Грамота"

Информация о возможности публикации статей в журнале размещена на Интернет сайте издательства: www.gramota.net
Вопросы, связанные с публикациями научных материалов, редакция просит направлять на адрес: phil@gramota.net

9. **Маркштайн Э.** Дом и котлован, или Мнимая реализация утопии // Андрей Платонов: в 2-х кн. М.: Современный писатель, 1994. Кн. 2. С. 284-302.
10. **Махлина С. Т.** Семиотика культуры повседневности. М.: Алетейя, 2009. 232 с.
11. **Степанов Ю. С.** Концепт «духовного странничества» в России XIX и XX века // Русское подвижничество / сост. Т. Б. Князевская. М.: Наука, 1996. С. 29-43.
12. **Топоров В. Н.** Петербургский текст / Отделение историко-филологических наук РАН. М.: Наука, 2009. 820 с.
13. **Хлынина Т. П.** Жилое пространство и особенности его организации в советской России в 1920-1940-е гг. // Былые годы. Российский исторический журнал. 2013. № 1 (27). С. 61-70.
14. **Цивьян Т. В.** Дом в фольклорной модели мира (на материале балканских загадок) // Труды по знаковым системам. Тарту, 1978. Вып. 10. С. 65-85.
15. **Черных А. И.** Жилищный передел. Политика 20-х годов в сфере жилья // Социологические исследования. 1995. № 10. С. 71-78.

ON THE ANALYSIS OF THE MOTIVE "THE HOUSING PROBLEM" IN THE RUSSIAN LITERATURE OF THE 1920-1930S

Kuvshinov Feliks Vladimirovich, Ph. D. in Philology
Lipetsk Institute of Cooperation (Branch) of Belgorod University of Cooperation, Economics and Law
fkuvshinov@yandex.ru

The article analyses the motive "housing problem", which spread in the Russian literature of the 1920-1930s because of the unsolved housing problem in the everyday life of the USSR. The topicality of the housing problem as a household reality was fixed by the pre-revolution literature, however after the revolution of 1917 this problem extremely aggravated and it received a new interpretation and representation in the kind of motive, the analysis of which reveals its ambivalent nature diachronically and in terms of the expression in the framework of opposition way of life / existence.

Key words and phrases: the motive "the housing problem"; household reality; opposition way of life/existence; description of a character; attitude to the living quarters; the connection between a man and a house.

УДК 81'33

Филологические науки

Обосновывается необходимость использования специальным образом составленных корпусов текстов при разработке методик проведения идентификационной и диагностической автороведческой экспертизы. Приводится краткий анализ существующих корпусов текстов, используемых для построения методик идентификации и диагностирования личности автора текста. Описывается структура корпуса текстов на русском языке Personality, содержащего метаразметку в виде информации об их авторах, и анализируются возможности его применения для решения задач судебного автороведения.

Ключевые слова и фразы: автороведение; лингвистическая экспертиза; автороведческая экспертиза; корпус текстов; лингвистическая статистика; диагностирование личности автора текста; естественная письменная речь; атрибуция текста.

Литвинова Татьяна Александровна, к. филол. н.

Литвинова Ольга Александровна

Воронежский государственный педагогический университет
centr_rus_yaz@mail.ru; olga_litvinova_teacher@mail.ru

Диброва Елена Владимировна

Южный федеральный университет
alena_dibrova@bk.ru

Рыжкова Екатерина Сергеевна

Воронежский государственный университет инженерных технологий
ryzhkova.katerina@yandex.ru

КОРПУСНЫЕ ИССЛЕДОВАНИЯ ПИСЬМЕННОЙ РЕЧИ В РЕШЕНИИ ЗАДАЧ СУДЕБНОГО АВТОРОВЕДЕНИЯ®

Исследование выполнено при поддержке гранта РФФИ № 15-36-50104 мол_нр «Формально-грамматические параметры письменного текста и индивидуально-личностные характеристики его автора: корпусное исследование», гранта РГНФ № 15-34-01221 «Детекция лжи в письменном тексте: корпусное исследование».

Установление личности человека является одной из главнейших задач, которым посвящена судебно-экспертная деятельность [6]. Хорошо известным видом криминалистической идентификации человека является идентификация личности по речи, проводимая в рамках почерковедческой и автороведческой экспертиз.

Как показал анализ оперативно-розыскной и следственной практики, проведенный в работе [30], ситуации, связанные с необходимостью решения автороведческих задач на основе исследования произведений речи, чаще всего возникают в ходе расследования таких серьезных преступлений, как убийство, похищение человека, похищение несовершеннолетних, похищение человека с целью вымогательства и др.

