

ЛИНГВИСТИКА ТЕКСТА

УДК Ш141.2 + Ш11
ББК 821.161.1 + 81'322.4

ГЕНДЕРНОЕ ПРОФИЛИРОВАНИЕ АВТОРА СУБСТАНДАРТНОГО ТЕКСТА

С.И. Красса, Е.Н. Калугина

Предлагается методика гендерного профилирования субстандартных текстов на основе разнообразных языковых и экстралингвистических характеристик: лексических, морфологических, синтаксических, стилистических, эрративных и фоносемантических параметров путем сужения показателей атрибуции текста сначала до гендерно идентифицирующих и затем до субстандартно идентифицирующих. Методика включает в себя три этапа, а именно: определение способов анализа и инструментария (используемое также в атрибуции текста); затем актуализация параметров, характерных для различия автора по полу; на последнем этапе происходит сужение на материале субстандартных текстов. При этом последовательность спецификации «гендер / субстандарт» может быть изменена.

Ключевые слова: атрибуция текста, гендер, текстовые параметры, компьютерная программа, субстандарт.

В гендерной лингвистике на современном этапе отмечается разнообразие идей относительно ключевых концептов научного направления, начиная от предложения ввести понятия «языковая личность мужчины» и «языковая личность женщины» [6] и заканчивая обособлением термина «гендерлект» [5], и это разнообразие предопределяет необходимость проведения гендерного профилирования текста с учетом разнообразных языковых и экстралингвистических характеристик.

Как известно, социальные роли, в том и числе гендерные, тесно связаны с речевыми особенностями человека. В речи отражается также социальный статус (профессия, уровень культуры, в том числе и речевой), являющийся более постоянным индикатором и составляющим социального «портрета» личности. Человек, являясь частью различных социальных объединений, выполняет разнообразные социальные роли, определяющие выбор используемых языковых средств [8]. С другой стороны, выполняя одну и ту же социальную роль, разные люди в силу имеющихся уровней речевой культуры, выразительности, психологических и других особенностей будут писать по-разному.

На характеристику речевого поведения могут оказывать влияние внутренние и внешние факторы, например эмоционально окрашенный стиль письменной речи проявляется в больших отступлениях от лексических и синтаксических норм, в увеличении количества ошибок и описок [12]. Индивидуальность в отношении письменной речи характеризуется наличием определенных показателей, а также спецификой речевых навыков конкретного человека, проявляющейся в уровне его интеллекта, характеристике его профессиональных и жизненных интересов, особенностях словарного

запаса, стилистических конструкций, наличием типичных слов и выражений [12].

В связи с этим одна из идей профилирования текста связывается нами с социолингвистической переменной – «величиной, которая зависит от некоторой нелингвистической переменной социального контекста: говорящего, слушающего, аудитории, обстановки и т. п.» [11, с. 150]. Лингвистические признаки, которые У. Лабов называет индикаторами, «образуют регулярное распределение по общественно-экономическим, этническим или возрастным группам, но в речи каждого индивидуума проявляются более или менее одинаковым образом в любом контексте» [9].

Е.И. Горошко, обобщая результаты исследований западных ученых по данной теме, отмечает, что в список стилистических показателей входят: 1) распределение частоты длины слов, 2) среднее количество слогов в словоформах, 3) длина слогов, 4) среднее количество словоформ в предложении, 5) особенности распределения частеречной принадлежности слов от их позиции в предложении, 6) число служебных слов, 7) индексы лексического разнообразия и показатель величины энтропии текста, а также 8) показатели лексической дистрибуции [3]. Особого внимания заслуживает статистический анализ тех признаков письменной речи, наличие которых не зависит от содержания текста. Полагаем, что все приведенные исследователем параметры относятся именно к таким признакам.

Е.И. Горошко провела также собственные исследования на материале русского языка и выявила следующие закономерности: словарь женщин по показателю лексического разнообразия оказался беднее, в речи мужчин было больше слов с час-

тотой один и два, в остальных показателях не было выявлено значительных различий [3].

Из сказанного следует, что проблема моделирования языковой личности автора письменного текста может решаться на различных теоретических основаниях. Вместе с тем в контексте лингвистической и автороведческой экспертизы текстов она по определению должна переходить из области теоретических построений в плоскость практических решений. Подобная облигаторность напоминает деятельность лексикографа, вынужденного принимать практические решения в далеко не однозначных теоретических областях, таких как разграничение полисемии и омонимии, конверсивных пар «существительное – глагол», например, в английском языке, коллокаций, которые подаются, как правило, внутри зоны значений, и идиом, представляемых в отдельной, «заромбовой» в русской словарной традиции, зоне.

