

Литвинова Татьяна Александровна

ФОРМАЛЬНО-ГРАММАТИЧЕСКИЕ КОРРЕЛЯТЫ ЛИЧНОСТНЫХ ОСОБЕННОСТЕЙ АВТОРА ПИСЬМЕННОГО ТЕКСТА

Статья представляет собой пилотное исследование зависимостей между формально-грамматическими параметрами текста и личностными особенностями его автора, т.е. раскрывает один из аспектов проблемы моделирования личности автора письменного текста. С применением методов математической статистики получены регрессионные модели, связывающие формально-грамматические характеристики текста и личностные особенности его автора. Материалом исследования послужил специально созданный корпус текстов, снабженный метаразметкой в виде информации о его авторах.

Адрес статьи: www.gramota.net/materials/2/2013/12-1/37.html

Источник

Филологические науки. Вопросы теории и практики

Тамбов: Грамота, 2013. № 12 (30): в 2-х ч. Ч. I. С. 132-135. ISSN 1997-2911.

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

Содержание данного номера журнала: www.gramota.net/materials/2/2013/12-1/

© 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: voprosy_phil@gramota.net

УДК 81'33

Филологические науки

Статья представляет собой пилотное исследование зависимостей между формально-грамматическими параметрами текста и личностными особенностями его автора, т.е. раскрывает один из аспектов проблемы моделирования личности автора письменного текста. С применением методов математической статистики получены регрессионные модели, связывающие формально-грамматические характеристики текста и личностные особенности его автора. Материалом исследования послужил специально созданный корпус текстов, снабженный метаразметкой в виде информации о его авторах.

Ключевые слова и фразы: текст; автороведение; лингвостатистика; математическая лингвистика; служебные слова; корреляционный анализ.

Литвинова Татьяна Александровна, к. филол. н.
Воронежский государственный педагогический университет
centr_rus_yaz@mail.ru

**ФОРМАЛЬНО-ГРАММАТИЧЕСКИЕ КОРРЕЛЯТЫ ЛИЧНОСТНЫХ
ОСОБЕННОСТЕЙ АВТОРА ПИСЬМЕННОГО ТЕКСТА**[©]

Исследование выполнено при поддержке гранта РФФИ № 13-06-00016 «Моделирование личности автора письменного текста», гранта РГНФ № 13-14-36001 «Речевой портрет воронежских студентов (на материале электронного корпуса текстов –Россия и мир глазами воронежских студентов)».

Введение. В настоящее время считается доказанным положение о том, что текст как продукт индивидуальной речевой деятельности несет информацию о личности его автора [5]. Однако в науке не выработано единого подхода к методике выявления информации о характеристиках автора текста на основе лингвистического анализа его параметров.

1. Постановка задачи. В последнее время в связи с бурным развитием средств автоматической обработки языка (морфологических и синтаксических парсеров), программ для статистической обработки данных представляется особенно перспективным стилеметрический подход к моделированию личности по тексту, основная идея которого состоит в том, что на большом корпусном материале с использованием методов статистической обработки данных вычисляются корреляции между подлежащими количественной оценке параметрами текстов и характеристиками их авторов [2; 6]. Следовательно, для решения этой задачи в свете указанного подхода необходимы три компонента:

– *корпус текстов*, специально созданный для решения данной задачи и содержащий метаразметку в виде социобиографической информации об их авторах (пол, возраст, образование, профессия, данные психотестов и пр.). Отметим, что создание таких корпусов само по себе является нетривиальной задачей, и общедоступных корпусов такого типа на русском языке, насколько нам известно, не существует [1];

– *перечень параметров текста*, которые могут быть информативными для диагностирования той или иной характеристики его продуцента. Как показывают современные исследования, личность автора проявляется на всех уровнях текста, однако количественный анализ текста на уровне семантики и лексики весьма трудоемок и на современном этапе развития науки не может быть полностью автоматизирован, в связи с чем в новейших исследованиях основное внимание уделяется параметрам текста на уровне морфологии и частично синтаксиса [3];

– *математические методы* выявления корреляций численных значений параметров текстов и характеристик личности их авторов. Для решения данной задачи используются методы статистической обработки данных, методы машинного обучения и пр. [4].

