

## ЭКСПРЕССИВНОСТЬ КАК МАРКЕР ГЕНДЕРНЫХ РАЗЛИЧИЙ КОМПЬЮТЕРНОЙ КОММУНИКАЦИИ (К ПРОБЛЕМЕ АВТОМАТИЧЕСКОЙ ГЕНДЕРНОЙ АТРИБУЦИИ ТЕКСТА)

Приведены результаты исследования гендерных различий в текстах компьютерно опосредствованной коммуникации с использованием лингвистических и статистических методов анализа. В качестве маркеров гендерных различий изучаются лексемы, выражающие различные виды семантики экспрессивности: единицы, содержащие 1) компоненты рациональной и эмоциональной оценки; 2) компоненты интенсивности проявления признака; 3) морфемные маркеры эмоционально-экспрессивной окрашенности, единицы; 4) графическое маркирование выражения эмоционального отношения. При наличии сочетания маркеров экспрессивности, лексема включается не в одну группу.

**Ключевые слова:** компьютерная коммуникация; гендерная атрибуция текста; экспрессивность; эмоциональность текста; статистический анализ.

### Введение

Гендерная лингвистика – направление, интенсивно развивающееся в настоящее время в российской лингвистике, включающее широкий спектр частных проблем. Наиболее противопоставленными, на наш взгляд, аспектами анализа общей проблемы проявления гендерных различий в языке являются исследование способов маркирования гендерных различий в структурах языка и различий в структуре коммуникативной деятельности, как следствие, в текстах, порождаемых мужчинами и женщинами. В первом направлении, прежде всего, обособляются изучение грамматического маркирования гендерных различий, например обсуждаются вопросы своеобразия категории рода в структурах разных языков (см., например, [1–4]), и выявление маркирования гендерных стереотипов в лексике, фразеологии различных языков [5–7]. В совокупности исследований речевой деятельности в гендерном аспекте противопоставляется изучение особенностей речи (текста), стратегий и тактик их развертывания, своеобразия использования единиц различных уровней языковой системы в текстах, порождаемых мужчинами и женщинами [8–10], и особенности восприятия и ассоциирования мужчинами и женщинами языковых единиц и текстов разного типа [11–14]. Очевидно, что при решении данных проблем лингвист неизбежно вступает в проблемное поле ряда смежных гуманитарных наук – социологии, культурологии, психологии, когнитивистики.

К настоящему времени в исследованиях особенностей речевой коммуникации и порождаемых в данных процессах текстов получены заслуживающие внимание данные о типах различий, противопоставляющих речь мужчин и женщин: различия касаются характера использования языковых единиц разных языковых уровней [8–10, 15–17], что, как правило, связывается авторами проведенных ранее исследований с различным самопозиционированием мужчин и женщин в коммуникации, реже – с когнитивными особенностями мужчин и женщин.

При исследовании особенностей текстов, порождаемых мужчинами и женщинами, в качестве яркого отличительного свойства определяется их повышенная экспрессивность, проявляющаяся в различии ис-

пользования языковых разнуровневых средств выражения экспрессивности.

Вместе с тем уже на ранних этапах развития гендерной лингвистики было отмечено, что гендерное речевое позиционирование не может осуществляться изолированно, вне других социальных ролей, репрезентируемых в тех или иных речевых практиках [18]. Различие гендерных ролей может актуализироваться или подавляться в сочетании с другими социальными ролевыми позициями говорящих, отражающимися в менах коммуникативных ролей. Вследствие этого проблема гендерно обусловленных различий в коммуникации не может решаться вне проблемы дискурсивно и жанрово обусловленных вариаций коммуникативной деятельности.

В настоящее время в русистике исследования гендерно маркированных различий в коммуникации ведутся на материале различных дискурсов: художественного, рекламного, публицистического, политического [19–22]. Активно вовлекаются в аналитические работы этой направленности и различные жанры компьютерно опосредствованной коммуникации [17, 23], что мотивировано не только потребностью в разрешении ранее обозначенной научно значимой теоретической проблемы, но и наличием определенного прагматического социального заказа, потребностью в определении авторства текста, автор которого намеренно скрывается, в том числе и в криминалистической практике [24, 25].

В данном и других случаях решения практически ориентированных задач в последнее время необычайно остро актуализируется задача автоматической гендерной атрибуции текстов (см., например, [26, 27], а также обзор в [28]). Необходимо отметить, что во всех случаях при сравнении мужских и женских текстов речь идет о преобладании единиц определенного типа в тексте, а не абсолютном отсутствии / наличии единиц какого-либо конкретного типа. Вследствие этого возникает проблема доказательности выводов о преобладании единиц какого-либо типа, использование которых противопоставляет речь мужчин и женщин. Это актуализирует проблему использования методов статистического анализа как способа проверки релевантности выводов, сделанных с использованием лингвистических методов анализа. Достоверность

выводов о характере использования каких-либо единиц в речи в качестве маркеров типологических различий в настоящее время проверяется в системах автоматической обработки текстов, в методах проверки лингвистических гипотез с опорой на статистические методы анализа, методы математического моделирования.

Все сказанное ранее определяет, на наш взгляд, актуальность обращения к выявлению маркеров гендерных различий в социально актуальных современных дискурсах и определение степени надежности использования определенного типа маркеров в системах автоматической обработки текстов.

### **Постановка задачи, материал и методы**

В статье с опорой на выработанные в гендерной лингвистике данные о повышенной экспрессивности женской речи, о наличии различий в характере эмоциональности мужчин и женщин, обнаруживаемых в коммуникации, мы анализируем различные типы экспрессивных единиц в качестве маркеров различий мужской и женской коммуникации, сравнивая силу маркирующих различий эмоциональности как прагматической и семантической категории.

Основная задача, результаты разрешения которой представлены в статье, – определение диагностирующей силы лексических маркеров эмоциональности и экспрессивности текста в системе автоматической гендерной атрибуции текста.