Полагаем, что гендерное профилирование субстандартного текста должно исходить из следующих теоретико-практических предположений.

1. Определение гендерных особенностей текста, в том числе и субстандартного, исходит из тех же предпосылок, что и атрибуция текста, поскольку при определении авторства имеет место опора на текст как проекцию языковой личности автора на текстовые параметры.

2. Из всего разнообразия текстовых параметров выбираются те, которые в минимальной степени способны контролироваться автором. Набор параметров опирается на идеи стилеметрии Н.А. Морозова [13], работы по установлению авторства «Тихого Дона» [14], по автороведческой экспертизе в современной лингвоюридической практике [7].

3. Нами предлагается оптимальное, как видится на настоящее время, сочетание текстовых параметров, слабо контролируемых автором, и программного компьютерного инструментария, способного обеспечить относительную объективность анализа большого массива данных и высокую производительность такого анализа. Использование современных компьютерных программ, дающих возможность осуществлять достоверный и эффективный анализ значительного количества данных, можно считать значительным шагом в совершенствовании инструментария атрибуции текста в автороведческой экспертизе.

Мы предлагаем проводить процедуру гендерного профилирования текста по следующим параметрам:

- лексическим (средняя длина слова, длина слова в слогах, индекс лексического разнообразия);
- морфологическим (частеречный состав исследуемых текстов);
- синтаксическим (длина предложения, позиция частей речи в предложении, тип предложений – простое, сложносочиненное, сложноподчиненное, сложное с разными видами связи);
- стилистическим (комплексные характеристики, представляющие стиль автора как целое и

идентифицирующие его в том или ином отношении – например инвариант Фоменко);

- фоносемантическим (подсознательное восприятие звуковой стороны текста) [7]. На стыке плана содержания и плана выражения находится отрасль языкознания, призванная исследовать звукоизобразительную (звукоподражательную, звуко-символическую) систему языка [2]. Значение звуков может быть описано с помощью шкал, включающих в себя прилагательные, относящиеся к понятиям различного порядка и формирующие субъективные семантические пространства;

- эрративным (устанавливающим типичные ошибки или отсутствие таковых, а также уникальные ошибки).

Данные для анализа могут быть получены с помощью такого компьютерного инструментария, как программа «Худломер», стандартной программы Microsoft Word (сервис «Статистика»), сайта Мультитран, программ Style Recognition System, «Лингвоанализатор», «Сервис анализа текстов и сайтов» и других [7]. Существуют также специальные компьютерные программы, с помощью которых возможно провести фоносемантический анализ текста. К ним относятся ВААЛ, DIATON и другие, в основу работы которых положены исследования Ч. Осгуда, алгоритмы А.П. Журавлева [4] и В.В. Левицкого [10]. Как утверждают разработчики, данные программы позволяют определить фоносемантические характеристики слова и текста, типы кодирования, звуко-цветовые ассоциации, связанные с текстом, ритмические характеристики текста, «золотое сечение» текста.

Ввиду того что программы используют различные алгоритмы, которые основываются на тех или иных теориях, в исследовании следует применять разные программные средства там, где это возможно, с целью минимизации погрешностей получения результатов профилирования. Анализ величины, характеризующей степень разброса количественных значений показателей статистической выборки (случайных величин) относительно среднего значения для этой выборки, может служить идентификационным фактором для определения гендерной принадлежности текста. Однако остается не установленным точно, какая именно величина является значимой в данном случае.

Комбинация объективного (в той мере, насколько это обеспечивается теоретическими основаниями и алгоритмами, положенными в основу программ) статистического компьютерного анализа и исследование малоконтролируемых индикаторов позволяют считать предлагаемую методику весьма эффективной и лишенной, насколько это возможно, субъективности. Мы полагаем, что данный подход к анализу текста может быть успешно использован и для неидентификационного исследования.

5. Сведение полученных результатов в единый «координационный» блок, в котором проводится «взвешивание» получаемых данных и их ранжирование. Сначала ранжируются данные для «типичного» мужского и «типичного» женского

текстов. Чем больше текстов будет проанализировано, тем надежнее окажутся полученные результаты. Затем анализируются результаты для нормативных и субстандартных текстов.

Таким образом, в поставленной проблеме гендерного профилирования субстандартного текста насущными задачами являются следующие:

1. Создание концепции комплекса компьютерных программ, помогающих в решении установления авторского профиля текста.