В настоящей работе мы ставим задачу описать результаты проведенного авторским коллективом эксперимента по выявлению корреляций между формально-грамматическими, поддающимися квантификации параметрами текста и свойствами личности его автора на материале специального созданного корпуса текстов с применением статистических методов обработки данных.

2. Методика эксперимента. Рассмотрим более подробно используемый в исследовании корпус текстов, набор формализуемых параметров текста, математические методы.

2.1. Корпус. Методика построения корпуса текстов для исследований по проблеме моделирования личности автора текста подробно изложена в работе О. В. Загоровской, Т. А. Литвиновой, О. А. Литвиновой [1]. В настоящее время наш корпус насчитывает 1025 текстов от 586 респондентов (респондентов просили написать два текста на заданную тему, но некоторые написали один), информацию о поле авторов и данные их психологического тестирования. Респондентами выступали студенты 1-5 курсов воронежских и московских вузов (как технических, так и гуманитарных специальностей). Предложенные темы текстов: «Письмо другу»; «Описание картинки»; «Что бы я сделал с миллионом долларов США?»; «Убедите работодателя, что именно Вас он должен взять на работу» и др. Для пилотного исследования мы отобрали 150 текстов от 75-ти респондентов (26-ти мужчин, 49-ти женщин), среднее число слов в текстах – 166. Также в анкетах респонденты

указывали свой пол, специальность (техническая/гуманитарная) и заполняли опросники двух психологических тестов – пятифакторного личностного опросника МакКрае – Коста (традиционно применяется для исследований по моделированию личности автора в англоязычных исследованиях), с помощью которого можно измерять степень выраженности каждого из пяти факторов (экстраверсия – интроверсия; привязанность – обособленность; самоконтроль – импульсивность; эмоциональная неустойчивость – эмоциональная устойчивость; экспрессивность – практичность), и теста «Методика диагностики коммуникативной установки В. В. Бойко».

2.2. *Параметры текста.* Методами автоматической обработки текстов (АОТ) (с использованием морфологических анализаторов, программ для подсчета частоты встречаемости языковых элементов) были извлечены числовые значения формально-грамматических параметров текста, список которых был составлен по материалам русскоязычной и англоязычной научной литературы, а также предпроектных исследований автора, всего 75 параметров текста.

Гипотеза исследования состоит в том, что для моделирования личности по тексту на основе его формализуемых параметров наиболее релевантными окажутся т.н. функциональные слова (*function words*) – местоимения, предлоги, союзы, частицы, вспомогательные глаголы, дейктические наречия. Считается, что функциональный класс слов характеризуется рядом признаков, а именно: функциональные слова не обладают номинативной функцией, морфологически неразложимы, семантически и синтаксически несамостоятельны. При этом функциональные слова, будучи морфологически неразложимыми и синтаксически зависимыми единицами, могут иметь ослабленное лексическое значение [7, р. 66]. Функциональные слова менее контролируемы сознанием, им уделяется меньше внимания в речи, чем смысловым словам, и они обрабатываются мозгом по-иному, чем полнзначные слова, о чем свидетельствуют исследования афазий.

Чтобы уйти от зависимости от длины текста, в качестве параметров брали соотношения, т.е. относительные частоты.

2.3. *Математическая обработка данных.* Для измерения тесноты и направления связи между параметрами текста и личности, установления аналитического выражения (формы) связи мы применили корреляционно-регрессионный анализ с использованием современных программных комплексов обработки и визуализации статистических измерений.

Главной целью наших исследований стала оценка функциональной зависимости условного среднего значения результирующего признака (Y) (пола, данных психологического тестирования авторов текстов) от факторных признаков (x_1, x_2, \dots, x_n) – параметров текста. Таким образом, искомое уравнение регрессии, или статистическая модель связи характеристик автора и количественных значений параметров текста, выражается функцией

$$Y_x = f(x_1, x_2, \dots, x_n),$$

где n – число факторов, включенных в модель; x_i – факторы, влияющие на результат Y.