Данная задача решается на материале текстов компьютерной коммуникации, текстов персональных страниц социальной сети «ВКонтакте». Выбор материала исследования мотивируется следующим. Как было отмечено ранее, в настоящее время в виртуальной коммуникации Интернета содержатся электронные аналоги практически всех типов текстов «реальной коммуникации», и в то же время Интернет – это пространство формирования новых дискурсивных и жанровых форм общения, в наиболее непосредственной форме сочетающих черты первичных форм устной речи – личностное обыденное общение, протекающие в онлайн форме, и черты, характерные для письменной речи, к чему следует отнести, прежде всего, ее фактуру – систему письменных, визуально воспринимаемых знаков [29–31]. К таким жанровым формам следует, на наш взгляд, отнести тексты персональных страниц социальной сети «ВКонтакте», которые интерпретируются нами как принадлежащие к сфере естественной письменной речи в концепции Н.Б. Лебедевой [30], как одна из жанровых форм личностно-ориентированных бытовых дискурсов в противопоставлении институциональным и личностно-ориентированным бытийным [32]. К характерным чертам личностно-ориентированных бытовых дискурсов, значимым при анализе текстов персональных страниц «ВКонтакте», относим то, что в них, с одной стороны, «говорящий выступает как личность во всем богатстве своего внутреннего мира», с другой стороны, «этот тип дискурса характеризуется спонтанностью, сильной ситуативной зависимостью, ярко выраженной

субъективностью» [32. С. 5–6]. Социальная сеть «ВКонтакте» исследователями в жанровом аспекте интерпретируется как гипержанр, персональная страница – как наджанровое макрообразование, включающая жанры «анкета», «статус», «записи на стене», «личные сообщения», «обсуждения», «комментарии» [33. С. 24], «типичными стилевыми чертами» которых «являются эмоциональность, субъективность и имитация разговорной спонтанности при помощи экспрессивно-окрашенной лексики, разговорного синтаксиса, звукового письма, эмотиконов и экспрессивной пунктуации» [Там же. С. 23]. Важно, что тексты данных жанровых форм порождаются спонтанно, в естественных условиях.

В представляемом исследовании тексты были отобраны для анализа после порождения речи, протекающей в естественных условиях<sup>1</sup>. Конкретным материалом исследования послужили 19 диалогов компьютерной коммуникации, корпус текстов был собран в рамках учебной практики студентов отделения фундаментальной и прикладной лингвистики филологического факультета Томского государственного университета. Объем каждого диалога составил 150–200 Кб, 4 000 слов. В целом объем проанализированных текстов составляет 120 страниц, 80 300 слов.

### **Метод**

При решении поставленной задачи мы опирались на сочетание лингвистического и математического анализа текста.

На первом этапе после предварительной технической обработки текстов применялись собственно лингвистические методы контекстуального анализа лексических единиц, в результате выявлен состав маркеров экспрессивности и эмоциональности текстов.

На втором этапе с помощью приемов статистического анализа была проведена количественная квалификация использования выявленных маркеров в мужских и женских текстах и выявлена степень статистической достоверности (значимости) выявленных количественных различий в использовании данных маркеров, на этой основе была дана оценка диагностирующей силы классов и отдельных единиц при маркировании гендерной специфики.

Охарактеризуем более подробно применение методов и их результаты.

Перед началом анализа исходный материал – диалоги – были разделены на мужские и женские реплики. Все тексты были нормализованы: лексические единицы были приведены к одному стилю, устранены неинформативные формальные признаки (знаки пунктуации, ссылки и т.п.) и произведена лемматизация (приведение единиц к начальным формам – леммам). Нормализация и лемматизация осуществлялись при помощи языка программирования R (пакет «Quanteda») и интегрированным в него стеммером – «Mystem».

В результате текст был преобразован в список лексических единиц, возведённых к начальным формам. Из данного списка были выделены единицы, которые, по гипотезе авторов исследования, могут



томатического, машинного распределения текстов на мужские и женские на основании выявленных статистических различий в использовании лексических маркеров рассматриваемых классов.

Для подсчета частот лексических маркеров по выбранным признакам и их объединения в группы для последующего определения степени статистической значимости выявленных количественных различий в использовании данных маркеров применялась формула веса *tf-idf*. Суть ее заключается в подсчете веса некоторого слова пропорционально коли-

честву употреблений этого слова в группе текстов и обратно пропорциональному частоте употребления слова в других. Например, если слово *молодец* чаще встречается в женских текстах, а реже в мужских, тогда вес этого слова в женских текстах будет стремиться к 1, а в мужских к 0. Итогом работы алгоритма стала частотная матрица (*tf-idf*) суммы относительных величин использования лексических единиц мужчинами и женщинами, которые были сгруппированы нами по вышеуказанным признакам. Результаты данного анализа представлены в таблице.

Таблица 1

Фрагмент матрицы сумм весов групп маркеров экспрессивности текста

Текст	Лексемы-интенсивы	Оценочные лексемы	Лексемы-димиутивы	Лексемы с графическими маркерами
dlg1_f.txt	42,77	165,76	31,95	319,89
dlg2_f.txt	0,00	0,84	1,08	3,32
dlg4_f.txt	0,78	6,84	0,60	2,54
dlg3_f.txt	2,84	40,28	6,95	23,91
dlg5_f.txt	56,42	273,64	48,84	58,48
dlg6_f.txt	1,56	23,37	3,24	7,47
dlg7_f.txt	8,62	44,15	10,67	20,49
dlg8_f.txt	7,25	21,31	15,59	28,64
dlg9_f.txt	16,43	57,61	10,84	18,95
dlg10_f.txt	0,00	10,70	0,00	2,89
dlg11_f.txt	7,54	45,32	7,18	64,76
dlg12_f.txt	0,60	4,51	0,00	1,86
dlg12_m.txt	54,93	208,20	41,18	219,61
dlg1_m.txt	0,00	3,27	0,00	2,16
dlg2_m.txt	0,00	5,44	0,00	0,00
dlg3_m.txt	14,16	54,36	5,22	32,11
dlg4_m.txt	51,53	195,96	36,74	51,31
dlg5_m.txt	6,11	23,49	2,51	15,60
dlg6_m.txt	12,52	56,20	10,67	15,06
dlg7_m.txt	7,68	19,74	12,65	29,04
dlg8_m1.txt	10,88	63,89	9,29	12,02
dlg9_m.txt	5,24	16,56	4,02	3,06
dlg10_m2.txt	11,26	85,47	28,31	12,20
dlg11_m2.txt	0,48	4,77	0,00	3,92

Данная матрица является исходным материалом для дальнейшего анализа текстов с применением других формально-количественных методов.