Следует отметить, что в настоящее время наблюдается рост числа назначаемых автороведческих экспертиз, что, по мнению исследователей, объясняется «не только усовершенствованием законодательства и изменением правоприменительной практики, но и тем, что вопрос установления авторства либо его отрицание может стать актуальным при рассмотрении любого уголовного и гражданского дела» [8, с. 756].

В качестве непосредственного объекта автороведческой экспертизы выступает письменный текст как единство формально-языковых и содержательно-смысловых параметров, отражающих авторскую индивидуальность [12]. При этом исследователи исходят из того, что устойчивость и уникальность признаков письменной речи конкретного человека обусловлены нейробиологическими механизмами его врожденной языковой способности [Там же].

Автороведческая экспертиза традиционно решает разнообразные задачи: идентификация автора текста, диагностирование его индивидуально-личностных характеристик (определение пола, возраста, родного языка, психологических особенностей и др.), установление эмоционального состояния автора в момент создания текста и др. [13; 16; 28]. На практике идентификационные и диагностические задачи нередко решаются в рамках единого комплексного исследования: первая фаза – диагностическая, в ходе которой происходит сужение круга лиц, которые могут быть авторами текста путем моделирования личности автора спорного текста (определяются данные о степени владения языковыми навыками, данные об образовании, родном языке, половозрастной принадлежности автора письменного текста и т.п. характеристиках); вторая фаза – сравнительная, в ходе которой сравниваются речевые образцы лиц, которые могут быть авторами текста, со спорным текстом; вывод же экспертизы является идентификационным [1].

Идентификационные задачи ранее активно решались в рамках почерковедческой экспертизы, которая, однако, теряет сейчас свою значимость (за исключением графологического исследования подписей). Как справедливо отмечается исследователями, «в условиях современной действительности тексты, выполненные от руки, не утратили актуальности только в образовательной сфере, являясь частью индивидуального процесса обучения, и лишь в незначительной степени сохраняются в рамках бытового общения» [24, с. 387].

Важной проблемой современного автороведения является проблема установления автора текста интернет-коммуникации [13; 28], поскольку «преступный мир активно использует виртуальный мир в своих целях» [8, с. 758]. Так, в Сети публикуются тексты, содержащие клевету, оскорбления, призывы к разжиганию розни, экстремистские призывы, призывы к терроризму и т.д. Особенно актуальными автороведческие исследования текстов интернет-коммуникации становятся в свете последних изменений в законодательстве. Если прежде следственная практика сталкивалась только с фигурой распространителя конфликтного текста, то отныне в поле профессионального зрения оперативных работников должна войти совершенно новая фигура – автор [1].

Таким образом, в настоящее время ни у кого не вызывает сомнений тот факт, что атрибуция текста и составление портрета его автора относятся к числу важнейших задач криминалистического исследования текста.

Однако, несмотря на то, что задача атрибуции текста разрабатывается в отечественной науке на протяжении как минимум последнего столетия (см., например, [23]), многие исследователи отмечают субъективность имеющихся методик атрибуции [25, с. 17], в то время как «...юридическая строгость предполагает, прежде всего, максимальную (насколько это возможно) правовую однозначность ответа» [9, с. 5]. Кроме того, методология автороведческой экспертизы текстов на русском языке разрабатывалась преимущественно на материале художественных текстов и малоприменима к текстам малого объема, которые нередко становятся объектом автороведческой экспертизы. Как справедливо отмечают исследователи, вовлечение в сферу автороведческой экспертизы текстов разных жанров и функциональных стилей объективно требует комплексного подхода и выработки новых правил принятия идентификационных решений при установлении авторства спорного, анонимного или псевдонимного текста [8, с. 760].

В частности, особенно остро встает вопрос о составе максимально контекстно устойчивых признаков текста и признаков, релевантных при анализе текстов определенного типа [26]. До настоящего времени значимость каждого лингвистического признака и его вклад в решаемую задачу (идентификации автора текста) количественно не определены, и оценка лингвистических признаков проводится в основном на качественном уровне без учета частоты встречаемости в речи конкретного языкового явления [7].

Кроме того, единицы далеко не всех языковых уровней подвергаются анализу в автороведческих исследованиях. Так, насколько нам известно, отсутствуют работы, в которых бы производился лингвистический анализ целостности и связности письменного текста с целью установления его авторства, хотя, как совершенно справедливо указывает Е. И. Галяшина, такой анализ крайне важен при разрешении сомнений в подлинности и письменных текстов, используемых как судебные доказательства [Там же].

