

УДК 811.111-26

DOI: 10.18384/2310-712X-2016-4-38-45

## ОПЫТ ГЕНДЕРНОЙ АТРИБУЦИИ РУССКИХ ТЕКСТОВ

**Мамаев М.М.***Московский государственный областной университет  
105005, г. Москва, ул. Радио, д. 10А, Российская Федерация*

**Аннотация.** В статье рассматриваются частотные характеристики гендерно значимых слов, отражающих гендерные особенности языка. Автором сделана попытка выделить этот пласт лексики в произведениях русской художественной литературы XIX и XX вв., написанных авторами-мужчинами и авторами-женщинами (общий объем проанализированного текста составляет более 200 тысяч слов). Анализ производился с использованием методов количественной лингвистики, а также апробированной ранее на английском материале процедуры гендерной атрибуции текста. Полученные результаты свидетельствуют о том, что в отличие от английского материала, в данном случае атрибуция текстов не может быть произведена для двух рассматриваемых эпох одновременно; их следует рассматривать по отдельности.

**Ключевые слова:** русский художественный текст, гендерные особенности, количественная лингвистика, гендерная атрибуция текста, показатель маскулинности.

## ON GENDER ATTRIBUTION OF RUSSIAN TEXTS

**M. Mamaev***Moscow Region State University  
10A, Radio str., Moscow, 105005, Russian Federation*

**Abstract.** The article studies the frequency characteristics of gender-intrinsic words, reflecting language gender peculiarities. The author makes an attempt to sort out this lexical layer while analyzing fiction, written in Russian in 19<sup>th</sup> and in the 20<sup>th</sup> centuries by male and female writers (the total amount of the material analyzed is more than 200 thousand words). The analysis was done with the usage of the methods of quantitative linguistics as well as the procedure of gender attribution which we tested earlier with the English material. The results received indicate that, as opposed to the English material, in this case text attribution can't be carried out for both time periods analyzed simultaneously; they should be considered separately.

**Key words:** Russian fiction, gender peculiarities, quantitative linguistics, gender text attribution, masculinity index.

Культурное своеобразие любого языка проявляется в полном объеме при рассмотрении гендерного фактора, который может по-разному проявляться в структуре как лексических, так и фразеологических единиц, иметь разную степень эксплицитности и влияния на внутреннюю форму слов и идиом. В целом,

можно сказать, что человек воспринимает реальность окружающего мира в виде представлений, которые формируют «картину мира» – целостный образ мира, представляющий собой результат всей его духовной активности [5, с. 19].

Анализ ряда параметров текста, выделяемых исследователями как наиболее значимые в гендерном отношении, позволяет говорить о том, что тексты, написанные авторами-мужчинами и авторами-женщинами, более всего различаются своим словарным составом. Это, в свою очередь, не означает, что первые употребляют слова, которые игнорируются вторыми, и наоборот. Речь идет о *частотной структуре* сопоставляемых текстов: степень употребительности определенных языковых единиц в тексте представителями того или иного пола может быть не только различной, но и достаточно специфичной.

Таким образом, основным исследовательским приемом в таких случаях является оценка употребительности: слова или лексические обороты являются «фемининными», если они существенно чаще встречаются в произведениях авторов-женщин. Успешно провести такую оценку позволяет использование элементов методов квантитативной лингвистики, таких как: а) выделение переменных или категорий исследуемых характеристик объекта до стадии исследования; б) оперирование усредненными показателями для обозначения переменных; в) обобщение на стадии анализа и стремление к выявлению универсальных закономерностей [3, с. 156].

Ранее нами была описана разработанная нами процедура определе-

ния показателя маскулинности **М** на материале англоязычных текстов [1]. Напомним, что статистический показатель степени маскулинности **М** позволяет дифференцировать авторов внутри заданной гендерной группы и сопоставить величину гендерных различий для авторов разных эпох.

Полученные с помощью разработанного нами алгоритма результаты гендерной принадлежности английских текстов, а особенно возникший вопрос о связи показателя маскулинности с эпохой создания произведения, подтолкнули нас рассмотреть вопрос гендерной атрибуции текста на материале русского языка, причем привлечь к анализу тексты писателей той же эпохи (XIX–XX вв.).

