

Key words: corpus, simultaneous interpretation, inferring, shell noun, contextual meaning.

*А.Ю. Хоменко (Нижний Новгород, Россия)  
Национальный исследовательский университет  
«Высшая школа экономики»  
akhomenko@hse.ru*

## **КОМПЬЮТЕРНЫЕ МЕТОДЫ АНАЛИЗА ДЛЯ ОПРЕДЕЛЕНИЯ ГЕНДЕРНОЙ ПРИНАДЛЕЖНОСТИ ТЕКСТА. ОПЫТ ПРАКТИЧЕСКОГО ИССЛЕДОВАНИЯ<sup>1</sup>**

В настоящей статье речь пойдет о применении интегративного подхода к определению гендера в рамках решения задач судебной лингвистики. Автор интегрирует методы когнитивной науки, корпусной и, шире, компьютерной лингвистики, а также классический структурный анализ текста для идентификации характеристик мужской и женской речи.

*Ключевые слова:* гендер, мужская и женская речь, судебная текстовая атрибуция, компьютерная лингвистика, корпусная лингвистика.

Вопросы гендерной принадлежности речевого материала решаются с помощью методов многих научных направлений: компьютерной лингвистики, стилистической лингвистики корпусов, когнитивной и классической структурной лингвистики. Тем не менее сама проблема разделения мужской и женской речи, естественно, имеет когнитивно-социальную основу. Гендер, с одной стороны, является продуктом развития культуры и социума [Кирилина 2002: 238], с другой – склонность к категоризации проявляется у человека с детства, представляя собой суть процесса когнитивного развития, а гендер, безусловно, является одним из основных критериев категоризации [Калугина 2013: 253].

На современном этапе развития науки женская и мужская речь достаточно успешно дифференцируются с помощью применения компьютерных технологий и методов стилистического анализа. Стилостатистика позволяет осуществить подсчет различных текстовых длин (длины предложения, слова), являющихся идентификационными параметрами для разграничения речи разных гендеров [Степаненко 2017]. Компьютерные же технологии позволяют установить материальные экспликации некоторых когнитивных категорий, реализуемых в речи через лексику и грамматику и присущих тезаурусу разных гендеров. Так, с помощью автоматической обработки большого объема текстовых данных, собранных в корпус, ученым удалось установить, что «показатели говорят об эгоцентричной коммуникации женщин, а также эксклюзивности, т.е. направленности в процессе коммуникации на себя» [Степаненко 2017: 22]. Маркером этого является преобладание местоимений «я-группы» (т.е. местоимений всех возможных разрядов первого лица единственного числа в прямых падежах и их супплетивных форм в косвенных падежах: я, меня, мне и пр., мой,

---

<sup>1</sup> Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-31-27001 (19-312-90022)

моя, мой): в женской речи таких местоимений встречается значительно больше, чем в мужской [Там же].

Указанные выше постулаты находят свое практическое подтверждение ходе исследования текстов в рамках решения локальных задач судебного автороведения. Так, именно указанные выше характеристики могут оказать решающее влияние при формировании вывода о гендере в рамках юридизированного исследования речевого материала, поскольку именно они, имея цифровые эквиваленты, позволяют объективировать процесс анализа. Естественно, важно использовать их в совокупности с анализом других текстовых параметров, исследование которых обычно имеет качественный, а не количественный характер.

Итак, методами квалификативного анализа в рамках производства досудебного автороведческого исследования удалось установить, что представленный на анализ текст отзыва о компании, несмотря на материальную женскую авторизацию, вероятно, принадлежит мужчине. Об этом свидетельствовал ряд факторов.

Первым из них является отсутствие прямого поименования эмоций и оценок, отнесенных к предмету речи. Автор эксплицирует их в основном в рамках метафорических структур: «Руководитель был отправлен в надлежające места, вместе с Патентом(!)», «всей этой эпопее», «но для меня... сразу отправляется в черный список», «А за неадекватность сотрудника – отдельный огромный минус». Все основные компоненты текста, которые выражают основную его мысль, связанную с оценкой качества предлагаемой продукции и оказываемых услуг, сконструированы с помощью тропов («называли», «товарищи» – эмоционально окрашенные лексемы с негативной коннотацией, выражающей презрительное отношение) и фигур речи («них есть Патент(!)» – ирония, близкая к сарказму, «так и хотелось сказать: «Остановись! Что ты несешь?» – имитация диалога, употребление ярко окрашенного оценочного глагола в качестве центрального экспликата авторского отношения), некоторым образом маскирующих истинную оценочность (резко негативную) речи автора.