Еще одной проблемой идентификационных автороведческих исследований является отсутствие работ, посвященных определению пределов интериндивидуальной и межиндивидуальной вариативности каждого конкретного лингвистического признака методами статистического анализа [Там же].

Как отмечают исследователи [2], теоретическое и практическое обеспечение портретных (диагностических) экспертиз также оставляет желать лучшего. Методики комплексного решения вопросов установления пола, возраста и профессиональной принадлежности лица, составившего текст, являются неразработанными [16]. Как подчеркивает А. Н. Баранов, существенную проблему, кроме чисто теоретических аспектов портретных экспертиз, представляет их лингвистическое обеспечение, в том числе корпусная поддержка [2]. Кроме того, совершенно очевидна необходимость привлечения данных и методов широкого круга дисциплин: психологии, психолингвистики, социолингвистики, нейролингвистики и пр. для разработки подобного рода методик.

При проведении автороведческого исследования также необходимо помнить о том, что автор текста может прибегать к искажению признаков письменной речи путем маскировки или имитации с целью затруднения идентификации авторства. Проблема выявления факта искажения признаков письменной речи является слабо разработанной в мировой науке (см. обзор в [31]), несмотря на то, что такое речевое поведение представляет собой распространенную в преступной среде практику, в том числе и при коммуникации в Интернете [13]. Исследования данной проблемы на материале русского языка также единичны [11]. Одной из самых значительных работ по данной проблематике является диссертационное исследование Т. В. Гомон 1990 года [10], где предпринимаются попытки составить список признаков письменной речи, которые можно имитировать, и список глубинных признаков, с трудом поддающихся имитации. В целом же следует отметить, что измененная или искаженная речь не выделена в качестве самостоятельного объекта автороведческой экспертизы, нет научно-методических разработок, касающихся сообщений с измененной речью, что выводит искаженную речь из поля зрения эксперта-автороведа [12].

Таким образом, рост числа назначаемых автороведческих экспертиз и многообразие материалов, поступающих на экспертизу, с одной стороны, и наличие ряда неразрешенных проблем в области теории, методологии и методики судебной экспертизы письменной речи – с другой требуют разработки «принципиально новых подходов к автороведческому исследованию с учетом современных реалий создания письменных текстов и современных возможностей осуществления коммуникации» [24, с. 387]. На наш взгляд, создание таких подходов предполагает обязательное использование методов статистического анализа языкового материала и применение современных средств автоматической обработки языка. В настоящее время на материале русского языка ведутся разработки методик идентификации автора текста с применением методов математического анализа и предпринимаются попытки создания программных средств для такого анализа (см. их обзор в [27]). Однако большинство этих программ предназначено для анализа текстов большого объема и разрабатывалось на основе анализа художественных произведений. Методики же, предназначенные для решения диагностических задач автороведческого исследования, практически не подкреплены статистическими исследованиями.

Таким образом, проблема разработки автороведческих методик, основанных на количественном анализе письменных текстов, продолжает оставаться крайне актуальной и требует своего дальнейшего развития [7]. Однако разработка данных методик невозможна без соответствующего корпусного материала. При этом под корпусом языка мы, вслед за создателями Национального корпуса русского языка, понимаем информационно-справочную систему, основанную на собрании текстов на некотором языке в электронной форме [29].

Как отмечается в [3], к проблемам, затрудняющим исследования в области идентификации автора текста, относится проблема составления выборки эталонных текстов, обязательным критерием пригодности которых для автороведческого исследования является «отсутствие элементов корректуры, редактирования, дописывания другим лицом (другими лицами)» [24, с. 387]. Исследования в области диагностического автороведения, в свою очередь, значительно страдают от ограниченности корпусного материала, поскольку сбор корпусов текстов с метаразметкой в виде информации об их авторах требует значительных усилий.

Между тем, в мировой науке приобрел большую популярность подход к разработке методик диагностирования личности по тексту, состоящий в создании специальных корпусов текстов, содержащих, помимо собственно текстов, метаразметку в виде информации о личности их авторов (поле, возрасте, баллах) по шкалам психологических тестов и пр.), разметку корпусов средствами автоматической обработки языка, извлечении численных значений поддающихся квантификации параметров текста, вычислении корреляций между этими значениями и характеристиками личности автора текста и построении на их основе математических моделей для диагностирования тех или иных характеристик автора текста [19].

Кратко охарактеризуем корпуса текстов, на материале которых проводятся подобные исследования. Как показывает анализ, такие корпуса делятся на две группы – составленные из текстов, уже существовавших до момента создания корпуса, и составленные из текстов, специально написанных респондентами по заданию исследователя.