Наша основная задача была та же, что и при работе с английским материалом: возможность в результате лексического анализа текста с той или иной степенью успеха осуществить гендерную атрибуцию, т.е. определить, был ли автор текста мужчиной или женщиной. Для сопоставительного анализа нам был также выбран для рассмотрения пласт служебных слов и местоимений – наиболее употребительных лексем в текстах.

Первым шагом стал отбор служебных слов и местоимений, характерных для авторов-мужчин и авторов-женщин, для чего было необходимо оценить частотность отобранных слов в текстах.

Для анализа частотности служебных слов и местоимений мы привлекли 24 произведения русской литературы. В выборке были примерно в равной степени представлены произведения наиболее значимых писателей XIX в. с одной стороны и XX в. – с другой.

Общий объем выборки текстов авторов-мужчин и авторов-женщин был примерно равным (в обоих случаях он составил чуть более 100 тыс. слов). Следует отметить, что в число писателей XIX века мы включили наиболее значимых представителя эпохи (А.С. Пушкин, М.Ю. Лермонтов, Т.Л. Щепкина-Куперник, А.Я. Панаева и др.), а в число писателей XX века – лауреатов престижных российских премий последнего десятилетия (Б. Акунин, В.Г. Сорокин, В.Г. Галактионова, Л.Е. Улицкая и др.).

Следующим нашим шагом был анализ употребительности наиболее частотных служебных слов и местоимений в текстах. В нашей работе мы приняли за основу трактовку понятия «словоформа» В.А. Плунгяна, определяющего словоформу как «автономный комплекс, состоящий из достаточно жестко связанных между собой морфем» [4, с. 28].

Вначале мы отобрали группу из 40 самых частотных слов (по данным «Частотного словаря современного русского языка» [2]) и для каждого из них определили частоту его встречаемости в текстах нашей выборки, суммировав данные в отдельности для текстов авторов-мужчин и авторов-женщин.

Сопоставив полученные данные, мы выявили словоформы, которые можно однозначно отнести либо к категории «маскулинных», либо к категории «фемининных». Однако таких слов оказалось очень мало. Более тщательный анализ показал, что в отличие от английского материала частотность служебных слов и местоимений с течением времени в значительной мере *меняется*: весьма употребительные слова XIX в. оказываются не столь употреби-

тельными в XX в., и наоборот. В этой ситуации объединение данных по всей выборке оказалось нецелесообразным.

Работая с английским материалом, мы не имели этой проблемы – в значительной мере потому, что английский литературный язык к XIX в. полностью сформировался и имел стабильный статус за последние два века, в том числе и относительно статистической структуры лексического состава. О русском языке этого сказать нельзя, поэтому мы остановились на данном вопросе более подробно.

Сначала мы сопоставили частоту служебных слов и местоимений в отдельности для текстов XIX в. и XX в., а также сравнили эти данные с данными частотного словаря русского языка. Поскольку объем выборок в этих трех источниках был разным, мы сравнивали не абсолютные данные о частоте, а частоты слов в расчете на 100 тыс. слов текста.

Первое, что бросается в глаза, это общее резкое падение употребительности наиболее частотных словоформ в современной литературе (ср. 19006 в XIX в. : 15643 в XX в.). Обращает на себя внимание и тот факт, что данные частотного словаря гораздо ближе к данным текстов XIX в. Как нам известно, данные словаря были получены на материале литературных источников XIX и XX вв. По-видимому, тексты XIX в. составляли существенную долю всей выборки.

Иная картина будет, если говорить об относительной частоте отдельных слов. Для дальнейшего анализа целесообразно было перейти от абсолютных величин к рангам. Легко заметить, что все три последовательности рангов мало отличаются друг от друга. Под-

твердим это с помощью известного коэффициента ранговой корреляции Спирмена ( $\rho_{\text{сп}}$ ).

$$\rho_{\text{сп}} = 1 - \frac{6 \sum_{i=1}^n (R_i - S_i)^2}{n^3 - n}$$

Здесь

$n$  – число ранжируемых объектов (у нас  $n = 24$ );

$R_i$  – ранг  $i$ -ого слова в последовательности  $R$ ;

$S_i$  – ранг того же  $i$ -ого слова в последовательности  $S$ .