Корреляционно-регрессионный анализ для профилирования автора письменного текста мы проводили в несколько этапов. На первом этапе мы сформулировали решаемую задачу исследования и определились с методикой измерения показателей и сбора информации, а также посредством корреляционного анализа с использованием критерия Пирсона нам удалось определить число факторов, связанных в жестко детерминированную систему «параметры текста – параметры автора», а также оценить достоверность всех характеристик корреляционной связи, при тесноте связи $r = 0,05$. Далее мы предположили, что форма связи (тип аналитической функции) между выбранными параметрами текста автора и его личностной характеристикой будет линейной. На третьем этапе с использованием методов регрессионного анализа программного пакета SPSS мы нашли искомые уравнения регрессии и провели анализ полученных параметров уравнения с целью определения ошибки найденных закономерностей на проверочной группе.

Прогнозирование развития анализируемой системы по уравнению регрессии показало достаточно высокий уровень диагностирования параметров личности автора текста на основе неподконтрольных автору, но поддающихся квантификации параметров текста и наших результатов корреляционно-регрессионного анализа.

Для логической регрессии (пол) принимали: 1 = муж., 0 = жен. Что касается обычной мультилинейной регрессии (результаты психотестов), то число, которое считает уравнение, лежит в пределах от 0 до 100, как и баллы по тесту. Отклонение, или ошибку считали усредненную по 75-ти показателям, предварительно определив среднее отклонение регрессии от конкретного результата и далее усреднив значение.

3. Результаты эксперимента. Ниже представлены полученные регрессии, описывающие зависимость между численными значениями параметров текстов и характеристиками личности.

3.1. Пол

Регрессия = - 0,231 - (0,0395x(1)) + (2,681x (13)) + (0,204 x (14)) - (1,301x (20)) - (0,658 x (21)) + (0,466 x (25)) - (2,214 x (48)) + (1,173 x (55)) - (1,832 x (59)).

Вероятность совпадения пола по результатам проверки модели ~ **60%**.

Таким образом, для определения пола автора текста релевантными оказываются следующие параметры-отношения: **1)** количество знаменательных слов / количество незнаменательных слов; **13)** кол-во имен сущ. / всего слов; **14)** число знаменат. словоупотр. / число сущ.; **20)** (указ. мест. + относит.-вопросит. мест. + личн. мест. + местоименные наречия) / всего слов; **21)** (мест. всех разрядов + предлоги + мест. наречия) / всего слов; **25)** бессоюзные сложные предложения / сложных предложений всего; **48)** (местоим. + союзы + частицы) / общ. число слов; **55)** число деепричастий / общее число слов; **59)** личные местоим. / число слов.

3.2. Баллы по тесту «Методика диагностики коммуникативной установки В. В. Бойко»

Регрессия = 65,263 - (13,116 * (26)) - (18,872 * (66)) + (86,626 * (67)).

Отклонение от реального результата ~ **10%**.

Таким образом, для признака «Число баллов по тесту» релевантными оказываются параметры: **26)** сложноподчиненные предложения / сложные предложения всего; **66)** имена собственные / всего слов; **67)** имена собственные / (всего сущ. + личн. мест.).

3.3. Экстраверсия – интроверсия

Регрессия = $63,740 - (0,107 * (6)) - (39,485 * (23)) - (1,499 * (24)) + (10,665 * (29)) - (120,792 * (42)) - (3,899 * (45)) - (623,818 * (54))$.

Отклонение от реального результата по данным проверки модели ~ **13-14%**.

Таким образом, для этого параметра личными релевантными оказываются параметры: **6)** кол-во простых предл. / предл. всего; **23)** общее количество причастий и деепричастий / всего слов; **24)** кол-во союзов / кол-во предл.; **29)** число указ. + вопросит.-относит. мест. / всего слов; **42)** предл. / число слов; **45)** число дееприч. оборотов + число прич. оборотов / число обособлений; **55)** дееприч. / число слов.

3.4. Привязанность – обособленность

Регрессия = $80,427 - (1,227 * (3)) - (15,140 * (9)) - (12,020 * (12)) - (1,452 * (24)) - (51,413 * (42)) - (766,367 * (54))$.

Отклонение от реального результата ~ **15%**.

Таким образом, релевантными оказываются параметры: **3)** кол-во слов / кол-во простых предложений; **9)** кол-во предл. / кол-во незнаменат. словоупотреблений; **12)** (мест. всего + предлоги + мест. наречия) / (союзы + мест. + междометия + предлоги + мест. наречия + частица + вспомогат. глаголы); **24)** кол-во союзов / кол-во предл.; **42)** предлоги / число слов; **54)** дееприч. / всего слов.

3.5. Самоконтроль – импульсивность

Регрессия = $55,472 - (2,689 * (19)) - (55,871 * (23)) + (8,077 * (29)) - (546,071 * (54))$.

Отклонение от реального результата ~ **18%**.

Таким образом, релевантными оказываются параметры: **19)** прилаг. / (наречий + мест. наречий); **23)** общее количество причастий и деепричастий / всего слов; **29)** число указ. + вопросит.-относит. мест. / всего слов; **54)** количество дееприч. / всего слов.

3.6. Эмоциональная устойчивость – неустойчивость

Регрессия = $55,201 - (2,697 * (19)) - (521,891 * (54))$.

Отклонение от реального результата ~ **18%**.

Таким образом, релевантными оказываются параметры: **19)** прилаг. / (наречий + мест. наречий); **54)** количество дееприч. / всего слов.

3.7. Экспрессивность – практичность

Регрессия = $60,238 + (14,697 * (10)) + (227,831 * (11)) - (76,134 * (23)) + (5,893 * (29)) - (2,515 * (30)) - (576,137 * (46)) - (580,465 * (54))$.

Отклонение от реального результата ~ **9%**.

Таким образом, релевантными оказываются параметры: **10)** кол-во союзов / кол-во незнаменат. словоупотреблений; **11)** всего частиц / (союзы + мест. + межд. + предлоги + мест. наречия + частицы + вспомогат. глаголы); **23)** общее количество причастий и деепричастий / всего слов; **29)** (указ. мест. + вопросит.-относит. мест.) / всего слов; **30)** кол-во имен существительных / всего местоимений; **46)** отношение частиц / всего слов; **54)** дееприч. / число слов.

4. Обсуждение результатов. Таким образом, мы видим, что в целом наш подход к моделированию текстов подтвердил свою эффективность. Получены модели, дающие вполне точные результаты, за исключением логистической регрессии (диагностирование пола автора), что, возможно, связано с тем, что выборка по полу была недостаточно сбалансирована. Наша гипотеза о значимости служебных слов и местоимений для моделирования личности по полу также подтвердилась: для определения большинства параметров личности полезными оказываются те или соотношения этих частей речи.

Значимым является также анализ синтаксического уровня текстов, в частности, структуры предложений, однако в настоящее время он недостаточно поддается автоматизации, в связи с чем было взято ограниченное число параметров текстов на синтаксическом уровне: количество простых предложений; количество сложных предложений; количество простых предложений в составе сложных; количество сложных предложений по типам связи между частями (бессоюзное; сложносочиненное, сложноподчиненное). Все эти параметры также оказываются значимыми для моделирования личности по тексту.

Отметим, что мы намеренно не проводили контент-анализ указанных текстов, так как нашей целью был поиск формально-грамматических параметров текстов, коррелирующих с характеристиками личности. Наше исследование показало, что такая корреляция действительно существует, и исследования в этом направлении должны быть продолжены; в частности, необходимо выработать единую теоретическую концепцию с привлечением данных не только лингвистики, но и психологии, объясняющую эффективность тех или иных формально-грамматических параметров для диагностирования по тексту того или иного свойства личности. Построение такой концепции неизбежно приведет к выявлению новых релевантных параметров текстов и, следовательно, к повышению эффективности построенных моделей.

Выводы. Отметим, что данное исследование является пилотным и лишь намечает направления поиска в области моделирования личности по тексту на основе формализуемых, неподконтрольных сознанию пишущего параметров текста. В дальнейшем необходимы более релевантная и корректная выборка по логистической регрессии, анализ поведения отдельного параметра текста как коррелирующего с той или иной характеристикой его продуцента, исследования на большом корпусном материале. Однако уже сейчас нам удалось подтвердить выдвинутую гипотезу о том, что наиболее релевантными параметрами для автоматизированного моделирования личности по тексту являются показатели частотности служебных слов и местоимений.

Список литературы

1. 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.
2. Литвинова Т. А. Установление характеристик (профилирование) автора письменного текста // Филологические науки. Вопросы теории и практики. 2012. № 2 (13). С. 90-94.
3. Литвинова Т. А. Языковые корреляты личностных особенностей автора письменного текста: алгоритм исследования // В мире научных открытий. Серия: Проблемы науки и образования. 2012. № 9.3 (33). С. 236-255.
4. Литвинова Т. А., Загоровская О. В., Середин П. В., Лантюхова Н. Н., Шевченко И. С. Профилирование автора письменного текста: подходы, методы и их оптимизация // Филология, искусствознание и культурология: актуальные вопросы и тенденции развития: материалы международной заочной научно-практической конференции (13 мая 2013 г.). Новосибирск: Изд. «СибАК», 2013. С. 69-79.
5. Фомина Н. А. Свойства личности и особенности речевой деятельности. Рязань: Узорочье, 2002. 412 с.
6. Lyons J. Linguistic Semantics. Cambridge: Cambridge University Press, 1995. 376 p.
7. Argamon Sh., Koppel M., Pennebaker James W., Schler J. Automatically profiling the author of an anonymous text // Communications of the ACM (Association for Computing Machinery). 2009. Vol. 52 (2). P. 119-123.

FORMAL-GRAMMATICAL CORRELATES OF AUTHOR'S PERSONAL CHARACTERISTICS OF WRITTEN TEXT

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

The article presents the pilot research of dependencies between the formal-grammatical parameters of the text and its author's personal characteristics, i.e. one of the aspects of the author's personality modeling problem of the written text is disclosed. The regression models linking the text formal-grammatical characteristics and its author's personal features are received when using the mathematical statistics methods. A specially created corpus of texts supplied with meta-marking as information about its authors was the material for the research.

Key words and phrases: text; study about authors; linguo-statistics; mathematical linguistics; function words; correlation analysis.

УДК 82.0

Филологические науки

В статье выявляются жанровые особенности неклассического военного произведения Новейшего времени. Предполагается, что предметом эпического видения становятся личностные конфликты, авторская позиция актуализируется в форме субъективного личного рассказа, читатель помещается в ситуацию экзистенциального этического выбора. Выявленные особенности рассматриваются на материале анализа военного рассказа В. Быкова «Короткая песня», в поэтике которого системную смыслопорождающую роль выполняет образный мотив песни.

Ключевые слова и фразы: классическая эпика; неклассическая военная проза; автор; читатель; повествователь; рассказчик; герой; персонажи; символ.

Максимов Владимир Владимирович, к. филол. н., доцент
Национальный исследовательский Томский политехнический университет
v_v_maksimov@rambler.ru

**«КОРОТКАЯ ПЕСНЯ» В. В. БЫКОВА.
ЖАНРОВЫЙ КОД НЕКЛАССИЧЕСКОГО ВОЕННОГО ПРОИЗВЕДЕНИЯ[©]**

Одним из ключевых моментов развития мировой литературы, по мнению многих авторитетных исследователей [1; 2; 3], является переход от классических типов литературного творчества к неклассическим формам словесности, фронтально обозначившийся в ареале европейских культур на границе XIX-XX веков. Этот момент качественного перехода обусловлен рядом факторов, в числе которых основополагающим принято считать открытие нового измерения творческой рефлексии и дискурсивных практик, а именно – измерения межличностного взаимодействия и взаимопонимания. Безусловно, и раньше оно привлекало внимание писателей, но этот локальный интерес всегда подчинялся иным способам и горизонтам художественно-эстетического видения. Внутри неклассического творческого мышления место человека стало осмысляться и оцениваться не столько функционально как нормативно обустроенная роль социального характера в общем миропорядке, сколько событийно как возможное присутствие индивидуальной личности в контексте становящихся связей незавершенного жизненного цикла.