К первой группе относятся, прежде всего, корпуса текстов интернет-коммуникации (блоги, чаты, твиты, сообщения в соцсетях и др.), собранные при помощи специальных программ. Помимо самих текстов, такие программы также собирают информацию об их авторах, указанную ими самими в своих профилях (обычно – пол, возраст). Таковы корпуса, используемые для тестирования методик на проводимой ежегодно международной конференции PAN, посвященной апробации методик идентификации и диагностирования личности по тексту [33]. Подобные корпуса имеют много «шума» из-за автоматического метода сбора: собранные тексты могут содержать неавторские фрагменты (различного рода цитирования, ссылки и пр.), что может снижать точность классификации авторов по полу и возрасту почти вдвое. Кроме того, информация об авторе, указанная

в авторском профиле, может быть недостоверной, что снижает ценность такого рода корпусов. К несомненным достоинствам такого метода сбора материала относится, однако, большой объем таких корпусов, а также малое время, затраченное на их составление. Так, один из самых больших корпусов блогов на английском языке с метаразметкой в виде информации о поле и возрасте авторов содержит 140 миллионов слов [34].

Существуют также корпуса текстов интернет-коммуникации, содержащие, помимо информации о поле и возрасте авторов, результаты пройденного ими в режиме онлайн психологического тестирования. Среди корпусов такого рода – *myPersonality3*, содержащий тексты и информацию о 250 пользователях Facebook, а также корпус сообщений и данные психологического тестирования пользователей Twitter [32]. Именно на нем команды из разных стран мира – участники конференции PAN-2015 будут тестировать свои методики диагностирования автора текста (язык твитов – английский, испанский, итальянский и датский) [33].

Помимо корпусов, составленных из текстов, размещенных в Интернете, существуют также корпуса текстов, созданных специально для исследовательских целей. Одним из самых известных корпусов текстов, созданных специально для изучения проблемы взаимосвязи личности и языка, является корпус текстов студенческих эссе на английском языке, составленный под руководством американского ученого Дж. Пеннебакера [Ibidem]. Корпус собирался на протяжении 7 лет (с 1997 по 2004 год) и насчитывает 2469 текстов (по одному от каждого участника) (1,9 миллиона слов, в среднем 770 слов в тексте) и метаданные в виде информации о поле автора текста и результатов тестирования по пятифакторному личностному опроснику. Данный корпус текстов используется как материал многих исследований, в том числе и как корпус для тестирования методик диагностирования личности участников конкурса-конференции PAN [32].

Аналогов данного корпуса в мировой лингвистике не существовало, пока в 2014 году не было заявлено о создании подобного корпуса текстов на датском языке – *Stylometry Investigation (CSI) Corpus* [35]. Корпус содержит тексты, созданные студентами университета Антверпена, и находится в открытом доступе. Он содержит 749 текстов (305 000 символов), планируется ежегодное расширение корпуса за счет текстов студентов-первокурсников. Тексты представляют собой эссе на различные темы (средняя длина – 1126 символов) и отзывы о продуктах (в среднем 128 символов). Для каждого автора известны пол, возраст, место рождения, данные по пятифакторному личностному опроснику. Также по желанию авторы текстов могли указать свою сексуальную ориентацию и пройти тестирование по опроснику Майерс-Бриггс.

Насколько нам известно, других специально созданных корпусов текстов, содержащих метаданные в виде информации об авторах – поле, возрасте, результатах психологического тестирования, в мировой лингвистике нет, несмотря на очевидную ценность таких корпусов для разработки и тестирования методик диагностирования личности по тексту.

До недавнего времени отсутствовал такой корпус текстов и на русском языке. В имеющихся корпусах текстов на русском языке практически полностью отсутствует метатекстовая разметка в виде информации о личности их авторов [5], в связи с чем в 2012 году в Региональном центре русского языка при ВГПУ был начат сбор корпуса русских письменных текстов *Personality*, содержащего метаразметку в виде информации об авторах – поле, возрасте, результатах психологического тестирования, уровне образования, а также – для некоторых подкорпусов – результатах нейропсихологического обследования, информации о профессии автора текста [14; 17].

Приведем краткую информацию о структуре и составе корпуса.

1. Характеристики авторов текстов

На 23.04.2015 в корпусе *Personality* содержались тексты 1 145 респондентов, общее число текстов – 1 867 (в зависимости от типа заданий респонденты писали один или два текста). Общий объем текстового материала составлял ок. 300 000 словоупотреблений, или ок. 1 800 000 символов. Средняя длина текста варьировалась от 56 слов (подкорпус текстов респондентов, страдающих психическими заболеваниями) до 230 (подкорпус текстов эссе «В чем я вижу смысл жизни»).