2. Разработка или модификация программного обеспечения лексического, морфологического, синтаксического и стилистического модулей, а также координационного блока.

3. Проведение экспериментов по выявлению величины различия в параметрах, свидетельствующих в пользу мужского или женского авторства.

4. Формирование расширенного списка параметров и их ранжирование с целью профилирования автора текста.

В заключение отметим, что получение гендерного профиля автора субстандартного текста возможно в три этапа. На первом этапе определяется методика анализа и возможные программные средства, пригодные для этого. По существу, этот этап ничем не отличается от определения авторства, т.е. процедуры атрибуции текста. Затем атрибуция текста сужается до актуализации параметров, характерных для различия автора по полу. Подобное достигается путем многократного анализа текстов с известным по полу авторством. Наконец, на третьем, заключительном, этапе происходит еще одно сужение – на материале субстандартных текстов. Следует признать, что построение профиля автора субстандартного текста может быть проведено и на втором этапе – тогда у исследователя будут в распоряжении отличия профиля от литературной нормы. И затем уже проводится сужение до различий автора по полу. Однако порядок в получении спецификации «гендер / субстандарт» не окажет, по нашему мнению, существенного влияния на итоговые результаты.

Литература

1. Авторский инвариант. – http://ru.wikipedia.org/wiki/Авторский_инвариант (дата обращения 5.04.2014).
2. Воронин, С.В. Основы фоносемантики /

С.В. Воронин. – Изд. 2-е. – М.: ЛЕНАНД, 2006. – 248 с.

3. Горошко, Е.И. Изучение электронного дискурса в судебном речеведении / Е.И. Горошко <http://www.textology.ru/article.aspx?ald=44> (дата обращения 07.04.2014).

4. Журавлев, А.П. Звук и смысл / А.П. Журавлев. – 2-е изд. – М.: Просвещение, 1991. – 160 с.

5. Йокояма, О.Б. Когнитивный статус гендерных различий в языке и их когнитивное моделирование / О.Б. Йокояма // *Wiener Slawistischer Almanach, Munchen, Sonderband 55 (2002) X-XX*. – С. 29–40.

6. Каменская, О.Л. Гендергетика – междисциплинарная наука / О.Л. Каменская // *Тезисы докладов Второй международной конференции «Гендер: язык, культура, коммуникация», МГЛУ, Москва, 22–23 ноября 2001 г.* – М.: МГЛУ, 2001. – С. 62–63.

7. Красса, С.И. Методика и инструментарий атрибуции текста в автороведческой экспертизе С.И. Красса // *Альманах современной науки и образования*. – 2013. – № 10 (7). – С. 106–108.

8. Красса, С.И. Гендерные лингвоконцентры субстандартного текста / С.И. Красса, Е.Н. Калугина // *Вестник ЮУрГУ. Серия «Лингвистика»*. – 2012. – № 25. – С. 27–31.

9. Лабов, У. Исследование языка в его социальном контексте / У. Лабов // *Новое в лингвистике. Вып. VII. Социолингвистика*. – М.: Прогресс, 1975. – С. 5–33.

10. Левицкий, В.В. Фонетическая мотивированность слова / В.В. Левицкий // *Вопросы языкознания*. – № 1. – 1994. – С. 26–37.

11. Литвинова, Т.А. Лингвистические основы неидентификационной судебно-автороведческой экспертизы / Т.А. Литвинова // *Вестник Челябинского государственного университета*. – 2012. – № 20 (274). – Филология. Искусствоведение. – Вып. 67. – С. 74–78.

12. Морозов, Б.Н. Особенности криминалистического исследования письменной речи / Б.Н. Морозов. – Ташкент, 1985. – 40 с.

13. Морозов, Н.А. Лингвистические спектры / Н.А. Морозов. – <http://www.textology.ru/library/book.aspx?bookId=1&textId=3> (дата обращения 11.04.2014).

14. Кто написал «Тихий Дон»? (Проблема авторства «Тихого Дона») / Г. Хьетсо, С. Густавсон, Б. Бекман, С. Гил. – М.: Книга, 1989. – 263 с.

Красса Сергей Иванович, кандидат филологических наук, доцент кафедры лингвистики и лингводидактики, Северо-Кавказский федеральный университет (Ставрополь), skrassa@yandex.ru

Калугина Елена Николаевна, кандидат филологических наук, заведующая кафедрой иностранных языков и межкультурной коммуникации, Ставропольский государственный аграрный университет (Ставрополь), kalugina.elena2009@yandex.ru

Поступила в редакцию 5 мая 2014 г.