Опуская подробности подсчетов, скажем, что в итоге мы пришли к выводу, что в терминах ранговой корреляции можно сказать, что относительная частота служебных слов и местоимений в текстах XIX и XX вв.

практически совпадает. Тем не менее, целый ряд слов имеет настолько различную частоту встречаемости в текстах XIX–XX вв., что их объединение сильно исказило бы общую картину.

Таким образом, мы приняли решение составлять списки частотности *отдельно* для текстов XIX в. и текстов XX в.

Напомним, что наш изначальный список включал 40 словоформ. Как и в случае с английским материалом, мы начали с того, что определили частоту каждого слова в отдельности по следующим категориям: авторы-мужчины XIX в.; авторы-мужчины XX в.; авторы-женщины XIX в. и авторы-женщины XX в. Затем мы обобщили данные об употребительности слов и представили их в обозримом виде. Фрагмент этой исходной матрицы данных представлен в табл. 1.

Таблица 1

#### Употребительность слов в текстах (фрагмент исходной матрицы)

Группа	Слово	XX век		XIX век	
		Тексты мужчин	Тексты женщин	Тексты мужчин	Тексты женщин
I	в	211	<b>233</b>	187	<b>216</b>
	ни	15	<b>24</b>	18	<b>26</b>
	очень	7	<b>14</b>	12	<b>20</b>
II	мы	<b>20</b>	12	15	17
	на	172	167	<b>159</b>	129
	вы	14	11	<b>31</b>	20
	кто	<b>16</b>	8	5	6
	а	66	72	<b>82</b>	52
	и	301	303	<b>373</b>	293
III	за	<b>49</b>	38	37	<b>49</b>
IV	никто	5	4	2	2
	не	209	206	205	198

Рассмотрев количественное соотношение лексических единиц в текстах авторов-мужчин и авторов-женщин, мы

выделили случаи наибольшей разницы между ними и распределили слова на маскулинные и фемининные (эти слу-

чаи отмечены в табл. 1 жирным шрифтом). Как мы и намеревались, при выделении маскулинных и фемининных слов мы рассматривали данные отдельно для авторов XIX в. и авторов XX в.

Представленные в табл. 1 слова разделены на четыре группы. Группу I составили слова, которые оказались маскулинными (или соответственно фемининными) и в XIX в., и в XX в.

Наиболее многочисленная группа II включает слова, которые в одном веке являются выражено маскулинными или выражено фемининными, а в другом веке встречаются в мужских и женских текстах примерно с равной частотой (например, *мы*, *кто* в XX в. или *а*, *и* – в XIX в.).

Самый парадоксальный случай – слово *за* (группа III), которое в XIX в. было маскулинным, а в XX в. – фемининным.

Наконец, большое число слов встретились в мужских и женских тек-

стах одинаково часто. Примеры этой гендерно-нейтральной лексики собраны в группе IV.

Было естественно для окончательного определения списка маскулинных и фемининных слов каждого века отобрать такие лексемы, в которых соответствующая разность была бы максимальной.

Следующим нашим шагом была работа над системой коэффициентов для маскулинных и фемининных слов. Понятно, что, учитывая разный количественный и качественный состав этих категорий слов, системы коэффициентов для авторов XIX и XX вв. были бы разными. Принцип приписывания коэффициентов мы уже описывали ранее [1], когда представляли результаты исследования английского материала. Результатом этой процедуры стало создание двух систем коэффициентов для двух эпох (табл. 2).

Таблица 2

## Сопоставление системы коэффициентов для текстов XIX и XX веков

XIX век				
Фемининные слова			Маскулинные слова	
№	слово	вес		слово
1	<b>очень</b>	40	50	<i>тотчас</i>
2	<b>ни</b>	25	24	<b>вы</b>
3	<i>за</i>	20	12	а
4	<i>но</i>	14	10	он
5	<b>я</b>	3	4	<b>на</b>
XX век				
1	<i>теперь</i>	40	50	<i>поэтому</i>
2	<b>очень</b>	40	40	никто
3	под	30	35	<b>вы</b>
4	будто	30	30	сразу
5	<b>ни</b>	20	30	кто
6	она	10	25	нет
7	<b>я</b>	5	25	мы
8	в	3	8	<b>но</b>

XIX век			
Фемининные слова		Маскулинные слова	
№	слово	вес	слово
9		4	за
10		4	из
11		3	на

Сопоставление систем коэффициентов для текстов XIX и XX вв. позволило нам прийти к следующим выводам:

1) Состав гендерно значимой фемининной и маскулиной лексики двух эпох весьма различен. В XIX веке было получено 5 фемининных и 5 маскулиных слов, в то время как в XX – 8 фемининных и 11 маскулиных.

2) Детальное рассмотрение списков демонстрирует наличие лексики, «мигрировавшей» из группы фемининных в группу маскулиных (эти слова выделены серым в табл. 6), но не наоборот. Вес этих слов при этом значительно уменьшается (ср.  $за_{XIX} - 20 : за_{XX} - 4$ ;  $но_{XIX} - 14 : но_{XX} - 6$ ).

3) Многие слова с изменением эпохи сохраняют за собой маскулиную / фемининную принадлежность, при этом наблюдаются лишь незначительные изменения их веса (эта лексика представлена в табл. 2 жирным шрифтом). Исключение здесь составляет

только местоимение *вы*, коэффициент которого вырос в XX в. с 24 до 35. Наконец, стоит отметить, что коэффициенты наиболее «весомых» лексических единиц (выделены в табл. 2 курсивом) не претерпели серьезных изменений.

Располагая описанными данными, мы получили возможность определить показатель маскулинности *М* каждого автора, применив апробированный нами ранее на английском материале алгоритм подсчета [1].

Расположение показателей маскулинности авторов двух эпох на шкале маскулинности (от наименее до наиболее маскулиных), позволило нам сделать вывод о корректно проведенной процедуре гендерной атрибуции, поскольку полученные результаты верно отражают гендерную принадлежность авторов. Затем мы сравнили данные писателей, представляющих XIX и XX вв. Табл. 3 демонстрирует обобщенные данные по русскоязычным авторам.

Таблица 3

## Сопоставление русскоязычных авторов разных эпох

	Авторы	Маск <sub>ai</sub>	Фем <sub>ai</sub>	Показатель М		
				максимальный	минимальный	в среднем по группе
XIX в.	мужчины	3676	2757	+ 0,840	- 0,111	+ 0,305
	женщины	2484	3757	- 0,011	- 0,730	- 0,352
XX в.	мужчины	3509	2448	+ 0,925	+ 0,152	+ 0,438
	женщины	2451	4168	- 0,284	- 0,634	- 0,437

Как мы видим из табл. 3, в данном случае разность между самым маскулинным и самым фемининным автором со временем *практически не меняется*: в XIX в. она составила:  $[+ 0,840] \div [- 0,730] = 1,57$ , а в XX в.:  $[+ 0,925] \div [- 0,634] = 1,56$ . Это коренным образом отличается от того, что мы обнаружили для англоязычных писателей [1]. Там гендерная дифференциация была явно отчетливой: в XIX в. разность составляла 0,74, а в XX в. – всего 0,45. Другими словами, у английских авторов налицо определенное сближение значений коэффициентов маскулинности мужчин и женщин. Происходит это за счет действия двух

разнонаправленных процессов: мужчины с течением времени становятся менее «маскулинными», а женщины – более «маскулинными».

В результате анализа гендерно значимой лексики русских текстов XIX–XX вв. было выявлено этнокультурное различие, заключающееся в необходимости создания систем коэффициентов отдельно для текстов XIX и XX вв., а также в наличии лексики, «мигрировавшей» из группы фемининных в группу маскулинных слов. Наличие вышеизложенных случаев подталкивает к мысли о том, что гендерная картина в языке сильнее меняется с течением времени.

#### ЛИТЕРАТУРА:

1. Василевич А.П., Мамаев М.М. Проблема выделения гендерно значимых параметров текста // Вестник Московского государственного областного университета. Серия: Лингвистика. 2014. № 2. С. 17–24.
2. Ляшевская О.Н., Шаров С.А. Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). 92 млн. словоупотреблений. М.: Азбуковник, 2009. 1112 с.
3. Кашеева А.В. Квантитативные и качественные методы исследования в прикладной лингвистике // Социально-экономические явления и процессы. Тамбов: Тамбовский государственный университет им. Г.Р. Державина, 2013. № 3 (049). С. 155–162.
4. Плунгян В.А. Общая морфология: Введение в проблематику: Учебное пособие. Изд. 2-е, исправленное. М.: Едиториал УРСС, 2003. 384 с.
5. Ражева Е.С. Комплексный лексико-семантический и лингвокультурологический анализ тематической группы «инсектофоны / insectophones» (на материале английского и русского языков): автореферат дис. ... канд. филол. наук. М., 2013. 24 с.

#### REFERENCES

1. Vasilevich A.P., Mamaev M.M. Problema vydeleniya genderno znachimykh parametrov teksta [The problem of the selection of gender sensitive text settings] // Vestnik Moskovskogo gosudarstvennogo oblastnogo universiteta. Seriya: Lingvistika. 2014. no. 2. pp. 17–24.
2. Lyashevskaya O.N., Sharov S.A. Chastotnyi slovar' sovremennogo russkogo yazyka (na materialakh Natsional'nogo korpusa russkogo yazyka). 92 mln. slovopotreblenii [A frequency dictionary of modern Russian language (on the materials of the National corpus of the Russian language). 92 million tokens]. M., Azbukovnik, 2009. 1112 p.
3. Kashcheeva A.V. Kvantitativnye i kachestvennye metody issledovaniya v prikladnoi lingvistike [Quantitative and qualitative research methods in applied linguistics] // Sotsial'no-ekonomicheskie yavleniya i protsessy. 2013. no. 3 (049). pp. 155–162.
4. Plungyan V.A. Obshchaya morfologiya: Vvedenie v problematiku: Uchebnoe posobie. Izd.

2-е, исправленное [General morphology: introduction to the problems: a textbook. Ed. 2-е, revised]. М., Editorial URSS, 2003. 384 p.

5. Razheva E.S. Kompleksnyi leksiko-semanticheskii i lingvokul'turologicheskii analiz tematicheskoi gruppy «insektofony / insectophones» (na materiale angliiskogo i russkogo yazykov): avtoreferat dis. ... kand. filol. nauk [Complex lexical-semantic and linguo-cultural analysis of the thematic group “insectophones” (in the English and Russian languages): abstract ... dis. kand. Philol. Sciences]. М., 2013. 24 p.

---

#### ИНФОРМАЦИЯ ОБ АВТОРЕ

*Мамаев Михаил Михайлович* – старший преподаватель кафедры английской филологии факультета романо-германских языков Института лингвистики и межкультурной коммуникации Московского государственного областного университета;  
e-mail: mikesp87@mail.ru

#### INFORMATION ABOUT THE AUTHOR

*Mikhail Mamaev* – senior lecturer at the department of the English Philology at the Faculty of Romanic and Germanic Languages in the Institute of Linguistics and Intercultural Communication of Moscow Region State University;  
e-mail: mikesp87@mail.ru

---

#### БИБЛИОГРАФИЧЕСКАЯ ССЫЛКА

*Мамаев М.М.* Опыт гендерной атрибуции русских текстов // Вестник Московского государственного областного университета. Серия: Лингвистика. 2016. № 4. С. 38-45.  
DOI: 10.18384/2310-712X-2016-4-38-45

#### BIBLIOGRAPHIC REFERENCE

*M. Mamaev.* On gender attribution of Russian texts // Bulletin of Moscow State Regional University. Series: Linguistics. 2016. no 4 pp. 38-45.  
DOI: 10.18384/2310-712X-2016-4-38-45