Пол. В корпусе содержатся тексты 386 лиц мужского пола и 751 – женского, 8 респондентов не указали свой пол.

Возраст. В создании корпуса принимали участие лица в возрасте от 12 до 79 лет, однако основную часть корпуса (1519 текстов) на данный момент составляют тексты лиц от 18 до 22 лет – студентов вузов Воронежа, Ростова-на-Дону, Москвы.

Родной язык. Для большинства респондентов русский язык является родным языком, однако в настоящее время создается подкорпус текстов лиц, для которых русский язык хотя и не является родным, но степень владения им достаточна для прохождения психологического тестирования и продуцирования связных текстов. На момент написания статьи собрано 48 текстов от 37 лиц, для которых русский язык не является родным.

Образование. Корпус содержит тексты лиц с неполным средним образованием (т.е. школьников, 96 человек), с неполным высшим образованием (т.е. студентов, 954 человека), высшим образованием (95 человек).

Психологические характеристики. Все респонденты, за исключением лиц, страдающих психическими заболеваниями, проходили психологическое тестирование. Использовались различные опросники: пятифакторный личностный опросник (192 респондента), методика определения коммуникативной установки В. Бойко (75 респондентов), Фрайбургский многофакторный личностный опросник (550 респондентов), методика А. Белова для определения преобладающего типа темперамента (124), тест на определение акцентуации Леонгарда-Шмишека (124), тест «Домино» (54), опросник «Стиль саморегуляции поведения» (51),

госпитальная шкала тревоги и депрессии HADS (117), опросник Айзенка «Самооценка психических состояний личности», адаптированный к подростковому возрасту (158). Кроме того, 275 человек прошли обследование на определение моторного, сенсорного и когнитивного латерального профиля.

Также создается подкорпус текстов лиц, страдающих психическими заболеваниями (в настоящий момент подкорпус состоит из 15 текстов лиц, страдающих шизофренией, и 2 текстов лиц, страдающих биполярным расстройством, общее число респондентов – 17 человек).

2. Характеристики текстов

Все представленные в корпусе тексты представляют собой образцы естественной письменной речи [15]. Респондентам было предложено описать картину (540 текстов), любой день из своей жизни (228), написать письмо другу (734 текста), эссе на темы «На что бы я потратил 1 млн долларов США?» (75), «В чем я вижу смысл жизни?» (124), а также описать свои лучшие качества для потенциального работодателя (166).

Корпус текстов *Personality* содержит также подкорпус правдивых и «ложных» текстов. В создании этого подкорпуса на момент написания статьи приняли участие 105 респондентов, которым давалось задание в первом тексте описать любой день из их жизни (например, день, предшествовавший написанию текста, либо же особый, памятный день) таким, как он был на самом деле, а во втором – сознательно исказив события, но таким образом, чтобы потенциальному читателю было сложно понять, какой из текстов правдивый, а какой «ложный». В настоящее время нами проводится пилотное исследование по выявлению статистически значимых различий между правдивыми и «ложными» текстами, полученными от одного и того же автора. Насколько нам известно, это первое исследование статистически значимых различий между ложными и правдивыми текстами на русском языке, тогда как за рубежом подобные исследования (в основном на материале английского языка) проводятся очень активно [20].

На материале корпуса текстов *Personality* был проведен ряд исследований, в ходе которых был составлен список формально-грамматических параметров текста, поддающихся извлечению средствами автоматической обработки языка, выявлены корреляции между некоторыми формально-грамматическими параметрами текста и психофизиологическими свойствами (полом и некоторыми индивидуально-психологическими характеристиками) автора текста и на их основе построены математические модели для диагностирования этих личностных характеристик на основе числовых значений параметров текста [17; 18; 21; 22].

Таким образом, нами был применен к русскому языку подход к решению проблемы моделирования личности по тексту, активно развивающийся в современной зарубежной науке и показавший свою эффективность, однако, на наш взгляд, подход к решению данной проблемы, предполагающий комплексный психолингвистический анализ текста на разных уровнях его порождения, в том числе анализ показателей связности и цельности текста, будет более эффективным и научно обоснованным.

Нами планируется также расширение корпуса текстов как путем привлечения респондентов, так и путем включения в него текстов интернет-коммуникации, для которых известна информация об авторах (как правило, пол, возраст) [4], однако, по нашему мнению, сбор такого рода материала невозможен без ручной обработки: в текстах могут содержаться цитаты, ссылки и другая информация, подлежащая удалению. Кроме того, в профилях авторов может быть указана, как говорилось ранее, заведомо ложная информация, в связи с чем к созданию корпусов такого рода следует подходить крайне взвешенно.

