

Рыжкова Екатерина Сергеевна

**ПРОБЛЕМЫ АВТОРОВЕДЧЕСКОЙ КЛАССИФИКАЦИОННОЙ ЭКСПЕРТИЗЫ ТЕКСТОВ
ИНТЕРНЕТ-КОММУНИКАЦИИ**

В статье рассматривается вопрос о возможности проведения автороведческой классификационной экспертизы письменных текстов, в том числе текстов современной русскоязычной интернет-коммуникации. С учетом имеющегося в зарубежной и отечественной науке исследовательского опыта анализируются корреляции между формально-грамматическими параметрами речевых произведений и личностными характеристиками их авторов - современных школьников - на материале специально создаваемого корпуса текстов.

Адрес статьи: www.gramota.net/materials/2/2015/6-1/35.html

Источник

Филологические науки. Вопросы теории и практики

Тамбов: Грамота, 2015. № 6 (48): в 2-х ч. Ч. I. С. 130-133. ISSN 1997-2911.

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

Содержание данного номера журнала: www.gramota.net/materials/2/2015/6-1/

© 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

УДК 81'33

Филологические науки

В статье рассматривается вопрос о возможности проведения автороведческой классификационной экспертизы письменных текстов, в том числе текстов современной русскоязычной интернет-коммуникации. С учетом имеющегося в зарубежной и отечественной науке исследовательского опыта анализируются корреляции между формально-грамматическими параметрами речевых произведений и личностными характеристиками их авторов – современных школьников – на материале специально создаваемого корпуса текстов.

Ключевые слова и фразы: автороведческая экспертиза текста; классификационная автороведческая экспертиза текста; естественная письменная речь; корпус текстов; Интернет; интернет-коммуникация.

Рыжкова Екатерина Сергеевна*Воронежский государственный педагогический университет**ryzhkova.katerina@yandex.ru***ПРОБЛЕМЫ АВТОРОВЕДЧЕСКОЙ КЛАССИФИКАЦИОННОЙ
ЭКСПЕРТИЗЫ ТЕКСТОВ ИНТЕРНЕТ-КОММУНИКАЦИИ[©]**

Известно, что речь является зеркалом, в котором отражается личность ее автора. Человек говорящий «реализует себя прежде всего в создании речевых произведений, текстовой деятельности» [13, с. 15].

Создаваемые носителем языка тексты представляют собой его речевой портрет, источник информации об его индивидуально-личностных особенностях. Выявление этой информации о человеке (половая, возрастная, социальная принадлежность, психическое и физиологическое состояние и другие параметры) на основе анализа созданных им речевых произведений имеет не только теоретическое, но и практическое значение, в частности, при разработке методик судебно-автороведческой экспертизы документов. При этом получение такой информации, разработка методик для проведения экспертизы текста с воссозданием «облика» его автора, по мнению современных исследователей, возможны лишь при совместных усилиях ученых-криминалистов, психологов и лингвистов [8].

Становление судебно-автороведческой экспертизы текстов на русском языке началось в 60-70 гг. XX века с работ С. М. Вула, в которых утверждалась, прежде всего, необходимость учета системной организации текста документа при установлении информации о личности его автора [1, с. 165]. При этом под текстом документа названный ученый понимал «совокупность предложений, связанных единством смысла, которая в частности служит материалом для криминалистического исследования состояния автора» [Там же, с. 162]. В качестве текста, по мнению С. М. Вула, может рассматриваться отдельное предложение или даже словосочетание, «если они удовлетворяют указанному условию» [Там же].

С момента становления судебно-автороведческой экспертизы письменных документов и до настоящего времени в рамках рассматриваемого направления изучения текстов сделано много работ, накоплен значительный объем материала.

Как отмечается в специальной научной литературе, судебно-автороведческая экспертиза является в настоящее время одним из интереснейших и актуальных направлений исследования текстов и представляет собой вид криминалистической экспертизы, призванный решать идентификационные и неидентификационные задачи [5]. Идентификационные задачи состоят в установлении автора текста из ряда предполагаемых лиц, причастных к его созданию, или отрицание авторства лица, подозреваемого в составлении какого-либо текста. Неидентификационные задачи судебно-автороведческой экспертизы подразделяются на классификационные и диагностические. К числу классификационных задач относятся установление определенных личностных характеристик создателя текста: возраста, профессии, родного языка, места формирования языковых навыков, половой принадлежности, – а также отношений между участниками диалога или монолога. Судебно-автороведческие диагностические задачи – это задачи, направленные на выявление состояния автора в момент создания текста и определение некоторых сопутствующих условиям создания речевого произведения особенностей [Там же].

В рамках нашего исследования основное внимание будет уделено автороведческой классификационной экспертизе *письменных* текстов. Цель данной работы состоит главным образом в обзоре имеющегося на сегодняшний день в отечественной и зарубежной науке опыта исследования взаимосвязи между параметрами письменных высказываний и индивидуально-личностными характеристиками их создателей.

С момента возникновения и до недавнего времени вопросы автороведческой экспертизы письменных речевых произведений решались преимущественно на материале рукописных и машинописных текстов. На сегодняшний день работа с такими источниками становится менее востребованной и все большую актуальность приобретает обращение к речевым произведениям интернет-коммуникации [2; 3]. Современные исследователи отмечают, что значимость вербального общения в настоящее время увеличивается, прежде всего, в силу растущей популярности Интернета, который может быть назван новой социальной реальностью [4]. Ср., например: «Роль языка как одного из базовых “инструментов” формирования идентичности в Интернете возрастает вдвойне, <...> существует однозначная связь между коммуникацией, идентификационными и речевыми процессами» [Там же].

Понятие интернет-коммуникации относится к числу дискуссионных в современной научной литературе. Сказанное в полной мере касается неустоявшейся терминологии при обозначении данного феномена и многообразия подходов к его рассмотрению [Там же]. Мы придерживаемся точки зрения Е. А. Земляковой, разделяемой Е. И. Горошко, согласно которой необходимо различать понятия «электронная», «компьютерная» и «интернет-коммуникация», связанные между собой гиперо-гипонимическими отношениями [Там же]. По мнению Е. А. Земляковой, термин «электронная коммуникация», определяемый исследователем как коммуникативное взаимодействие, опосредованное электронным каналом, является наиболее широким, родовым, понятием. Данное понятие включает в себя более узкое понятие «компьютерная коммуникация», под которым понимается коммуникативное взаимодействие, осуществляемое с помощью компьютера. В предметную область названного понятия входит и понятие «интернет-коммуникация» – вербальное и невербальное взаимодействие коммуникантов в сети Интернет [Там же].

В настоящее время проблема классификационной автороведческой экспертизы текстов, в том числе текстов интернет-коммуникации, особенно активно разрабатывается в зарубежной науке [14-17].

Главной целью подобных исследований является создание методик, позволяющих диагностировать возраст и пол автора текста интернет-коммуникации. Анализ речевых произведений коммуникантов при этом заключается, как правило, в установлении корреляций между формально-грамматическими (*style-based*) и лексическими (в зарубежной терминологии – *content-based*) параметрами текстовых сообщений и индивидуально-личностными характеристиками (возрастом и полом) их авторов с помощью использования автоматизированных методов анализа текста и методов математической статистики. К числу формально-грамматических параметров текстовых сообщений в трудах зарубежных ученых относятся, в первую очередь, такие, как число в тексте функциональных слов (предлогов, союзов, артиклей, местоимений и модальных глаголов), средняя длина предложений и подобные, причем такие параметры оказываются более информативными, чем лексические (например, доля в тексте слов определенных тематических групп). Объем используемых исследователями текстовых корпусов насчитывает миллионы слов; в выборке представлены тексты сообщений коммуникантов разного возраста: от 13 до 40 лет и старше. Полученные по результатам подобных работ математические модели имеют достаточно высокий процент точности и позволяют ученым говорить о существовании зависимости использования языка от возраста и пола человека – создателя текста [Ibidem].

Вопросы классификационной автороведческой экспертизы текстов на русском языке наиболее детально разработаны в трудах криминалистов, которые использовали тексты «реальной» устной и письменной коммуникации (работы С. М. Вула, А. И. Бастрыкина, Е. И. Галяшиной, Е. И. Горошко и др.). На материале интернет-коммуникации названные вопросы остаются практически не изученными. Можно назвать лишь единичные исследования, посвященные данной проблеме, и, прежде всего, работу А. С. Романова и Р. В. Мещерякова «Определение пола автора короткого электронного сообщения» [12] и работу Е. И. Горошко «Изучение электронного дискурса в судебном речеведении», представляющую собой пилотное исследование гендерных особенностей электронного письма [3].

Актуальность и значимость автороведческой классификационной экспертизы текстов интернет-коммуникации, позволяющей получить информацию о коммуникантах Интернет-пространства (в первую очередь, об их возрастной и половой принадлежности), зарубежные и российские исследователи видят в возможности ее применения в судебной практике, в сфере маркетинга и коммерции, при подборе персонала в рекрутменте, в целях разработки эффективных механизмов таргетирования Интернет-пользователей, а также для определения истинной идентичности пользователей Сети [10; 14; 15; 17]. Последнее особенно важно в свете существующей и весьма распространенной в информационно-коммуникативной среде проблемы педофилии. Известно, что пользуясь широкими возможностями, предоставляемыми современной интернет-коммуникацией, преступники-педофилы скрывают свою истинную идентичность и создают ложный профиль (т.е. указывают ложные имя, возраст, пол и местонахождение) для того, чтобы преследовать своих жертв [15]. Кроме того, важность получения информации о личностных характеристиках авторов текстовых сообщений интернет-коммуникации вполне очевидна в связи увеличением опасности разжигания национальной, расовой или религиозной вражды, возникновением политических, социальных и других конфликтов, а также в связи с возрастанием террористической угрозы в современном мире [2; 10; 12].

В свете сказанного выше представляется очевидным, что исследование соотносительности пола и возраста русскоязычных пользователей интернет-коммуникации с формально-грамматическими показателями создаваемых ими текстов на русском языке имеет несомненную научно-теоретическую и прикладную значимость.

Следует отметить, что в немногочисленных работах российских исследователей, посвященных анализу параметров текстовых сообщений интернет-коммуникации с целью диагностирования личности их авторов, упор делается на проблему определения пола создателя речевого произведения, в то время как вопрос об установлении возрастных характеристик авторов текстов остается за рамками исследовательской деятельности. Кроме того, существенным недостатком подобных исследований является то, что они опираются на данные о личностных характеристиках, которые указывают сами авторы текстов (подобный подход характерен и для англоязычных научных работ, посвященных рассматриваемой проблеме). Между тем очевидно, что личностные характеристики, в том числе пол и возраст, которые размещают в Интернете сами авторы, могут быть ложными, и случаи искажения авторской идентичности в интернет-коммуникации достаточно частотны. В связи с этим представляется, что большей диагностической ценностью будут обладать методики, построенные на корпусах таких текстов, данные об авторах которых (пол, возраст, психологический профиль) являются абсолютно достоверными.

Как представляется, для корректного прогнозирования пола и возраста авторов речевых произведений интернет-коммуникации требуется создание корпуса текстов, содержащего образцы *естественной письменной речи*. Под *естественной письменной речью* мы вслед за Н. Б. Лебедевой понимаем «речевую деятельность (и ее результат – тексты), которая занимает свое место в парадигме, построенной по координатам «устная / письменная» и «естественная / искусственная» (*искусная*) речь», и обладает такими признаками, как письменная форма, спонтанность и непрофессиональность исполнения, что, безусловно, характерно и для большинства текстов интернет-коммуникации [7]. Особенно актуальным, на наш взгляд, в настоящее время является создание корпуса текстов естественной письменной речи молодых носителей русского языка, которые принадлежат к числу самых активных пользователей Интернета.

Для выявления взаимосвязи между параметрами текста и характеристиками личности его автора наиболее правомерным представляется использование подхода, предложенного Т. А. Литвиновой [9] и включающего в себя следующие этапы:

1) сбор исследовательского корпуса текстов – образцов естественной письменной речи и его перевод в электронный вид;

2) составление, на основе предпроектных исследований и изучения научной литературы, списка параметров текста, которые могут коррелировать с теми или иными параметрами личности;

3) использование математических методов для обработки собранного материала с целью выявления корреляций между численными значениями параметров письменных текстов и характеристиками личности их авторов (пол, возраст, результаты психологического тестирования и пр.).

Названный подход к составлению исследовательского корпуса текстов предполагает включение в данный корпус, помимо собственно текстов (образцов естественной письменной речи), также еще и метаданных в виде социобиографической информации об авторах речевых произведений (пол, возраст, результаты психологического тестирования и др.).

На протяжении нескольких последних лет сотрудниками Регионального центра русского языка при Воронежском государственном педагогическом университете ведется работа по формированию корпуса текстов – образцов естественной письменной речи, содержащего метаданные в виде информации об их авторах [6]. Исследователями уже установлен ряд значимых корреляций между формально-грамматическими параметрами текстов и характеристиками их авторов [9; 11]. В настоящее время ставится задача дополнения названного корпуса новым подкорпусом, содержащим образцы естественной письменной речи современных российских школьников среднего и старшего возраста. С учетом существующего опыта зарубежных и российских ученых на материале нового подкорпуса планируются углубленное изучение преимущественно формально-грамматических (морфологических и частично синтаксических) параметров письменного текста, коррелирующих с личностными характеристиками автора, и проверка достоверности полученных в рамках исследований воронежской научной школы корреляций между названными текстовыми параметрами и индивидуальными характеристиками продуцента текста. Отказ от специального анализа лексико-семантических параметров текста объясняется тем, что подобный анализ на современном этапе развития науки, во-первых, не может быть полностью автоматизирован [8; 9], во-вторых, – не может быть свободен от ошибок, обусловленных возможностью имитации некоторых содержательных признаков речевых произведений [10].

Представляется, что исследования на материале специально созданных корпусов текстов – образцов естественной русской письменной речи, направленных на выявление корреляций между неподдающимися контролю сознания автора параметрами текста и личностными характеристиками продуцента речи, могут внести существенный вклад в разработку методики диагностирования личности по ее речевой продукции, в том числе по текстам, функционирующим в современной русскоязычной интернет-коммуникации. Необходимость создания специальных корпусов текстов для исследований подобного рода не вызывает сомнений. В настоящее время такие работы уже ведутся не только в зарубежной, но и в российской науке и имеют конкретные результаты. В частности, современными исследователями были установлены взаимосвязи между полом автора письменного речевого произведения и количеством употребленных в нем местоимений и служебных слов [9, с. 134]. Мы надеемся получить подобные результаты (соотнесенности формально-грамматических показателей текстов с полом, возрастом и психологическим профилем их создателей) на нашем материале, сбор которого уже проводится.

Список литературы

1. Вул С. М. Текст документа как источник информации о личности его автора // Криминалистика и судебная экспертиза. Республиканский межведомственный сборник научных и научно-методических работ. Киев: РИО МВД УССР, 1971. Вып. 8. С. 162-168.
2. Галяшина Е. И. Судебная автороведческая экспертиза [Электронный ресурс]. URL: <http://isemgua.ru/info/avtor/> (дата обращения: 23.01.2015).
3. Горошко Е. И. Изучение электронного дискурса в судебном речеведении [Электронный ресурс]. URL: <http://www.textology.ru/article.aspx?aId=44> (дата обращения: 23.01.2015).
4. Горошко Е. И. Современные интернет-коммуникации: структура и основные характеристики [Электронный ресурс]. URL: <http://www.textology.ru/article.aspx?aId=232> (дата обращения: 15.02.2015).
5. Горошко Е. И. Судебно-автороведческая классификационная экспертиза [Электронный ресурс]. URL: <http://www.textology.ru/article.aspx?aId=98> (дата обращения: 27.02.2015).
6. 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, образования. Горно-Алтайск, 2012. № 3. С. 387-389.

7. Лебедева Н. Б. Естественная письменная русская речь как объект лингвистического исследования [Электронный ресурс]. URL: <http://www.uni-altai.ru/Journal/vestbspu/2001/gumanit/PDF/lebedeva.pdf> (дата обращения: 27.02.2015).
8. Литвинова Т. А. Лингвистические основы неидентификационной судебно-автороведческой экспертизы // Вестник Челябинского государственного университета. Сер. Филология. Искусствоведение. 2012. Вып. 67. № 20 (274). С. 74-78.
9. 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.
10. Литвинова Т. А., Лантюхова Н. Н., Рыжкова Е. С., Шевченко И. С. Профилирование автора текста как одно из стратегических направлений исследований // Вестник Воронежского института ГПС МЧС России. Воронеж, 2013. Вып. 1 (6). С. 38-41.
11. 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 государственного лингвистического университета. 2014. № 2. С. 8-12.
12. Романов А. С., Мещеряков Р. В. Определение пола автора короткого электронного сообщения [Электронный ресурс] // Компьютерная лингвистика и интеллектуальные технологии: сб. Междунар. конф. по компьютерной лингвистике «Диалог». 2011. URL: <http://www.dialog-21.ru/digests/dialog2011/materials/ru/pdf/55.pdf> (дата обращения: 27.02.2015).
13. Седов К. Ф. Становление структуры устного дискурса как выражение эволюции языковой личности: дисс. ... д. филол. наук. Саратов, 1999.
14. Argamon S., Koppel M., Pennebaker J. W., Schler J. Mining the Blogosphere: Age, Gender and the Varieties of Self-Expression [Электронный ресурс]. URL: <http://firstmonday.org/ojs/index.php/fm/article/view/2003/1878> (дата обращения: 25.11.2014).
15. Peersman C., Daelemans W., Vaerenbergh L. Van. Predicting Age and Gender in Online Social Networks [Электронный ресурс]. URL: <http://www.cnts.ua.ac.be/~walter/papers/2011/pdv11.pdf> (дата обращения: 24.11.2014).
16. Rangel F., Rosso P. Use of Language and Author Profiling: Identification of Gender and Age [Электронный ресурс]. URL: <http://www.kicorangel.com/wp-content/uploads/2013/10/Use-of-Language-and-Author-Profiling-Identification-of-Gender-and-Age.pdf> (дата обращения: 27.11.2014).
17. Schler J., Koppel M., Argamon S., Pennebaker J. Effects of Age and Gender on Blogging [Электронный ресурс]. URL: <http://www.aaai.org/Papers/Symposia/Spring/2006/SS-06-03/SS06-03-039.pdf> (дата обращения: 24.11.2014).

PROBLEMS OF AUTHORSHIP CLASSIFICATION EXAMINATION OF TEXTS IN INTERNET COMMUNICATION

Ryzhkova Ekaterina Sergeevna
Voronezh State Pedagogical University
ryzhkova.katerina@yandex.ru

The article considers the possibility of conducting the authorship classification examination of written texts, including the texts of the modern Russian-speaking Internet communication. In view of the existing in foreign and domestic science research experience, the author of the article analyzes the correlation between the formal grammatical parameters of speech products and personal characteristics of their authors – modern schoolchildren – by the material of the specially created corpus of texts.

Key words and phrases: authorship examination of text; classification authorship examination of text; natural written speech; corpus of texts; Internet; Internet communication.

УДК 821.111

Филологические науки

В жизни и творчестве английского писателя Лоренса Даррелла (1912-1990) Греция всегда занимала особое место. Книги Даррелла о греческих путешествиях позволили читателям узнать его с новой стороны: не только как неординарного романиста, но и как наблюдательного и остроумного путешественника, настоящего «эллинофила». Интерес к Греции и греческому античному наследию в разные исторические эпохи был неодинаков, в фокусе оказывались разные проблемы и стороны греческой культуры и жизни. К середине XX века в Англии формируется новый тип эллинофильства, обращающий особое внимание на современную Грецию и греческую культуру. В своих травелогах Даррелл стремится постичь феномен Греции, её культуры и природы.

Ключевые слова и фразы: путевая проза; травелог; эллинофильство; эллинизм; греческая культура; греческое античное наследие; философия «островомании», «эллинофильства» и «духа места».

Савельева Инна Геннадиевна, к. филол. н.

Московский государственный областной социально-гуманитарный институт
savelieva.inna@gmail.com

ОТРАЖЕНИЕ ЭЛЛИНИСТИЧЕСКОЙ ТРАДИЦИИ И ИДЕЙ ЭЛЛИНОФИЛЬСТВА В ТВОРЧЕСТВЕ ЛОРЕНСА ДАРРЕЛЛА[©]

Путевая проза Лоренса Даррелла (1912-1990) практически целиком посвящена Греции и греческим островам, за исключением книг «Сицилийская Карусель» (1977) и «Прованс» (1990).

[©] Савельева И. Г., 2015