THE PROCEDURE OF THE AUTHOR'S GENDER PROFILING

S.I. Krassa, North-Caucasus Federal University, Stavropol, Russian Federation, skrassa@yandex.ru

E.N. Kalugina, Stavropol State Agrarian University, Stavropol, Russian Federation, kalugina.elena2009@yandex.ru

The article proposes the procedure of author's gender profiling in the nonstandard texts. It offers lexical, morphological, syntactic, stylistic, errative and phonosemantic parameters for carrying it out. The first narrowing is up to the gender identifying indicators and the second narrowing is up to the substandard identifying indicators. The technique includes three stages, namely the definition of analysis methods and tools (also used in the attribution of the text), then actualization of parameters specific for author identification by gender; the last stage is narrowing to substandard text. The sequence of specification "gender / substandard" can be changed.

Keywords: text attribution, gender, text parameters, computer program, nonstandard.

References

1. *Avtorskiy invariant* [Author's invariant]. Available at: http://ru.wikipedia.org/wiki/Авторский_инвариант (date of access 5.04.2014).
2. Voronin S.V. *Osnovy fonosemantiki* [Fundamentals of phonosemantics]. 2nd edition. Moscow: LENAND, 2006, 248 p.
3. Goroshko E.I. *Izycheniye elektronnoy diskursa v sudebnom rechevedenii* [Research of electronic discourse in court speech studies]. Available at: <http://www.textology.ru/article.aspx?aId=44> (date of access 28.03.2014).
4. Zhuravlev A.P. *Zvuk i smysl* [Sound and meaning]. 2-nd edition. Moscow: Prosveshcheniye, 1991, 160 p.
5. Yokojama O.T. *Kognitivnyi status gendernykh razlichiy v yazyke i ih kognitivnoye modelirovaniye* [Cognitive status of gender differences in the language and their cognitive modeling] // Wiener Slavistischer Almanach, Munchen, Sonderband 55 (2002) X-XX, pp. 29–40.
6. Kamenskaya O.L. *Gendergetika – mezhdistsiplinarnaya nauka* [Gendergetics is an interdisciplinary science]. Second International Conference «Gender: language, culture, communication», MSLU, Moscow, November 22-23, 2001. Moscow, MSLU, 2001, pp. 62–63.
7. Krassa S.I. *Metodika i instrumentariy atributsiyi teksta v avtorovedcheskoy ekspertize* [Methodology and tools of author attribution]. Almanac of modern science and education. Tambov: Gramota Publ., 2013, № 10 (7), pp. 106–108.
8. Krassa S.I., Kalugina E.N. *Gendernye lingvokontsentry substandarta* [Gender linguistic center of non-standard speech]. *Bulletin of South-Ural State University. Linguistics*, 2012, № 25, pp. 27–31.
9. Labov W. *Issledovanie yazyka v ego sotsialnom kontekste* [The study of language in its social context]. *New in linguistics, №.VII, Sociolinguistics*. Moscow: Progress Publishers, 1975, pp. 5–33.
10. Levitsky V.V. *Foneticheskaya motivirovannost slova* [Phonetic motivation of words]. *Issues of linguistics*, 1994, № 1, pp. 26–37.
11. Litvinova T.A. *Lingvisticheskiye osnovy neidentifikatsyonnoy sudebno-avtorovedcheskoy ekspertizy* [Linguistic basics of nonidentificational judicial author attribution]. *Bulletin of Chelyabinsk State University*, 2012, № 20 (274), Philology. Study of art. Issue 67, pp. 74–78.
12. Morozov B.N. *Osobennosty kriminalisticheskogo issledovaniya pismennoy rechi* [Peculiarities of criminological analysis of written language]. Tashkent, 1985, 40 p.
13. Morozov N.A. *Lingvisticheskiye spektry* [Linguistic spectra]. Available at: <http://www.textology.ru/library/book.aspx?bookId=1&textId=3> (date of access 11.04.2014).
14. Kjetsaa G., Gustavsson S., Beckman B., Gil S. *Kto napisal "Tikhiy Don" (Problemy avtorstva "Tikhogo Dona")* [The Authorship of The Quiet Don]. Moscow, Kniga. 1989. 263 p.

Sergei I. Krassa, PhD (Russian language), associate professor of Linguistics and Linguo-Didactics department (North-Caucasus Federal University), skrassa@yandex.ru

Elena N. Kalugina, PhD (Theory of language), head of Foreign Language and Intercultural Communication department (Stavropol State Agrarian University), kalugina.elena2009@yandex.ru

Received 5 May 2014