Таким образом, корпус русских письменных текстов *Personality* в настоящее время может быть использован для разработки и тестирования методик решения следующих актуальных задач судебного автороведения:

- 1) идентификации автора текста;
- 2) диагностирования характеристик автора текста;
- 3) установления наличия в тексте намеренно искаженной информации.

Корпусные исследования письменных текстов позволят проводить дальнейшие исследования вариативности единиц письменного текста и разрабатывать эффективные приемы их формально-языкового описания с целью выявления комплекса объективных параметров структурирования письменно порождаемого текста на русском языке, что является одной из важнейших задач судебного автороведения.

Список литературы

1. Араева Л. А., Осадчий М. А. Языковая личность экстремиста (о специфике автороведческой экспертизы по криминальным проявлениям экстремизма) // Юрислингвистика. 2008. № 9. С. 182-194.
2. Баранов А. Н. Лингвистическая экспертиза текста: теория и практика: учеб. пособие. М.: Флинта; Наука, 2007. 592 с.
3. Багура Т. В. Формальные методы определения авторства текстов // Вестник НГУ. Сер.: Информационные технологии. 2012. Т. 10. № 4. С. 81-94.
4. Беликов В. И., Копылов Н. Ю., Селегей В. П., Шаров С. А. Дифференциальная корпусная статистика на основе неавтоматической метатекстовой разметки // Компьютерная лингвистика и интеллектуальные технологии: материалы ежегодной междунар. конф. «Диалог». М.: Изд-во РГГУ, 2014. Вып. 13 (20). С. 54-67.
5. Беликов В. И., Селегей В. П., Шаров С. А. Прологомены к проекту Генерального интернет-корпуса русского языка (ГИКРЯ) // Компьютерная лингвистика и интеллектуальные технологии: материалы ежегодной междунар. конф. «Диалог». М.: Изд-во РГГУ, 2012. Вып. 11 (18): в 2-х т. Т. 1. С. 37-49.
6. Бугаев К. В. Тенденции судебной экспертизы в России // Вестник Сибирского института бизнеса и информационных технологий. 2013. № 4 (8). С. 29-32.