На следующем этапе была подсчитана сумма весов лексических единиц всех групп мужских и женских реплик. Отметим, что при этом мы объединили единицы всех четырех групп маркеров, удалив повторяющиеся лексемы, т.е. лексемы, имеющие более

чем один компонент. Например, текстовая единица *ужаааас*, входящая в три класса слов с семантикой оценочности, интенсивности, которые содержали графические маркеры интенсивности, в данном подсчете учитывалась один раз. Сумма весов лексических единиц оставила 1 529,3 в женских репликах, а в мужских – 458,8. Результат анализа представлен в диаграмме на рис. 1.

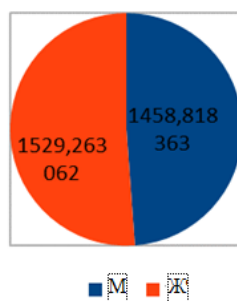


Рис. 1. Соотношение сумм весов экспрессивных единиц в мужских и женских репликах

Как видно из данных диаграммы, мужская и женская речь отличается в количестве используемых

экспрессивных единиц, женские реплики содержат больше экспрессивных и эмоциональных лексиче-

ских единиц (сегмент f на диаграмме), чем мужские (сегмент m). Однако проверка статистической значимости данных различий показала, что выявленный результат не является значимым: анализ на основании критерия Манна – Уитни выявляет уровень значимости  $p = 0,72$  ( $p > 0,05$ ), что свидетельствует о случайном характере выявленного распределения.

На следующем этапе исследования были подсчитаны в мужских и женских репликах суммы весов лексических единиц выделенных групп, выявлено преобладание маркеров экспрессивности в женских репликах (наблюдается в трех группах: лексем-интенсивы, лексем-димиутивы, оценочные лексем и отсутствует в группе лексем с графическими маркерами экспрессивности). Однако проверка статистической значимости выявленных количественных различий показала, что только относительно группы единиц-димиутивов уровень значимости составил  $p = 0,04$  ( $p > 0,05$ ). В использовании остальных групп лексических единиц в репликах мужчин и женщин

статистическая значимость не была выявлена ( $p > 0,05$ ), а следовательно, все различия случайны.

Так как полученный уровень значимости для лексем-димиутивов близок к пороговому уровню значимости 0,05, мы провели дополнительный статистический анализ сравнения дисперсий критерием Фридмана.

Цель дисперсионного анализа заключалась в проверке гипотезы о равенстве средних значений использования экспрессивов в зависимости от гендерной принадлежности участника коммуникации. Дисперсия характеризует разброс значений относительных частот использования анализируемых групп лексических единиц в текстах, противопоставленных по гендерной принадлежности автора. Результат анализа показан на рис. 2: как можно видеть, дисперсия различает группы лексем, но не различает гендерно противопоставленные тексты (уровень значимости составил 0,88, т.е. значительно больше 0,05).

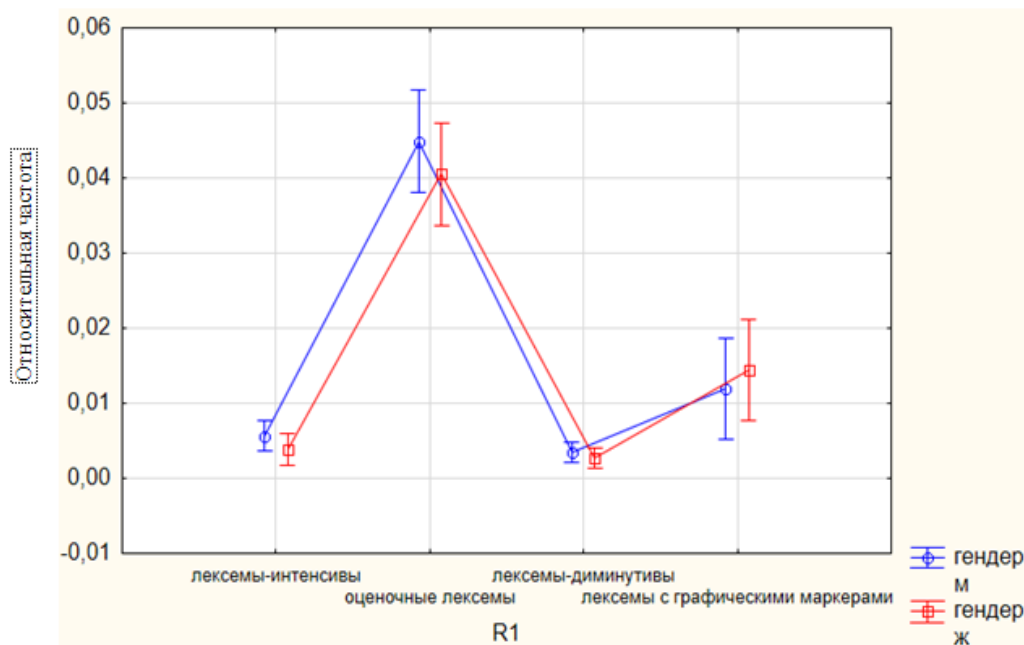


Рис. 2. Дисперсии относительных частот использования групп экспрессивных единиц

Полученные данные явились основанием предположения, что подобный результат вызван различиями внутригрупповых дисперсий: вероятно, частоты использования экспрессивных единиц могут отличаться в каждом диалоге. Чтобы проверить данное утверждение, мы провели дисперсионный анализ одной из групп лексем – лексем с графическими маркерами. Результат анализа представлен на рис. 3.