Доказано, что эмоциональный градус сам по себе не является признаком мужской или женской речи: и мужская, и женская речь могут быть эмоциональными. Тем не менее эмоциональная окрашенность проявляется у мужчин и женщин по-разному. Так, для женщин значительно более характерно прямое называние эмоций и высказывание своего отношения: «Прямые номинации чувств в мужских текстах встречаются в 2 раза реже, чем у женщин (веселились, соболезную, уважаю, хорошее настроение, разочаровал, бодность). Мужчины предпочитают не именовать чувства напрямую, особенностью проанализированных текстов является тенденция передавать чувства через метафорические выражения (пробирает до костей, душу выворачивает, слова зацепили, уровень Биби Кинга, переворачивает, проворачивает мясо, рубит душу просто “влет”») [Колпакова 2016: 98].

Еще одним синтаксическим элементом, дифференцирующим мужскую и женскую речь, являются вставные конструкции (ВК), а точнее, их смысловая наполненность. Так, для женской гендерной идентичности более характерно использование вставных конструкций, выражающих эмоцио-

нальное отношение к объекту или предмету речи [Кучмезова 2001: 176], «в мужской прозе доминирующее положение занимают ВК, передающие дополнительные сведения» [Там же]. В анализируемом отзыве однозначно преобладают конструкции, содержащие дополнительную информацию: «особенности которых я опишу в заключительной части», «почему обоснованны? Да потому, что для устранения неполноценности данный производитель имеет “уникальные разработки”. Но об этом ниже и подробнее», «если они вообще существуют», «термическим», «Си, Zп», «мы работаем только с комплектами, длиной 15 м». Из 6 ВК, присутствующих в тексте, 5 относятся к эксплицирующим дополнительную информацию и лишь одна выражает эмоциональное отношение к предмету речи («на свою голову»).

Указанные выше текстовые характеристики, безусловно, характеризуют гендер, дифференцируют его, тем не менее на ограниченном объеме материала (646 словоформ) этого недостаточно для формирования категорического вывода о гендерной принадлежности материала. Объективировать результаты качественного анализа и дополнить его в данном случае можно с помощью количественных характеристик текста.

Так, после стилеметрических подсчетов удалось выяснить, что средняя длина предложения анализируемого отзыва равна 20,13 слова. Это значит, что автор изъясняется достаточно длинными синтаксическими конструкциями. Такая особенность характерна более для мужчин, чем для женщин [Степаненко 2017: 19-20]. Средняя длина «женского» предложения равна примерно 15,17 слова [Там же].

Одним из важнейших аспектов, подтверждающих мужскую гендерную принадлежность автора анализируемого текста, является дисперсия местоименных сущностей в отзыве, связанная с когнитивными составляющими, рознятсями мужчин и женщин. В анализируемом отзыве наблюдаем явное преобладание местоимений «мы-группы» (т.е. местоимений всех возможных разрядов первого лица множественного числа в прямых падежах и их форм в косвенных падежах) более, чем в два раза над местоимениями «я-группы» (т.е. местоимений всех возможных разрядов первого лица единственного числа в прямых падежах и их супплетивных форм в косвенных падежах) (19:8) (см. рис. 1, 2, 3).

1	яко[ямы]объявлять[объект]к[к]слову[слова]ость[ость]ны[ны]и[и]используем[использовать]системы[система]
2	яко[ямы]объект[объект]к[к]слову[слова]ость[ость]ны[ны]и[и]используем[использовать]системы[система]вед
3	не[не]делаю[делаю]длина[длина]производитель[производитель]нам[нам]согласны[согласны]использовать[использовать]пр
4	кав[ка]наме[на]и[и]подробнее[подробнее]подробный[добрый]ны[ны]принять[принять]присылать[присылать]и[и]тому[т
5	яко[ямы]и[и]подробней[подробней]подробный[добрый]ны[ны]принять[принять]присылать[присылать]и[и]тому[т
6	яко[ямы]объект[объект]як[як]свой[свой]счет[счет]чтобы[чтобы]ны[ны]применил[применил]это[это]и[и]его[его]и[и]сравн
7	яко[ямы]объект[объект]як[як]свой[свой]счет[счет]чтобы[чтобы]ны[ны]применил[применил]это[это]и[и]его[его]и[и]сравн
8	яко[ямы]другой[другой]производитель[производитель]и[и]от[от]нас[нас]требовалось[требовалось]написать[написать]только
9	[системы]компания[компания]тапа[тап]тип[тип]ИРСК[ИРСК]ны[ны]для[для]длин[длин]на[на]место[место]и[и]некогда[и]не
10	твоей[твоей]компания[компания]тапа[тап]тип[тип]ИРСК[ИРСК]ны[ны]для[для]длин[длин]на[на]место[место]и[и]некогда[и]не
11	еще[еще]равнение[равнение]доберутся[доберутся]и[и]до[до]нас[нас]потому[потому]нас[нас]длин[длин]кванты[кванты]
12	[кото]длин[длин]и[и]и[и]есть[есть]выть[есть]патент[патент]ны[ны]и[и]не[не]иметь[иметь]право[право]правый[правый]
13	горяч[горяч]у[у]ны[ны]и[и]есть[есть]Патент[патент]ны[ны]и[и]не[не]иметь[иметь]право[право]правый[правый]Одним[одним]
14	идея[идея]так[так]как[как]при[при]и[и]монтирует[монтирует]ны[ны]работает[работает]только[только]с[с]комплект
15	идея[идея]так[так]как[как]при[при]и[и]монтирует[монтирует]ны[ны]работает[работает]только[только]с[с]комплект

Рис. 1. Конкордантный список местоимения «мы», построенный в корпусном менеджере AntConc после лемматизации текста утилитой Mystem

```

1  менеджеры[менеджер]{названиями[называть]}в[в]нашу[наш]организацию[организация]{я[я]}назначив[назначив
2  ользовать]}свое[свой]производство[произудца]}в[в]наших[наш]проектах[проект]}и[и]на[на]наших[наш]объектах[о
3  дия]}в[в]наших[наш]проектах[проект]}и[и]на[на]наших[наш]объектах[объект]}к[к]слову[слово]хотят[хотят]мы[и]
4  [и]менеджеры[менеджер]{позвоним[позвонить]}в[в]нашу[наш]организацию[организация]{я[я]}место[место]типи

```

**Рис. 2.** Конкордантный список местоимения «наш», построенный в корпусном менеджере AntConc после лемматизации текста утилитой Mystem

```

1  ли[позвонить]}в[в]нашу[наш]организацию[организация]{и[и]}место[место]типичного[типичный]двух[два]грёк
2  {позвонить]}в[в]нашу[наш]организацию[организация]{я[я]}место[место]типичного[типичный]двух[два]грёк
3  олько]выслать[выслать]каталог[каталог]}но[но]име[в]удалось[удаваться]переговорить[переговаривать]}с[с
4  ампазон]}особенности[особенность]которых[который]}и[и]о[о]нашу[описывать]}в[в]закончательной[закончители
5  выпрон]}особенности[особенность]которых[который]}и[и]о[о]нашу[описывать]}в[в]закончательной[закончители
6  овы[высоке]звонке[звонке]звонкий]}и[и]тигерь[тигерь]}и[и]общаться[общаться]уже[уже]узкий|узко|ужк}{с[с]тре
7  ь[вновь]звонке[звонке]звонкий]}и[и]тигерь[тигерь]}и[и]общаться[общаться]уже[уже]узкий|узко|ужк}{с[с]треть
8  самостоятельно[самостоятельный]}но[но]для[для]меня[я]меньше[производитель[производитель]}который[который

```

**Рис. 3.** Конкордантный список местоимения «я», построенный в корпусном менеджере AntConc после лемматизации текста утилитой Mystem

Так, с когнитивной точки зрения центром повествования не является «я» говорящего, что дает возможность в сопряжении с другими текстовыми параметрами говорить о тяготении письменной речи отзыва к мужскому гендеру.

Доказать это в рамках судебного автороведения становится возможным после интеграции качественных и количественных методов анализа гендерной лингвистики.

### Литература

Калугина Е.Н. Гендер в антропоориентированных науках // Актуальные проблемы гуманитарных и естественных наук. 2013. № 5. С. 251-254.

Кирилина А.В. Новый этап развития отечественной лингвистической гендерологии // Гендерные исследования и гендерное образование в высшей школе: материалы международной научной конференции, Иваново, 25–26 июня 2002 г.: в 2 ч. Ч. II. История, социология, язык, культура. Иваново: Иван. гос. ун-т, 2002. С. 238-242.

Колтакова Л.В. Лексические средства репрезентации эмоций (гендерный аспект) // Вестник Томского государственного педагогического университета. 2016. № 7 (172). С. 96-100.

Кучмезова Л.О. Вставные конструкции в аспекте гендерной лингвистики // Вестник Костромского государственного университета им. Н.А. Некрасова. 2011. № 2. С. 175-179.

Степаненко А.А. Гендерная атрибуция текстов компьютерной коммуникации: статистический анализ использования местоимений // Вестник Томского государственного университета. 2017. № 415. С. 17-25. Doi: 10.17223/15617793/415/3.

**A.Yu. Khomenko** (Nizhny Novgorod, Russia)  
National Research University Higher School of Economics

### COMPUTER ANALYSIS FOR DETERMINING GENDER. PRACTICAL RESEARCH EXPERIENCE

This article will focus on an integrative approach to the definition of gender in the framework of solving the problems of forensic linguistics. The author integrates methods of cognitive science, corpus and, more broadly, computer linguistics, as well as classical structural analysis of the text to identify the characteristics of male and female speech.

Key words: gender, male and female speech, judicial text attribution, computer linguistics, corpus linguistics.