7. **Галяшина Е. И.** Лингвистическое обеспечение криминалистического исследования устного и письменного текста [Электронный ресурс] // Компьютерная лингвистика и интеллектуальные технологии: материалы ежегодной междунар. конф. «Диалог». URL: <http://www.dialog-21.ru/en/digest/archive/2001/?year=2001&vol=22725&id=6839> (дата обращения: 26.04.2015).
8. **Галяшина Е. И., Приводнова Е. В.** Автороведческая экспертиза в российском судопроизводстве // *Lex Russica*. 2006. Т. LXV. № 4. С. 755-761.
9. **Голев Н. Д.** От редактора. Актуальные проблемы юрислингвистической экспертизы // *Юрислингвистика-3. Проблемы юрислингвистической экспертизы: межвуз. сб. науч. тр. / под ред. Н. Д. Голев. Барнаул: Изд-во Алт. ун-та, 2002. С. 5-13.*
10. **Гомон Т. В.** Исследование документов с деформированной внутренней структурой: автореф. дисс. ... к. юрид. н. М., 1990. 22 с.
11. **Грасник Л. В.** Установление автора текстов интернет-коммуникаций при умышленном искажении признаков письменной речи // Актуальные проблемы российского права. 2011. № 2. С. 401-408.
12. **Губаева Т., Ситдикова Р.** Определяет ли автороведческая экспертиза автора? // *Интеллектуальная собственность. Авторское право и смежные права*. 2006. № 8. С. 23-31.
13. **Дроздова Т. Н.** Диагностические и классификационные задачи в автороведческой экспертизе блогов // Актуальные проблемы российского права. 2010. № 2 (15). С. 394-404.
14. **Загоровская О. В., Литвинова Т. А., Литвинова О. А.** Электронный корпус студенческих эссе на русском языке и его возможности для современных гуманитарных исследований // *Мир науки, культуры и образования*. 2012. № 3 (34). С. 387-389.
15. **Лебедева Н. Б.** Естественная письменная русская речь как объект лингвистического исследования // *Вестник БГПУ. Сер.: Гуманитарные науки*. 2001. № 1. С. 4-10.
16. **Литвинова Т. А.** Лингвистические основы неидентификационной судебно-автороведческой экспертизы // *Вестник Челябинского гос. университета. Серия: Филология. Искусствоведение*. 2012. Вып. 67. С. 74-79.
17. **Литвинова Т. А.** Русский письменный текст как носитель информации об индивидуально-личностных характеристиках его автора (на материале корпуса текстов нового типа Personality) // *Известия ВГПУ. Сер.: Педагогические науки; Гуманитарные науки*. 2015. Т. 266. № 1. С. 196-198.
18. **Литвинова Т. А.** Формально-грамматические корреляты личностных особенностей автора письменного текста // *Филологические науки. Вопросы теории и практики*. Тамбов: Грамота, 2013. № 12 (30): в 2-х ч. Ч. 1. С. 132-135.
19. **Литвинова Т. А.** Языковые корреляты личностных особенностей автора письменного текста: алгоритм исследования // *В мире научных открытий. Сер.: Проблемы науки и образования*. 2012. № 9.3 (33). С. 236-255.
20. **Литвинова Т. А., Литвинова О. А.** Исследование текста на предмет наличия в нем намеренно искаженной информации: проблемы и перспективы // *Известия ВГПУ. Сер.: Педагогические науки; Гуманитарные науки*. 2015. № 2 (267). С. 189-192.
21. **Литвинова Т. А., Литвинова О. А., Середин П. В.** Частоты встречаемости последовательностей частей речи в тексте и психофизиологические характеристики его автора: корпусное исследование // *Вестник Иркутского гос. лингвистического ун-та*. 2014. № 2. С. 8-12.
22. **Литвинова Т. А., Середин П. В., Литвинова О. А., Загоровская О. В., Сердюк М. Е.** Диагностирование индивидуально-психологических особенностей автора письменного текста, детерминирующих склонность к аутоагрессивному поведению // *Вестник Воронежского гос. ун-та. Сер.: Лингвистика и межкультурная коммуникация*. 2015. № 3.
23. **Морозов Н. А.** Лингвистические спектры. Новое орудие объективного исследования древних документов [Электронный ресурс]. URL: <http://www.textology.ru/library/book.aspx?bookId=1&textId=3> (дата обращения: 26.04.2015).
24. **Назарова Т. В., Ростовская А. В.** О методических разработках по производству автороведческих экспертиз рукописных текстов // *Речевые информационные технологии*. 2011. С. 387-392.
25. **Напреенко Г. В.** Идентификация текста по его авторской принадлежности на лексическом уровне (формально-количественная модель) // *Вестник Томского гос. университета*. 2014. № 379. С. 17-23.
26. **Резанова З. И., Романов А. С., Мещеряков Р. В.** О выборе признаков текста, релевантных в автороведческой экспертной деятельности // *Вестник Томского гос. университета. Филология*. 2013. Вып. № 6 (26). С. 38-52.
27. **Романов А. С.** Методика и программный комплекс для идентификации автора неизвестного текста: автореф. дисс. ... к. техн. н. Томск, 2010. 26 с.
28. **Рыжкова Е. В.** Проблемы автороведческой классификационной экспертизы текстов интернет-коммуникации // *Филологические науки. Вопросы теории и практики*. Тамбов: Грамота, 2015. № 6 (48). Ч. 1. С. 130-133.
29. **Что такое Корпус?** [Электронный ресурс]. URL: <http://www.ruscorpora.ru/corpora-intro.html> (дата обращения: 26.04.2015).
30. **Шахтарина О. И.** Криминалистическая диагностика в аспекте розыска автора анонимного текста // *Известия Тульского государственного университета. Сер.: Экономические и юридические науки*. 2012. № 2-2. С. 296-305.
31. **Afroz S.** Deception in Authorship Attribution [Электронный ресурс]: PhD thesis. URL: <https://www.eecs.berkeley.edu/~sa499/thesis.pdf> (дата обращения: 26.04.2015).
32. **Celli F., Pianesi F., Stillwell M., Kosinski D.** Workshop on Computational Personality Recognition (Shared Task) [Электронный ресурс] // *Proceedings of the Workshop on Computational Personality Recognition (Shared Task)*. Boston, 2013. URL: www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/viewFile/6190/6306 (дата обращения: 26.04.2015).
33. **PAN 2015 Author Profiling** [Электронный ресурс]. URL: <http://www.uni-weimar.de/medien/webis/events/pan-15/pan15-web/author-profiling.html> (дата обращения: 26.04.2015).
34. **Schler J., Koppel M., Argamon S., Pennebaker J.** Effects of Age and Gender on Blogging [Электронный ресурс] // *Proceedings of 2006 AAAI Spring Symposium on Computational Approaches for Analyzing Weblogs*. URL: <http://u.cs.biu.ac.il/~koppel/papers/springsymp-blogs-07.10.05-final.pdf> (дата обращения: 26.04.2015).
35. **Verhoeven, B., Daelemans W.** CLiPS Stylometry Investigation (CSI) Corpus: A Dutch Corpus for the Detection of Age, Gender, Personality, Sentiment and Deception in Text [Электронный ресурс] // *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*. Reykjavik, Iceland (May, 2014). URL: www.cnts.ua.ac.be/sites/default/files/paper_csi_lrec_1.0.0.bv_2014-03-21.final_.pdf (дата обращения: 26.04.2015).