Как видно из рис. 3, во-первых, для дисперсий характерен большой размах, во-вторых, в то время как дисперсии по всем лексемам словаря с графическими маркерами отличаются, медианы единиц данного словаря почти равны как для мужчин, так и для женщин.

Результат анализа дисперсий по каждому из текстов позволяет сделать предположение о том, что различия в использовании экспрессивов определяются не только гендерной принадлежностью авторов, но и другими признаками коммуникации. Такое предположение коррелирует с идеями гендерной лингвистики о том, что гендер как социокультурная характеристика коммуникантов не может не вступать во взаимодействие с другими социально значимыми ролями коммуникантов [18], а также с классическими положениями теории дискурса о множественности факторов дискурсообразования (см., например, [39. С. 205–210]).

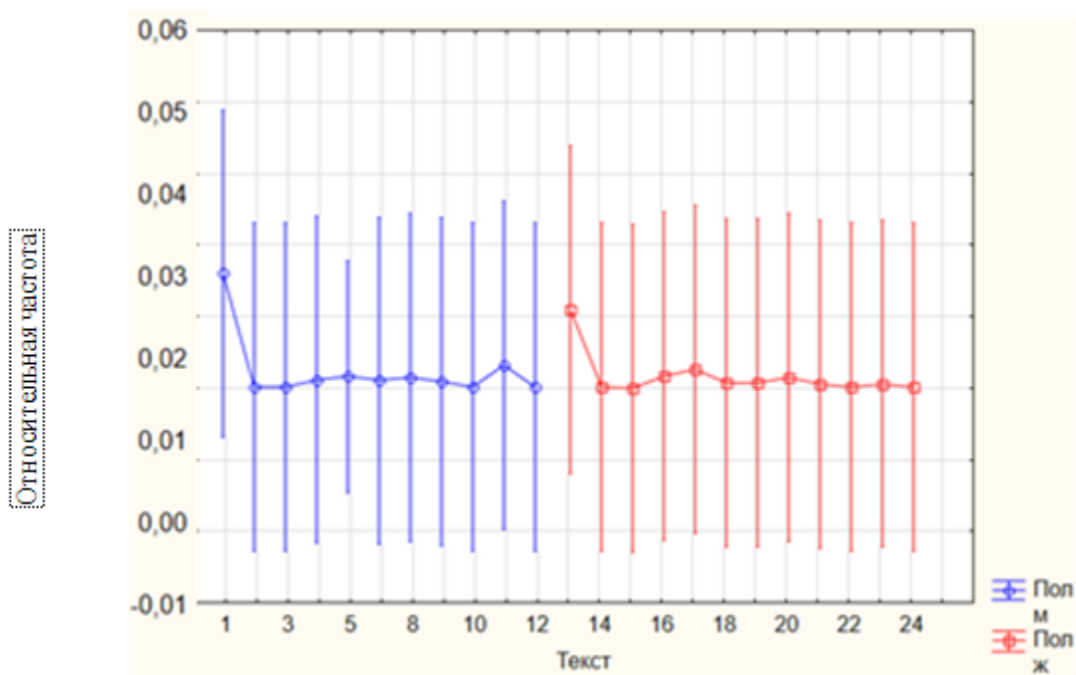


Рис. 3. Дисперсии относительных частот использования лексем с графическими маркерами экспрессивности

Для проверки этой гипотезы мы провели кластерный анализ «Уорда» [40. С. 298]. Под «кластерами» понимаются однородные (схожие) подгруппы, в которых дисперсия минимизирована внутри групп и максимизирована между группами. На вход подавалась матрица экспрессивных лексических единиц с относительными величинами без маркирования класса, результате тексты распределились на два основных кластера. В пер-

вом кластере (*dlg0 f1*, *dlg0 m2*, *dlg2 f2*, *dlg2 m1*) наблюдается превалирование экспрессивных лексических единиц в сравнении со вторым. Что касается второго кластера, то он противопоставляется первому кластеру (за счет максимизирования дисперсии между группами) и одновременно распадается на два подкласса, имеющие менее значительные различия в характере использования экспрессивных единиц (рис. 4).

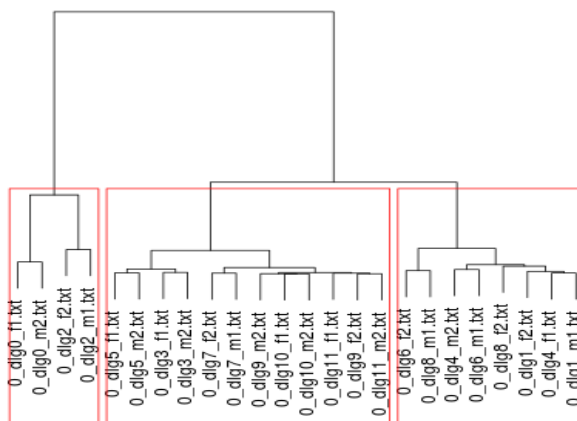


Рис. 4. Кластерная дендрограмма распределения текстов компьютерной коммуникации по признаку использования экспрессивных лексем

Как было отмечено, в первом кластере (*dlg0 f1*, *dlg0 m2*, *dlg2 f2*, *dlg2 m1*) наблюдается преобладание экспрессивных лексических единиц всех четырех классов в сравнении со вторым. Дискурсивный анализ текстов, вошедших в первый класс, выявил, что тексты первого кластера противопоставляются второму по двум дискурсивным признакам – тема коммуникации и личностные отношения коммуникантов. Реплики, вошедшие в первый кластер, при-

надлежат мужчинам и женщинам, имеющим длительный опыт личного общения, темой коммуникации являются по преимуществу межличностные эмоционально наполненные отношения. Так, в диалоге *dlg0* участники коммуникации (*f1* и *m2*), судя по репликам, знают друг друга продолжительное время, имеют общие интересы в спорте и учебе, они обсуждают личные отношения, отношения общих друзей, рассуждают о жизни, совместном времяпровожде-

нии, например, Ж: *Напьешься поди, а мне тебя до-мой тащить?*, М: *«Красава»*). )))). Или реплики из dlг 2: 1) М: *«Идешь завтра гулять?»*. Ж: *«Ааа, блин я завтра только в часов в 7 освобожусь»*. 2) М: *«Скучаю По тебе Скорее бы завтра услышать твой голос»*; Ж: *«Нуу) Я тоже скучаю по тебе»*). Объединяющей характеристикой диалогов второго кластера является тема деловых отношений, учебы, не исключающая выражения эмоционального отношения к данным темам, например, типичный диалог: dlг 1: Ж: *«Какими способами? Или как учится?»*) *Аааа, английский и испанский;*); М: *«Ого) Испанский охренеть) А почему Испанский?»*). Или dlг 4: Ж: *«А ты не мог бы сфотать свою тетрадь тему и прислать через вк?»*; М: *«прости я тогда не писал мы с сашкой сидели»*).

Таким образом, сочетание методов лингвистического и статистического анализа позволило нам сделать вывод о том, что различия в использовании экспрессивных единиц в текстах компьютерной коммуникации, противопоставленных по признаку гендерной принадлежности авторов, отмечаются, но не являются статистически значимыми, т.е. признак гендерного противопоставления не обнаруживается как устойчивый и воспроизводимый. Применение кластерного анализа позволило выявить взаимодействие гендерного признака авторов коммуникации с другими социально значимыми параметрами компьютерной коммуникации – темой текста, ролевыми и социальными позициями коммуникантов в диалоге, которые оказываются более значимыми факторами, определяющими характер использования экспрессивных единиц.

## ПРИМЕЧАНИЯ

<sup>1</sup> Извлечение диалогов из социальных сетей осуществлялось с согласия их авторов, которые, в соответствии с нормами регламента Этического комитета междисциплинарных исследований ТГУ (<http://lab.tsu.ru/cognitivestudies/node/14>) и в соответствии с Федеральным законом № 152 РФ «О персональных данных», были проинформированы о целях проводимого исследования и о гарантиях анонимности предоставленных персональных данных, после чего были заполнены «Формы информированного согласия», в структуру которых были включены метаданные участников диалогов: пол, возраст, социальный статус.

<sup>2</sup> Дискурсивно-жанровая специфика анализируемых текстов обусловила наличие значительного количества просторечной, сленговой, в том числе ненормативной лексики. Для уточнения значения данных единиц наряду со словарем литературного языка [37] мы использовали словарь Т.Г. Никитиной [38].

<sup>3</sup> В таблице приняты следующие обозначения: dlг-диалог, цифрами обозначается номер текста, f.txt и m.txt – маркеры гендерной атрибуции авторов текстов. Например, dlг0 – первый диалог, dlг0\_f – женские реплики из первого диалога; dlг0\_m – мужские реплики первого диалога.

## ЛИТЕРАТУРА

- Boroditsky L., Schmidt L. A., Phillips W. Sex, syntax, and semantics // *Language in mind: Advances in the study of language and thought*. 2003. P. 61–79.
- Kurinski E., Jambor E., Sera M.D. Spanish grammatical gender: Its effects on categorization in native Hungarian speakers // *International Journal of Bilingualism*. 2016. Vol. 20, № 1. С. 76–93.
- Landor R. Grammatical Categories and Cognition across Five Languages: The Case of Grammatical Gender and its Potential Effects on the Conceptualisation of Objects: Thesis (PhD Doctorate). Griffith University, Brisbane, 2014. 310 p.
- Резанова З.И., Некрасова Е.Д. Влияние грамматической категории рода на бимодальное восприятие имен существительных болгарского языка // *Русин*. 2015. № 3 (41). С. 241–255.
- Кирилина А.В. Гендерные стереотипы, общение и пол говорящего // *Женщина в российском обществе*. М., 1999. № 2. С. 27–45.
- Соловьева Н.С. Динамика гендерных стереотипов в английской и русской языковых картинах мира: на материале фразеологии : автореф. дис. ... канд. филол. наук. Волгоград, 2008.
- Каменева В.А. Гендерно обусловленные стереотипы в публицистическом дискурсе: на материале американской прессы : автореф. дис. ... канд. филол. наук. Кемерово, 2005.
- Базылев В.Н., Сорокин Ю.А. Феминилект и маскулинилект: модусы существования // *Пол и его маркировка в речевой деятельности* / под ред. Е.Н. Шовгеля. Кривой Рог : МИЦ ЧЯКЦ, 1996. С. 4–18.
- Вернер Ф. Речевое поведение женщин и мужчин // *Языкознание*. РЖ ИНИОН РАН. 1984. Сер. 6. С. 116–135.
- Земская Е.А., Китайгородская М.А., Розанова Н.Н. Особенности мужской и женской речи // *Русский язык и его функционирование*. М. : Наука, 1993. С. 90–136.
- Горошко Е.И. Особенности мужских и женских ассоциаций // *Пол и его маркировка в речевой деятельности* / под ред. Е.Н. Шовгеля. Кривой Рог : МИЦ ЧЯКЦ, 1996. С. 65–88.
- Резанова З.И., Некрасова Е.Д. Семантика диминутивных суффиксов в восприятии носителей русского языка: влияние контекстных и социальных факторов // *Вестник Томского государственного университета*. 2017. № 421. С. 12–21.
- Васильева А.В. Особенности когнитивной обработки диминутива мужчинами и женщинами: экспериментальное исследование // *Наука. Технологии. Инновации* : сб. науч. тр. в 9 ч. / под ред. ст. преп. О.Е. Цыганковой. Новосибирск : Изд-во НГТУ, 2015. Ч. 8. С. 121–123.
- Горошко Е.И. Особенности восприятия и порождения текста, обусловленные половой принадлежностью индивида // *Язык: Антропоцентризм и прагматика*. Москва ; Кривой Рог, 1995. С. 82–91.
- Котов А.Е. Гендерное своеобразие функционирования дискурсивных элементов в английском и русском языках: экспериментально-сопоставительное исследование на материале разностилевых устных текстов : автореф. дис. ... канд. филол. наук. Пенза, 2003.
- Антинескул О.Л. Гендер как параметр текстообразования : автореф. дис. ... канд. филол. наук. Пермь, 2000. 19 с.
- Васильева А.В. Коммуникативно-прагматические аспекты проявления экспрессивности в мужских и женских электронных сообщениях // *Вестник науки Сибири*. 2014. № 4 (14). С. 190–195
- Кирилина А.В. Гендер: лингвистические аспекты. М. : Ин-т социологии РАН, 1999. 189 с.
- Балакина Л.В. Проявление гендерного фактора в художественном тексте : автореф. дис. ... канд. филол. наук. Орел, 2005.
- Витлицкая Е.В. Лингвистическая репрезентация гендерных стереотипов в рекламе: на материале англоязычных и русскоязычных текстов : автореф. дис. ... канд. филол. наук. Волгоград, 2005.
- Спирюшкина Е.В. Проявление гендерного фактора в немецком языке : На материале публицистики : автореф. дис. ... канд. филол. наук. Н. Новгород, 2006.
- Верзун А.Б. Гендерная агональность политического дискурса : автореф. дис. ... канд. филол. наук. Волгоград, 2005.
- Захарова Т.Н. Семиотические средства выражения гендера в тексте на электронном носителе: на материале немецких чатов : автореф. дис. ... канд. филол. наук. М. : Моск. гос. лингвист. ун-т, 2006.

24. Вул С.М., Горошко Е.И. Судебно-автороведческая классификационная диагностика: установление половой принадлежности автора документа // Современные достижения науки и техники в борьбе с преступностью : матер. науч.-практ. конф. Минск, 1992. С. 139–141.
25. Ощепкова Е.С. Выявление идентификационных признаков мужской и женской письменной речи при искажении текстов // Теория та практика експертизи і криміналістики. Харків : Право, 2002. Вип. 2. 221–226
26. Романов А.С., Мещеряков Р.В. Определение пола автора короткого электронного сообщения // Диалог. 2011. С. 620–626.
27. Степаненко А.А. Гендерная атрибуция текстов компьютерной коммуникации: статистический анализ использования местоимений // Вестник Томского государственного университета. 2017. № 415. С. 17–25.
28. Резанова З.И., Романов А.С., Мещеряков Р.В. Задачи авторской атрибуции текстов в аспекте гендерной принадлежности (к проблеме междисциплинарного взаимодействия лингвистики и информатики) // Вестник Томского государственного университета. 2013. № 370. С. 24–28.
29. Вавилова Е.Н. Жанровая квалификация виртуального дискурса телеконференций Фидонет : автореф. дис. ... канд. филол. наук. Томск, 2001.
30. Алтухова Т.В., Лебедева Н.Б. Виртуальное общение: новый этап развития письменной коммуникации // Вестник Кемеровского государственного университета. 2012. № 1 (49). С. 105–111.
31. Картины русского мира: современный медиадискурс. Томск, 2011.
32. Карасик В.И. О типах дискурса // Языковая личность: институциональный и персональный дискурс : сб. науч. тр. Волгоград : Перемена, 2000. С. 5–20.
33. Алтухова Т.В. Социальная компьютерная сеть «ВКонтакте»: жанровая характеристика // Вестник Кемеровского государственного университета. 2012. № 4 (52), Т. 3. Филология. С. 21–25
34. Лингвистический энциклопедический словарь. М. : Сов. энциклопедия, 1990. 685 с.
35. Лукьянова Н.А. Экспрессивная лексика разговорного употребления: проблемы семантики. Новосибирск : Наука, Сиб. отд-ние, 1986. 230 с.
36. Бельская Е.В. Интенсивность как категория лексикологии (на материале говоров Среднего Приобья). автореф. дис. ... канд. филол. наук. Томск, 2002. 20 с.
37. Большой академический словарь русского языка / гл. ред. К.С. Горбачевич, А.С. Герд. М., СПб. : Наука, 2004. 571 с.
38. Никитина Т.Г. Так говорит молодежь. Словарь молодежного сленга. СПб., 1998. 592 с.
39. Макаров М.Л. Основы теории дискурса. М. : Гнозис, 2003. 280 с.
40. Deza M. Encyclopedia of Distances / Deza M., Deza E. // Springer Dordrecht Heidelberg. London; New York, 2009. 722 p.

Статья представлена научной редакцией «Филология» 15 апреля 2018 г.

#### EXPRESSIVENESS AS A MARKER OF GENDER DIFFERENCES IN COMPUTER COMMUNICATION (THE PROBLEM OF AUTOMATIC GENDER ATTRIBUTION OF THE TEXT)

*Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*, 2018, 433, 38–46.

DOI: 10.17223/15617793/433/5

**Andrei A. Stepanenko**, Tomsk State University (Tomsk, Russian Federation). E-mail: stepanekone@mail.ru

**Zoya I. Rezanova**, Tomsk State University (Tomsk, Russian Federation); Tomsk Polytechnic University (Tomsk, Russian Federation). E-mail: resso@rambler.ru; resso@mail.tsu.ru

**Keywords:** computer communication; gender attribution of text; expressiveness; text emotionality; statistical analysis.

The article solves the problems of gender attribution of the text, which is one of the most intensively developing directions in linguistics. This field of science uses methods of linguistics, logic, mathematics and computer science. The article describes the results of a statistical analysis of the use of expressive lexemes in the texts of computer communication. The purpose of the analysis was to test the hypothesis about the differences in their use depending on the gender of the communicants. The authors investigated the use of lexemes of four expressiveness types as markers of the gender invariant in the text: units containing (1) components of rational and emotional evaluation, (2) components of the feature intensity manifestation; (3) morphemic markers of emotionality and expressiveness, (4) graphic markers of emotionality and expressiveness. The material of the study was 19 dialogues of computer communication. The volume of each dialogue is ~ 150–200 Kb, ~ 4,000 words. In general, the volume of the analyzed texts is 120 pages, 80,300 words. Texts of computer communication were conducted using linguistic and mathematical methods. Methods of linguistic analysis of the text were applied at the first stage after preliminary technical processing of texts. The composition of markers of expressiveness and emotionality of texts was revealed using methods of distributive analysis of lexemes. At the second stage, quantitative qualification of the use of the detected markers in male and female replicas with the help of statistical analysis techniques was made, the statistical significance of the revealed quantitative differences in the use of these markers was revealed, and cluster analysis was carried out. On this basis, the assessment of the diagnostic strength of classes and individual units in marking gender specificities was given. As a result of the study, the authors conclude that there are differences in the use of expressive units in texts of computer communication, which are opposed to the gender of the authors of texts, but these differences are not statistically significant, that is, the sign of gender opposition is not stable and intelligible. The application of cluster analysis made it possible to identify the interaction of the gender sign of the authors of communication with other socially important parameters of computer communication – the topic of the text, the role and social positions of communicants in the dialogue.

#### REFERENCES

1. Boroditsky, L., Schmidt, L.A. & Phillips, W. (2003) Sex, syntax, and semantics. In: Gentner, D. & Goldin-Meadow, S. (eds) *Language in mind: Advances in the study of language and thought*. Cambridge: MIT.
2. Kurinski, E., Jambor, E. & Sera, M.D. (2016) Spanish grammatical gender: Its effects on categorization in native Hungarian speakers. *International Journal of Bilingualism*. 20(1). pp. 76–93. DOI: 10.1177/1367006915576833
3. Landor, R. (2014) *Grammatical Categories and Cognition across Five Languages: The Case of Grammatical Gender and its Potential Effects on the Conceptualisation of Objects*. Thesis (PhD Doctorate). Brisbane: Griffith University.
4. Rezanova, Z.I. & Nekrasova, E.D. (2015) The influence of grammatical gender on the bimodal perception of Bulgarian nouns. *Rusin*. 3 (41). pp. 241–255. (In Russian). DOI: 10.17223/18572685/41/17
5. Kirilina, A.V. (1999) Gendernye stereotipy, obshchenie i pol govoryashchego [Gender stereotypes, communication and gender of the speaker]. *Zhenshchina v rossiyskom obshchestve*. 2. pp. 27–45.
6. Solov'eva, N.S. (2008) *Dinamika gendernykh stereotipov v angliyskoy i russkoy yazykovykh kartinakh mira: na materiale frazeologii* [Dynamics of gender stereotypes in English and Russian language pictures of the world: on the material of phraseology]. Abstract of Philology Cand. Dis. Volgograd.



7. Kameneva, V.A. (2005) *Genderno obuslovlennyye stereotypy v publitsisticheskom diskurse: na materiale amerikanskoy pressy* [Gender-related stereotypes in journalistic discourse: on the material of the American press]. Abstract of Philology Cand. Dis. Kemerovo.
8. Bazylev, V.N. & Sorokin, Yu.A. (1996) Feminolekt i maskulinolekt: modusy sushchestvovaniya [Feminolect and masculinolect: modes of existence]. In: Shovgel, E.N. (ed.) *Pol i ego markirovka v rechevoy deyatel'nosti* [Gender and its marking in speech activity]. Krivoy Rog: MITs ChYaKTs.
9. Verner, F. (1984) Rechevoe povedenie zhenshchin i muzhchin [Speech behavior of women and men]. *Yazykoznanie. RZh INION RAN. Ser. 6*. pp. 116–135.
10. Zemskaya, E.A., Kitaygorodskaya, M.A. & Rozanova, N.N. (1993) Osobennosti muzhskoy i zhenskoy rechi [Features of male and female speech]. In: Zemskaya, E.A. & Shmelev, D.N. (eds) *Russkiy yazyk i ego funktsionirovanie* [Russian language and its functioning]. Moscow: Nauka.
11. Goroshko, E.I. (1996) Osobennosti muzhskikh i zhenskikh assotsiatsiy [Features of male and female associations]. In: Shovgel, E.N. (ed.) *Pol i ego markirovka v rechevoy deyatel'nosti* [Gender and its marking in speech activity]. Krivoy Rog: MITs ChYaKTs.
12. Rezanova, Z.I. & Nekrasova, E.D. (2017) Semantics of diminutive suffixes in the perception of native speakers of the Russian language: influence of contextual and social factors. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 421. pp. 12–21. (In Russian). DOI: 10.17223/15617793/421/2
13. Vasil'eva, A.V. (2015) Osobennosti kognitivnoy obrabotki diminutiva muzhchinami i zhenshchinami: eksperimental'noe issledovanie [Features of cognitive processing of the diminutive by men and women: an experimental study]. In: Tsygankova, O.E. (ed.) *Nauka. Tekhnologii. Innovatsii* [Science. Technologies. Innovations]. In 9 parts. Part 8. Novosibirsk: Novosibirsk State Technical University.
14. Goroshko, E.I. (1995) Osobennosti vospriyatiya i porozhdeniya teksta, obuslovlennyye polovoy prinadlezhen'st'yu individa [Peculiarities of perception and generation of the text conditioned by the individual's gender]. In: Kholod, A.M. (ed.) *Yazyk: antropotsentrizm i pragmatika* [Language: anthropocentrism and pragmatics]. Kiev, Moscow; Krivoy Rog: MITs ChYaKP.
15. Kotov, A.E. (2003) *Gendernoe svoeobrazie funktsionirovaniya diskursivnykh elementov v angliyskom i russkom yazykakh: eksperimental'no-sopostavitel'noe issledovanie na materiale raznostilevykh ustnykh tekstov* [Gender features of the functioning of discursive elements in English and Russian: an experimental comparative study on the material of oral texts of different styles]. Abstract of Philology Cand. Dis. Pyatigorsk.
16. Antineskul, O.L. (2000) *Gender kak parametr tekstoobrazovaniya* [Gender as a parameter of text formation]. Abstract of Philology Cand. Dis. Perm.
17. Vasil'eva, A.V. (2014) Kommunikativno-pragmaticheskie aspekty proyavleniya ekspressivnosti v muzhskikh i zhenskikh korotkikh elektronnykh soobscheniyakh [Communicative and pragmatic aspects of expressivity in male and female short text messages]. *Vestnik nauki Sibiri – Siberian Journal of Science*. 4 (14). pp. 190–195
18. Kirilina, A.V. (1999) *Gender: lingvisticheskie aspekty* [Gender: the linguistic aspects]. Moscow: Institute of Sociology RAS.
19. Balakina, L.V. (2005) *Proyavlenie gendernogo faktora v khudozhestvennom tekste* [Manifestation of the gender factor in the literary text]. Abstract of Philology Cand. Dis. Orel.
20. Vitlitskaya, E.V. (2005) *Lingvisticheskaya reprezentatsiya gendernykh stereotipov v reklame: na materiale angloyazychnykh i russkoyazychnykh tekstov* [Linguistic representation of gender stereotypes in advertising: on the material of English and Russian texts]. Abstract of Philology Cand. Dis. Volgograd.
21. Spiryushkina, E.V. (2006) *Proyavlenie gendernogo faktora v nemetskom yazyke: Na materiale publitsistiki* [Manifestation of the gender factor in the German language: On the material of journalism]. Abstract of Philology Cand. Dis. Nizhny Novgorod.
22. Verzun, A.B. (2005) *Gendernaya agonal'nost' politicheskogo diskursa* [Gender agonality of political discourse]. Abstract of Philology Cand. Dis. Volgograd.
23. Zakharova, T.N. (2006) *Semioticheskie sredstva vyrazheniya gendera v tekste na elektronnom nositele: na materiale nemetskikh chatov* [Semiotic means of expressing gender in the text on an electronic medium: on the material of German chats]. Abstract of Philology Cand. Dis. Moscow.
24. Vul, S.M. & Goroshko, E.I. (1992) [Court authorship classification diagnostics: establishing the gender of the author of the document]. *Sovremennyye dostizheniya nauki i tekhniki v bor'be s prestupnost'yu* [Modern achievements of science and technology in the fight against crime]. Proceedings of the Conference. Minsk: NIIPKK i sudebnykh ekspertiz. pp. 139–141. (In Russian).
25. Oshchepkova, E.S. (2002) Vyyavlenie identifikatsionnykh priznakov muzhskoy i zhenskoy pis'mennoy rechi pri iskazhenii tekstov [Determination of identification signs of male and female written speech in distorted texts]. *Teoriya ta praktika ekspertizi i kriminalistiki*. 2. pp. 221–226
26. Romanov, A.S. & Meshcheryakov, R.V. (2011) Gender identification of the author of a short message. *Dialog-21*. pp. 620–626. (In Russian).
27. Stepanenko, A.A. (2017) Gender attribution in social network communication: the statistical analysis of pronouns frequency. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 415. pp. 17–25. (In Russian). DOI: 10.17223/15617793/415/3
28. Rezanova, Z.I., Romanov, A.S. & Meshcheryakov, R.V. (2013) Tasks of author attribution of text in the aspect of gender (on interdisciplinary interaction of linguistics and computer science). *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 370. pp. 24–28. (In Russian).
29. Vavilova, E.N. (2001) *Zhanrovaya kvalifikatsiya virtual'nogo diskursa telekonferentsiy Fidonet* [Genre qualification of the virtual discourse of Fidonet teleconferences]. Abstract of Philology Cand. Dis. Tomsk.
30. Altukhova, T.V. & Lebedeva, N.B. (2012) Virtual communication: a new stage of writing communication. *Vestnik Kemerovskogo gosudarstvennogo universiteta – Bulletin of Kemerovo State University*. 1 (49). pp. 105–111. (In Russian).
31. Rezanova, Z.I. (ed.) (2011) *Kartiny russkogo mira: sovremennyy mediadiskurs* [Images of the Russian world: modern media discourse]. Tomsk: ID SK.
32. Karasik, V.I. (2000) O tipakh diskursa [On the types of discourse]. In: Karasik, V.I. & Slyshkin, G.G. (eds) *Yazykovaya lichnost': institutsional'nyy i personal'nyy diskurs* [The language personality: institutional and personal discourse]. Volgograd: Peremena.
33. Altukhova, T.V. (2012) Social computer network VKontakte: genre characterization. *Vestnik Kemerovskogo gosudarstvennogo universiteta – Bulletin of Kemerovo State University*. 4 (52):3. pp. 21–25. (In Russian).
34. Yartseva, V.N. (ed.) (1990) *Lingvisticheskiy entsiklopedicheskiy slovar'* [Linguistic encyclopedic dictionary]. Moscow: Sov. entsiklopediya.
35. Luk'yanova, N.A. (1986) *Ekspressivnaya leksika razgovornogo upotrebleniya: problemy semantiki* [Expressive colloquial vocabulary: problems of semantics]. Novosibirsk: Nauka, Sib. otd-nie.
36. Bel'skaya, E.V. (2002) *Intensivnost' kak kategoriya leksikologii (na materiale govorov Srednego Priob'ya)* [Intensity as a category of lexicology (on the material of dialects of the Middle Ob region)]. Abstract of Philology Cand. Dis. Tomsk.
37. Gorbachevich, K.S. & Gerd, A.S. (eds) (2004) *Bol'shoy akademicheskiy slovar' russkogo yazyka* [The big academic dictionary of the Russian language]. St. Petersburg: Nauka.
38. Nikitina, T.G. (1998) *Tak govorit molodezh'. Slovar' molodezhnogo slenga* [This is what the youth say. Dictionary of youth slang]. St. Petersburg: Folio-Press.
39. Makarov, M.L. (2003) *Osnovy teorii diskursa* [Fundamentals of discourse theory]. Moscow: Gnozis.
40. Deza, M. & Deza, E. (2009) *Encyclopedia of Distances*. London; New York: Springer-Verlag Berlin Heidelberg.

Received: 15 April 2018