CORPUS STUDY OF WRITTEN SPEECH IN SOLVING PROBLEMS OF JUDICIAL AUTHORSHIP

Litvinova Tat'yana Aleksandrovna, Ph. D. in Philology
Litvinova Ol'ga Aleksandrovna
Voronezh State Pedagogical University
centr_rus_yaz@mail.ru; olga_litvinova_teacher@mail.ru

Dibrova Elena Vladimirovna
Southern Federal University
alena_dibrova@bk.ru

Ryzhkova Ekaterina Sergeevna
Voronezh State University of Engineering Technologies
ryzhkova.katerina@yandex.ru

The article substantiates the necessity of using in a special way the corpus of texts in the development of techniques for the identificatory and diagnostic authorship examination. A brief analysis of the existing corpus of texts used to create the methods of identification and diagnosis of the text author's personality is presented. The paper describes the corpus of texts in the Russian language *Personality*, containing meta-marking in the form of information about their authors, and analyzes the possibility of its application for solving problems of the judicial authorship.

Key words and phrases: authorship; linguistic examination; authorship examination; corpus of texts; linguistic statistics; diagnosing of text author's personality; natural written speech; attribution of text.

УДК 811.133.1

Филологические науки

В статье рассматривается понятие фразеологизма во французской и русской лингвистике, приводятся особенности фамиллярной и просторечной фразеологии. Все теоретические утверждения подтверждены многочисленными иллюстративными примерами. Особенности перевода французских фразеологизмов выявлены посредством сопоставительного анализа оригинала и перевода романов А. Нотомб «Антихриста» и «Токийская невеста».

Ключевые слова и фразы: фразеологизм; разговорная и просторечная фразеология; библеизм; перевод; французский язык; русский язык.

Лиходкина Ирина Александровна, к. филол. н.
Военный университет Министерства обороны Российской Федерации
ilizka@rambler.ru

**ФРАЗЕОЛОГИЗМЫ КАК ЭЛЕМЕНТ РАЗГОВОРНОЙ РЕЧИ И ОСОБЕННОСТИ ИХ ПЕРЕВОДА
(ФРАНЦУЗСКО-РУССКИЕ ПАРАЛЛЕЛИ)[©]**

Во французской и русской разговорной речи разговорные выражения и обороты занимают важное место среди выразительных средств: воплощают в себе накопленную народную мудрость, в них находят отражение народный юмор, склонность к гиперболлизации и иронии, стремление к острому и эффектному выражению мысли. Они могут быть отнесены как к различного рода фразеологическим единицам, так и к свободным сочетаниям слов.

Фразеология (от греч. *phrasis*, род. п. *phraseos* – выражение и *logos* – слово, учение) – раздел языкознания, изучающий фразеологический состав языка в его современном состоянии и историческом развитии, а также совокупность фразеологизмов данного языка, то же, что фразеологический состав [4].

Фразеологизмы отличаются образностью, экспрессивностью, стилистической окрашенностью. У них всегда есть аналоги в нейтральной лексике, тогда как сами они принадлежат к разговорно-фамиллярному, просторечному или литературно-художественному стилям и удовлетворяют потребность языка придать речи тот или иной оттенок, выйти за строгие рамки обыденности.

Ученые давно интересовались вопросами фразеологии, фиксировали устойчивые словосочетания в словарях, однако самостоятельным разделом языкознания фразеология стала относительно недавно, а над определением ее объекта до сих пор ломают голову как отечественные, так зарубежные лингвисты. Во французском языке имеется много терминов, называющих объект фразеологии: *locution*, *locution figée*, *expression idiomatique*, *gallicisme*, *idiotisme* и др. Однако этими же терминами могут обозначаться свободные словосочетания и обороты, а иногда и различные грамматические конструкции. Русскому термину «фразеологическая единица» (ФЕ) больше всего соответствует выражение «locution phraséologique», предложенное швейцарским лингвистом Ш. Балли в начале XX в. В своем трактате «Traité de stylistique française», впервые опубликованном в 1909 г. и до сих пор переиздающимся, Ш. Балли выделил две группы фразеологизмов: