

А.В. Колмогорова

ЭМОЦИОНАЛЬНО НЕЙТРАЛЬНЫЕ ИНТЕРНЕТ-ТЕКСТЫ: В ПОИСКАХ СПЕЦИФИКИ

Исследование выполнено при поддержке гранта РФФИ (проект «Разработка классификатора русскоязычных интернет-текстов по критерию их тональности на основе модели эмоций “Куб Лёвхейма”» № 19-012-00205).

Рассматривается специфика эмоционально нейтральных интернет-текстов. По результатам анализа двух подкорпусов – вики-текстов и текстов из паблика «Подслушано» в ВКонтакте, размеченных асессорами как нейтральные, делается вывод о том, что в интернет-коммуникации категория эмоциональной нейтральности представляет собой континуум, верхнюю границу которого очерчивают информационно-справочные тексты, а нижнюю – тексты жанра «интернет-откровение». Однако и для тех, и для других существует ряд общих вербальных «маркеров нейтральности».

Ключевые слова: интернет-тексты; sentiment-анализ; эмоционально нейтральные тексты; корпусные методы в лингвистике; вики-тексты.

1. Введение

Современная коммуникация в интернете характеризуется в исследовательской литературе как дискурсивное пространство новой чувствительности [1], которая призвана освободить человека цифровой эры от переживаний, связанных с насильем и эксплуатацией, вернув ему интерес к собственным инстинктивным экзистенциальным ощущениям [2. Р. 54]. Однако слоган «ты то, чем ты делишься» [3. С. 46], став основой идеологии новых медиа, породил волну так называемого киберэкзбиционизма [4], в том числе и эмоционального. Гипертрофированная резкость оценок и откровенность эмоций, став неременной характеристикой интернет-коммуникации, спровоцировали желание социальных стейкхолдеров параметризовать чувства и эмоции в интернете в целях их последующего мониторинга. Возникла исследовательская парадигма sentiment-анализа [5], а затем – эмоционального анализа текстовых данных [6].

Выполняемый исследовательской группой проект находится в рамках последнего из упомянутых направлений. Цель проекта – разработка компьютерного классификатора, способного автоматически определять ведущую эмоцию, вербализованную в интернет-тексте на русском языке.

За основу была взята восьмичленная классификация эмоций, предложенная шведским исследователем Г. Лёвхеймом [7], так называемый «куб Лёвхейма». Семь эмоций в данной классификации имеют двойную номинацию, где первая часть является обозначением слабой степени выраженности эмоционального состояния, а вторая – самой сильной (исключение составляет эмоция удивления): Интерес / Возбуждение; Удовольствие / Радость; Удивление; Страдание / Тоска; Гнев / Ярость; Страх / Ужас; Презрение / Отвращение; Стыд / Унижение.

На первом этапе проекта определены морфологические, синтаксические, лексические и семантические группы вербальных маркеров текстов, принадлежащих восьми эмоциональным классам [8]. Однако опыт организации эмоциональной разметки корпуса

текстов с помощью группы асессоров, а также тестирование пилотной версии классификатора продемонстрировали необходимость создания 9-го класса текстов – нейтрального.

Серия исследовательских вопросов, раскрываемых в той части исследования, результаты которой будут изложены в данной публикации, сформулирована следующим образом: имеют ли нейтральные тексты свою собственную языковую специфику относительно эмоциональных текстов? зависит ли эта специфика от жанра интернет-текста? насколько «далеко» отстоят тексты, оцениваемые информантами как нейтральные, от тех, которые ими же классифицируются как эмоциональные?

В попытке ответить на поставленные вопросы мы последовательно рассмотрим материал и методы исследования (п. 2), три аспекта категории нейтральности, нашедшие отражение в лингвистических штудиях, – нейтральность стилистическая, оценочная и эмоциональная (п. 3); проведем корпусный лингвистический анализ двух подкорпусов нейтральных текстов разной жанровой принадлежности и сравним полученные результаты с теми, что были получены при анализе эмоционально окрашенных текстов (п. 4); наконец, сделаем вывод о наличии / отсутствии собственных дискриминантных черт у класса нейтральных текстов, зависимости этих черт от жанра текстов и о специфике категории нейтральности в интернет-коммуникации.

2. Материал и методы

Цель данной публикации – описать результаты исследования, направленного на выявление вербальных маркеров эмоционально нейтральных текстов.

Материалом исследования послужили три корпуса интернет-текстов: тексты из Википедии, рассматриваемые а priori как эмоционально нейтральные и составившие корпус данных в объеме 136 184 токенов (далее – НТ а priori); тексты из пабликов «Карамель», «Подслушано», «Палата № 6» русскоязычной социальной сети «ВКонтакте», размеченные асессорами

как нейтральные в процессе эмоциональной разметки датасета эмоциональных текстов – корпус а posteriori нейтральных текстов (далее – НТ а posteriori); наконец, 8 подкорпусов эмоциональных текстов (далее – ЭТ) из указанных выше пабликов в объеме 1 092 327 токенов, которые служат фоном для контрастивного описания специфики нейтральных текстов.

Выбор Википедии в качестве источника данных для класса НТ а priori обусловлен принадлежностью текстов к информационному дискурсу, предполагающему стремление авторов текстов избегать оценочности. Так, по словам В.В. Дементьева, «Википедия старается избегать давать оценки – любые» [9. С. 241]. Кроме того, исследователи отмечают, что в Википедии образовалась критическая среда для развития коллективного мышления постнеклассического идеала рациональности [10].

Нейтральные тексты а posteriori были отобраны по итогам процедуры ассессмента, проведенной на одной из краудсорсинговых платформ. Зарегистрированным на данной платформе 2 тыс. информантам было предложено разметить 3 920 текстовых фрагментов объемом от 80 до 120 слов, т.е. по 490 образцов из восьми эмоциональных классов текстов, согласно восьми базовым эмоциям, выделенным Г. Лёвхеймом.

Упомянутые 8 классов текстов были первоначально сформированы путем сплошной выборки постов со страниц проектов «Подслушано», «Карамель», «Палата № 6» русскоязычной социальной сети ВКонтакте. Выборка производилась по хештегам, содержащим вербализованную отсылку к соответствующему эмоциональному состоянию (например, #_фуу – эмоция отвращения или #Грустненько) или номинацию ситуации, обычно вызывающей определенную эмоцию (#Предательство или #Одиночество). Общий объем выборки составил 15 тыс. текстов, из которых затем случайным образом были отобраны уже упомянутые 3 920 текстов.

В процессе аннотирования информанты должны были разметить тексты, пользуясь шкалой и инструкцией.

Инструкция: Внимательно прочитайте текст. Если необходимо, прочитайте его несколько раз. Какие эмоции автор выражает в тексте? На каждой из шкал поставьте отметку ближе к той эмоции, которая сильнее выражена в тексте. Поставьте отметку настолько близко, насколько очевидна и сильна эта эмоция в тексте. Например, один шаг от центра – оттенок эмоции присутствует, но выражен слабо; три шага – если эмоция явно присутствует; пять шагов – если эмоция, без сомнения, доминирует. Если в тексте нет эмоций, обозначенных на шкале, оставьте отметку в среднем положении.

Далее информантам предлагалось 4 шкалы, между полюсами которых можно было ставить отметку-маркер: Грусть – Радость, Гнев – Отвращение, Стыд – Оживление, Страх – Удивление. Такое конструирование шкал было обусловлено моделью эмоций «Куб Лёвхейма», согласно которой данные эмоции на биохимическом уровне формируют оппозиции. Так, например, Грусть «запускается» при условии низкого уровня гормонов серотонина и дофамина в крови че-

ловека, в то время как Радость иницируется высоким уровнем этих нейротрансмитеров.

В результате проведенной аннотации мы получили ограниченный 412 фрагментами, но, тем не менее, показательный корпус текстов, размеченный информантами как «нейтральные тексты», объемом 14 тыс. токенов.

Подчеркнем, что в предварительную выборку данные тексты были включены как «эмоциональные», но информанты расценили их как нейтральные.

Для анализа мы намеренно взяли два корпуса интернет-текстов разной жанровой принадлежности: а priori нейтральные тексты из Википедии и а posteriori нейтральные тексты из паблика «Подслушано». Первые репрезентируют информационный жанр «справочная информация», а вторые – жанр «интернет-откровение» [11]. Предполагается, что необходимо выделить, с одной стороны, такие дискриминантные черты нейтральных текстов, которые не зависят от жанра, а с другой – посмотреть, зависят ли вообще маркеры нейтральности от жанра.

В качестве ведущего используется метод корпусного лингвистического анализа, предполагающий исследование коллокаций, ключевых слов, конструкций, окружения исследуемой единицы, а также использование квантитативных методов [12. С. 20]. Технологический базис исследования составил корпусный менеджер Sketch Engine – поисковая система для работы с данными корпуса, получения статистической информации и предоставления пользователю результатов в удобной форме [13. С. 141]. Помимо множества функций, статистических методов, поддержки большого количества языков и удобного интерфейса, корпусные менеджеры позволяют исследователю работать с большими собственными коллекциями данных, которые предварительно индексируют для ускорения процесса поиска.

Объектом анализа является категория эмоционально нейтральных текстов, а предметом – их дискриминантные черты. Последние понимаются как такие доступные параметризации характеристики группы текстов, которые отличают ее от других групп текстов. В нашем проекте дискриминантные черты текстов каждого из эмоциональных классов используются в качестве дополнительных факторов для формирования функций-предикторов в процессе машинного обучения по прецедентам, которое проходит компьютерный классификатор текстов на основе размеченной выборки.

3. О понятии нейтральности в лингвистике

Понятие нейтральности в лингвистике далеко неоднозначно. В целом следует констатировать, что для лингвистов категория нейтральности (будь то в стилистике, семантике или этиологии) практически всегда является точкой отсчета для описания какого-то иного объекта, качества или характеристики: стилистического значения лексемы, степени экспрессивности или оценочности конструкции, выраженности эмоционального концепта. Однако практически никогда нейтральность не рассматривалась как самостоя-

тельный объект исследования. Тем не менее есть определенная доля истины в известном афоризме – «чтобы быть эксцентричным, надо знать, где находится центр». Рассмотрим специфику трактовки категории нейтральности в трех аспектах: стилистическая нейтральность, оценочная нейтральность и эмоциональная нейтральность.

3.1. Стилистическая нейтральность

В традициях стилистики нейтральность текста или составляющих его единиц трактуется в самом общем смысле как их «невыведенность», создающая фон для появления у некоторой единицы или совокупности единиц некоторой стилистической окраски или стилистического значения.

Так, М.В. Панов отмечает [14. С. 179], что только окрашенные единицы создают стилистическую характерность текстов, – «нейтральные же единицы, встречающиеся в любом тексте (курсив наш. – А.К.), не могут быть носителями характерности», однако первые проявляют свою характерность, только находясь в сочетании со вторыми.

Р.Г. Пиотровский сужает границы нейтральных текстов, специфицируя их невыведенность новым ограничением – это должны быть некие типовые, стереотипные контексты, где лексические единицы встречаются в своем привычном окружении, поскольку «даже наиболее стилистически нейтральные слова, словосочетания и грамматические конструкции при перенесении их в другую речевую среду становятся стилистически окрашенными» [15. С. 22].

Эти «обычность» и «невыведенность» являются в том числе следствием соблюдения в нейтральных текстах базовых норм языка, предопределяющих отсутствие каких бы то ни было препятствий к их пониманию [16. С. 275]. И наоборот, всякое нарушение норм языка, согласно стилистике отклонений [17], трактуется его носителем – реципиентом некоторого текста как «отклонение», уже этим привлекает к себе внимание читателя, становясь «выделенным», а значит, приобретая стилистическую окраску.

Невыведенность, типизированность контекста и стереотипность ситуации употребления, нормативность, отсутствие проблем понимания – основные признаки нейтрального текста, рассмотренного в стилистической перспективе.

3.2. Оценочная нейтральность

Оценка, по мнению Н.Д. Арутюновой, является результатом соотнесения человеком мира реального и его идеализированной самим же человеком модели [18. С. 182–183], что порождает «разнообразные иллюкативные силы» – «оценочное высказывание стремится повлиять на адресата, а через него и на ход практической жизни». Следовательно, оценка находится в зоне прагматики языкового знака. Рассматривая ее в лексикографическом аспекте, Ю.Д. Апресян [19. С. 66–69] выделял два типа информации, составляющей данную зону: 1) культурные коннотации и семантические ассоциации, фиксирующие отношение

национально-лингво-культурного коллектива к обозначаемой словом сущности; 2) «модальную рамку», отражающую оценку описываемой конкретной ситуации говорящим или слушающим.

И.В. Арнольд связывает оба вида прагматической информации с понятием потенциальных или импликационных сем [20], которые могут быть актуализированы в контексте, но могут и оставаться в латентном состоянии. Во втором случае лексема может, по видимому, приниматься за нейтрально-оценочную.

Таким образом, к нейтрально-оценочным языковым единицам принято относить лексемы, либо не имеющие в данном национально-лингво-культурном коллективе специфических культурных коннотаций, либо употребленные в таком контексте, который не актуализирует эти семы, а также в случае, если в контексте данные лексемы не имеют специфической модальной рамки, позволяющей увидеть через нее индивидуальное восприятие говорящим / слушающим описываемой ситуации.

Соответственно, нейтрально-оценочные тексты – это тексты, которые не создают контекстных условий для актуализации какой бы то ни было оценки (ни социальной, ни индивидуальной) описываемой в них ситуации, минимизируя тем самым прагматическую направленность речевого произведения.

3.3. Эмоциональная нейтральность

Наконец, третий аспект – собственно эмоциональная нейтральность. Эмоции долгое время не включались в число объектов лингвистического анализа, поскольку с «легкой руки» Э. Сепира считалось, что, хотя языковые символы и соотносились, возможно, когда-то с эмоциями, сейчас эта связь «уже непосредственно не просматривается» [21. С. 263]. Однако, когда появилась отдельная отрасль лингвистики – эмотиология, сконцентрировавшаяся на изучении проявления эмоциональности в языке, обнаружилась противоположная тенденция – стали высказываться мнения о том, что «весь язык эмотивен», что «не существует эмотивно нейтральной лексики» [22. С. 25].

Тем не менее о возможности существования эмоционально нейтральных текстов писал уже Ш. Балли в начале XX в. Он отмечал, что во всяком акте речевого производства говорящий «придает своим мыслям либо объективную, рассудочную форму, максимально соответствующую действительности, либо чаще всего вкладывает в выражение в самых различных дозах эмоциональные элементы» – индивидуальные чувства, часто претерпевающие влияние чувств социальных [23. С. 29].

В языковых сообществах сложились определенные жанровые формы, которые тяготеют к «объективации интеллектуальной сферы духовной жизни человека» [23. С. 29]. Это в первую очередь жанры информативного дискурса [24]: научный текст (монография, статья, доклад), информационно-справочные тексты (словари, энциклопедии, справочники).

С распространением интернет-дискурса корпусы текстов с так называемым нейтральным сентиментом (тональностью) также формируются и на основе по-

стов в социальных сетях, отзывов о товаре или услуге на интернет-платформах. Это происходит в рамках достаточно новых парадигм анализа, использующих технологический потенциал компьютерной и математической лингвистики.

Первое направление известно как извлечение мнений (opinion mining) [25, 26]. Оно также включает анализ тональности (sentiment analysis), где под тональностью понимается эмоционально-оценочное отношение к объекту оценки (товару или услуге), выраженное в тексте [27, 28]. Оно может быть позитивным, негативным или нейтральным. Текст считается нейтрально окрашенным, если он не содержит эмоциональной составляющей [29. С. 152]. Тональность в рамках такого подхода определяется либо по словарям тональностей, либо на основе вероятностных статистических моделей, выстраиваемых при помощи технологии машинного обучения. Если используется технология машинного обучения по прецедентам, то для тренировки алгоритма привлекается обучающая выборка, в которой каждый текстовый фрагмент имеет «эмоциональную этикетку», приписанную ему информантами-ассессорами.

Второе направление – эмоциональный анализ текстов. Его объектом является уже конкретная эмоция, вербализованная в тексте: страх, радость, гнев и др. Но в качестве нейтральных, как и в первом случае, рассматриваются тексты, не содержащие эмоциональной составляющей.

Источником знаний, позволяющих квалифицировать некоторые тексты как эмоционально нейтральные, как в первом, так и во втором случае выступают статистические модели, выстраиваемые компьютерной системой самостоятельно (машинное обучение без учителя), либо с опорой на разметку, предварительно проведенную информантами (машинное обучение с учителем). Кроме того, существенно повысить точность классификации может подача «на вход» алгоритма

неких дискриминантных черт (в англоязычной терминологии – features) класса нейтральных текстов, выделяющих его среди других и способных выступить в качестве основы функций-предикторов, определяющих работу программы-классификатора.

Проведенный анализ теоретической литературы позволил сформулировать рабочую гипотезу, предопределяющую алгоритм последующей работы по поиску дискриминантных черт класса нейтральных текстов: вербальная специфика НТ может быть маркирована 1) соответствием стилистической норме (отсутствие обсценизмов); 2) невыраженностью модальной рамки высказывания, что скажется на функционировании маркеров чужого слова, функционировании слов-дейктиков, а также частотных предикатах; 3) отсутствием или статистической нерелевантностью «вокабуляра эмоций».

4. Корпусный анализ нейтральных выборок на фоне эмоциональной выборки

4.1. Сниженная и оценочная лексика

Проведенный анализ частотности четырех наиболее употребительных обсценизмов в трех подкорпусах показал отсутствие подобных лексем (что в вики-текстах ожидаемо), поскольку у ресурса есть команда модераторов с правом блокировки пользователей и удаления записей, к тому же это публичный ресурс.

Однако результаты анализа подкорпуса текстов из социальной сети ВКонтакте, оцененных информантами как нейтральные, продемонстрировали, что, когда речь идет об интернет-текстах, несоответствие стилистической норме не означает автоматическую не-нейтральность текста – в данном подкорпусе обсценизмы есть, хотя значения их частотности в большинстве случаев ниже, чем в эмоциональных текстах. В табл. 1 указана частотность четырех наиболее распространенных обсценизмов.

Таблица 1

Относительная частотность обсценизмов в трех подкорпусах

Подкорпус / частотность наречия в ipm (items per million)	Обсценизм № 1	Обсценизм № 2	Обсценизм № 3	Обсценизм № 4
НТ a priori	0	0	0	0
НТ a posteriori	260	0	43	130
ЭТ	708	68	72	79

Что касается оценочных прилагательных, то в каждом подкорпусе их доля в списке 50 наиболее частотных прилагательных также отличается: НТ a posteriori – 18 из 50 (*красивый, счастливый, нужный, хороший, огромный, страшный, крутой, любимый* и т.д.), НТ a priori – 3 из 50 (*хороший, настоящий, основной*) и ЭТ – 24 из 50 (*хороший, счастливый, отвлеченный, прекрасный, уютный*, и т.д.).

4.2. Способы передачи «чужого слова»

Первая категория вербальных маркеров, верифицированная на материале НТ, – это категория ксенопоказателей. Термин был введен Н.Д. Арутюновой [30] для

обозначения частиц *де*, *мол*, *дескать* как знаков чужого голоса, отчуждаемой речи, чужого мира. Воспроизведение говорящим речи «другого», несущее груз двойной субъективности, выполняет функцию нивелирования социальной ответственности говорящего за передаваемое сообщение в его диктумном или модусном аспекте, что позволяет отнести анализируемые частицы к категории показателей «субъективной цитации» [31].

Как показал анализ, *де* и *дескать* не используются ни в ЭТ, ни НТ, по-видимому, тому виной оттенок книжности, который для них характерен. Что касается ксенопоказателя *мол*, то если в восьми эмоциональных подкорпусах он входит в число частотных слу-

жебных слов в диапазоне рангов от 13 до 64, то в обоих нейтральных подкорпусах (НТ а posteriori и а priori) нет ни одного употребления данной единицы.

Уже имеющиеся у коллектива результаты свидетельствуют [32] о том, что ЭТ чувствительны к ксенопоказателям. По-видимому, эмоциональность часто объективируется путем апелляции к чужому слову, а нейтральность – ее отсутствием.

4.3. Глагольность и бытийность

Сравнение частотности глагольных форм в девяти подкорпусах показало, что в корпусах НТ значительно меньше глаголов, чем в ЭТ (табл. 2).

Таблица 2

Значения относительной частотности глаголов
в подкорпусах ЭТ и НТ

Подкорпус	ЭТ	НТ а posteriori	НТ а priori
Частоты в ipm	от 137 613 до 160 123	101 907	84 061

Среди глаголов обоих подкорпусов самым частотным является глагол *быть*.

Предварительный интерпретативный анализ корпусов НТ а posteriori позволил сформулировать предположение о том, что внимание нарратора в данных текстах сосредоточено, по-видимому, не на своем эмоциональном состоянии, а на некоторой наблюдаемой ситуации, в которой третьи лица, возможно, испытывают некоторые эмоции, но нарратор о них может только догадываться и эмпатично им сопереживать, однако это не его личные чувства. В [33] подобное явление названо эмотивной тональностью объектного типа. Глагол *быть* в нейтральных текстах в «Подслушано» используется в имперфективном и перфективном стативном употреблении [34]. Сравните примеры текстов (орфография и пунктуация пользователей ВКонтакте сохранены).

1. Когда мне было лет 6 мы жили в р-не рядом с роцей. У нас во дворе часто гуляла пожилая женщина со взрослым сыном. Это сейчас я понимаю, что он был умственно-отсталым, а тогда всем детям во дворе он казался просто интересным и странным взрослым и конечно мы не считали его опасным. Как-то его заметили гуляющим уже без матери, а потом перестали встречать на улице его вовсе. И вот однажды наша ровесница пропала. Ее не могли найти целый день. А вечером кто-то увидел этого парня, бредущего за руку с пропавшей малышкой в сторону роцы. Оказалось, что целый день они были в его квартире где он «знакомил» ее со своей мамой, которая лежала мертвая на кровати уже 3 недели. Умерла она своей смертью. Наверное он искал замену, чтобы было с кем гулять (корпус НТ а posteriori).

Нарратор никак не эксплицирует свое эмоциональное состояние, хотя тот факт, что процитированный пост был изначально помещен редакторами публика под хештегом #Одиночество, имплицитно эмоцию грусти, эмпатически, возможно, переживаемую автором поста, но не замеченную менее эмпатичным ассессором, оценившим его как НТ.

В нейтральных же вики-текстах (НТ а priori) глагол *быть* используется преимущественно в конструкциях с глаголами в пассивном залоге для констатации фактивной информации, например:

2. В 1922 г. уезды преобразовали в округа. 11 октября 1923 г., согласно постановлению ВЦИК, в административное деление Крымской АССР были внесены изменения, в результате которых округа были ликвидированы, основной административной единицей стал Джанкойский район и село включили в его состав (НТ а priori).

4.4. Дейктики

В подкорпусах ЭТ и НТ проведен анализ значений меры семантической близости МІ (коэффициент взаимной информации) глагола *быть* и четырех дейктиков: *я, он, она, они*. Мера МІ относится к точечным оценкам силы ассоциации. Она сравнивает зависимые контекстно-связанные частоты с независимыми (при случайном появлении слов в контексте) по формуле

$$MI(n, c) = \log_2 \frac{f(n, c)}{f(n) \cdot f(c)},$$

где n – ключевое слово; c – коллокат; $f(n, c)$ – частота встречаемости ключевого слова n в паре с коллокатом c ; $f(n)$, $f(c)$ – абсолютные (независимые) частоты ключевого слова n и коллоката c в корпусе (тексте) [35].

Оказалось, что значение МІ для *быть* и указанных дейктиков в НТ а priori нулевое, это означает, что их совместная встречаемость статистически не значима, она случайна. Кроме того, частотность каждого из указанных местоимений в сочетании с *быть* не превышает 200 ipm, что является чрезвычайно низким показателем по сравнению с другими подкорпусами.

В подкорпусе НТ а posteriori ситуация кардинально иная.

На рис. 1 показаны в сравнении значения МІ по восьми подкорпусам ЭТ и НТ а posteriori: значения данного коэффициента выше всего в подкорпусе НТ а posteriori, но если для *я+быть* этот рост незначителен, то для *он+быть* и *они+быть* скачок значений более выражен. Иначе говоря, «неслучайность» сочетания *быть* с местоимениями 3-го лица в корпусе НТ выше, чем в подкорпусах ЭТ.

Таким образом, наблюдается значимое расхождение в функционировании местоимений-дейктиков в связке с глаголом *быть* в корпусах НТ а posteriori и а priori: в одном они тесно и статистически неслучайно связаны, в другом – практически не употребляются вместе, хотя сам глагол одинаково частотен в обоих подкорпусах. Наблюдаемый контраст можно объяснить тем, что в случае с НТ а posteriori нарратор является одновременно наблюдателем ситуации, происходящей с третьими лицами, – наблюдение требует достаточно длительной фиксации внимания на объекте. Это и влечет за собой, во-первых, употребление *быть* в его имперфективном или перфективном стативном значениях, во-вторых, использование местоименной репризы во избежание повторов (см. текст 1): *он был умственно-отсталым; целый день они были в его квартире* и притяжательных конструкций (*в его квартире*).

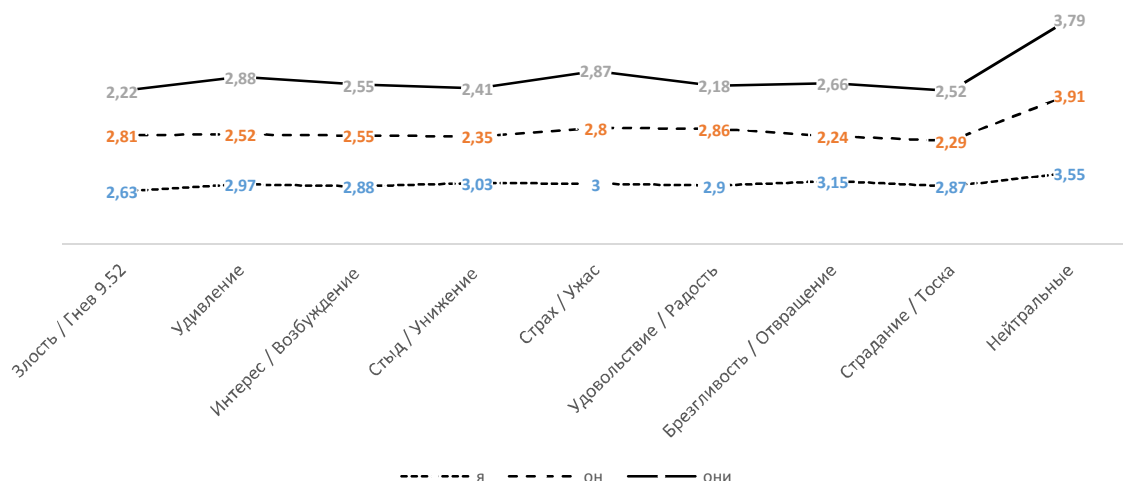


Рис. 1. Сопоставление значений MI лексемы *быть* и коллокатов *я, он, они* в левом контексте в девяти подкорпусах текстов

В корпусе же НТ а priori нарратор констатирует факты либо описывает ситуации «с чужих слов», поэтому глагол *быть* в вики-текстах используется в качестве служебного для формирования пассивного залога, который, в свою очередь, просто фиксирует глагольный признак, направленный на семантический объект (см. текст 2: *округа были ликвидированы*).

Другая категория дейктиков – наречия времени. В табл. 3 показана относительная частотность группы наречий времени, передающих «присутствие» говорящего-наблюдателя: *однажды, вчера, вдруг, потом, сразу*.

Прослеживается тенденция наименее значительного их присутствия в вики-текстах (НТ а priori), большей частотности – в НТ а posteriori и максимальной – в ЭТ.

Таблица 3

Относительная частотность наречий времени в трех подкорпусах текстов

Подкорпус / частотность наречия в ipm	Однажды	Вчера	Вдруг	Потом	Сразу
НТ а priori	5	0	10	47	120
НТ а posteriori	390	86	216	267	260
ЭТ	105	92	264	953	490

Исключение составляет наречие *однажды*, частотность его выше всего в НТ а posteriori (текст 3). Вероятно, это связано с тем, что *однажды* – типичный зачин истории, рассказываемой нарратором-наблюдателем, не являющимся экспериенцером чувства:

2. *Однажды катались на аттракционах с дочкой) так вот, сели на детские, такие американские горки, а за нами мамочка с дочкой, так пристегнуться из-за размеров и не смогли! Мамочка позже пыталась пересест в другую кабинку, но даже находясь одна в ней, не смогла из-за живота опустить ограничитель)) в итоге вышла! Дамы ну вы хоть мужей своих не стыдите) а то он бедный рядом стоял, смотрел на все это... И задумался наверно...* (НТ а posteriori).

4.5. Присутствие «вокабуляра эмоций»

Корпусный менеджер Sketch Engine располагает функцией выявления ключевых биграмм (key bigrams)

одного корпуса на фоне другого (reference corpus). В табл. 4 представлены по три наиболее частотные и релевантные для корпусов ЭТ биграммы, выявленные при фоновом сопоставлении с корпусами НТ. Биграммы же релевантные для НТ выявлены путем сопоставления их с ЭТ.

Согласно полученным данным (табл. 4), каждый подкорпус ЭТ имеет в качестве отличительных и одновременно частотных биграмм сочетания, номинирующие а) саму эмоцию (*стыдно за, меня бесят, я счастлива, в шоке*); б) ситуацию, провоцирующую эмоцию (*я одна, у меня нет*); в) обстоятельства (место и время), где чаще всего человек переживает эту эмоцию (*в туалете, в больнице, год назад*) и г) действия, реакции, сопровождающие эмоцию (*говорить*). Для обоих НТ характерны биграммы, не имеющие отношения к сценарию переживания эмоции, – это сочетания служебных слов (*не с, как же, с кем*), клишированных сочетаний (*до н.э., в настоящее время, в связи с*).

Тройки наиболее частотных и релевантных для подкорпусов биграмм

№	Подкорпус	Релевантные биграммы (на фоне подкорпуса НТ)	Частотность в целевом подкорпусе (ipm) / коэффициент релевантности
1	Злость / Гнев	(как) же бесят	629,3/ 63,9
		бесят люди	589, 5/ 60
		меня бесят	463, 7/ 47,4
2	Удивление	об этом	285/29,5
		в шоке	280,1/29
		в голове	230 / 24,1
3	Интерес / Возбуждение	и говорит	208,9 / 21,9
		об этом	178,4/ 18,8
		со словами	158,16,8
4	Стыд / Унижение	стыдно за	976/ 977
		(до сих) пор стыдно	695/ 696
		очень стыдно	615 / 616
5	Страх / Ужас	в больнице	312,7/ 32,3
		год назад	234,6/ 24,5
		два года	189,9 / 20
6	Удовольствие / Радость	я счастлива	836,2/ 84,6
		от счастья	513,7/ 52,4
		(как) же это	334, 5/ 34,5
7	Брезгливость / Отвращение	в говне	364,3 / 37,4
		в туалете	314,7/32
		не ем	215,3/ 22,5
8	Страдание / Тоска	(у) меня нет	567,9/ 57,2
		от одиночества	478, 2/ 48,8
		я одна	269/ 27,9
9	Нейтральный a posteriori (на фоне восьми корпусов ЭТ)	не с	216, 7/22,7
		как же	346,8/ 3,8
		с кем	346,8/ 4,4
10	Нейтральный a priori (на фоне восьми корпусов ЭТ)	до н.э.	194/ 20,5
		в настоящее время	110/ 12
		в связи с	105/ 11,5

5. Обсуждение результатов корпусного анализа

Анализ показывает, что категория эмоциональной нейтральности в интернет-текстах представляет собой сложный феномен.

Во-первых, текстовое пространство нейтральности легко «ложится» на градуированную шкалу, на одном полюсе которой находятся тексты информационного жанра (вики-тексты), а на другом – определенная группа текстов жанра «интернет-откровение». Следовательно, эмоционально-нейтральные тексты, даже в рамках одного коммуникативного пространства – пространства интернет – находятся под влиянием жанровых ограничений. Так, например, одни дискриминантные черты совершенно нехарактерны или характерны в минимальной степени для вики-текстов (НТ а

priori) (обсценизмы, дейктические наречия времени, статистическая значимость он / она / они + быть), в то время как для текстов из ВКонтакте, оцененных как нейтральные (НТ а posteriori), эти же признаки присутствуют: в рамках жанра «интернет-откровение» обсценизмы – это норма, наличие дейктиков естественно, поскольку нарратор рассказывает от первого лица (гомодиегезис), длительно фокусируясь на эмоциональном состоянии другого человека, что приводит к частотности предиката *быть* с местоимениями 3-го лица.

Во-вторых, у эмоциональной нейтральности все-таки есть универсальные текстовые манифестации, независимые от жанра: отсутствие ксенопоказателей, низкая частотность глаголов, отсутствие среди статистически значимых биграмм лексики, относящейся к «вокабуляру эмоций» (табл. 5).

Таблица 5

Сопоставительный анализ дискриминантных черт НТ а priori НТ а posteriori и ЭТ

Дискриминантная черта / подкорпус	НТ а priori	НТ а posteriori	ЭТ
Обсценизмы	–	+	++
Оценочная лексика	+	+	++
Ксенопоказатели	–	–	+
Частотность глаголов	–	–	+
Стат. значимость он / она / они + быть	–	++	+
Высокая частотность дейктических наречий времени	–	+	++
Вокабуляр эмоций	–	–	+

В-третьих, по-видимому, «нижняя граница», отделяющая нейтральные тексты от эмоциональных, обусловлена типом нарратора: если нарратор-наблюдатель рассказывает о себе, то эмоциональность в той или иной степени будет присутствовать, если он рассказывает о других людях, то следует, скорее, констатировать потенциальную эмоциогенность текста, которая может восприниматься реципиентами и как нейтральность. А «верхняя граница» нейтральности, видимо, тесно связана с ситуацией, когда нарратор не является наблюдателем описываемой ситуации или положения дел.

Кроме того, анализ подтверждает тот факт, что интернет-тексты в социальных сетях гипертрофированно эмоциональны, что приводит к изменению на этом фоне границ и прототипов нейтральности. Тексты из социальной сети ВКонтакте, оцененные асессорами как нейтральные (см. тексты 1 и 3), являются таковыми только на фоне эмоционального киберэкзбиционизма, пропитавшего собой социальные сети. Будучи извлечены из данного коммуникативного пространства, они могут произвести на асессоров впечатление эмоциональности. Возможно, на результаты оценки информантами повлияло то, что вики-тексты, составляющие верхнюю границу нейтральности в интернет-пространстве, не были включены в выборку, представленную информантам.

6. Заключение

Исследованный материал показывает, что эмоциональная нейтральность связана с нейтральностью стилистической и оценочной, но этими двумя не исчерпывается. По-видимому, в интернет-пространстве коммуникации складывается своя шкала эмоциональность–нейтральность.

Предположение о том, что одним из признаков эмоционально нейтральных интернет-текстов является их стилистическая невыделенность, соответствие

внутренней норме жанра, в целом подтвердилось. Однако исследование показало, что интернет-тексты разнородны и их внутрижанровые нормы могут значительно различаться. Кроме того, в интернет-коммуникации сложилось свое представление о нормативности, которое существенно отличается от содержания данного концепта в общеязыковой перспективе: обесценизмы и оценочная лексика, обычно неприемлемые для стилистически нейтральных текстов с точки зрения нормы языка, могут функционировать естественным образом в ряде сегментов интернет-коммуникации.

Слабая выраженность модальной рамки, характерная для оценочно нейтральных текстов, выступает, пожалуй, наиболее значимым критерием и для эмоционально нейтральных текстов. Кроме того, для различения нейтральных и эмоциональных интернет-текстов важную роль играет фигура нарратора.

Поскольку интернет-коммуникация в целом более персонализирована, чем, например, иные опосредованные виды коммуникативных процессов, нарратив в ней доминирует. Однако, как показал анализ, различение, с одной стороны, нарратора-экспериментатора, а с другой стороны, нарратора-наблюдателя и нарратора-наблюдателя позволяет прогнозировать специфику дискриминантных черт эмоциональных интернет-текстов и эмоционально нейтральных интернет-текстов и очерчивать границу между ними.

Таким образом, отвечая на исследовательский вопрос «имеют ли нейтральные тексты собственную языковую специфику относительно эмоциональных текстов», отметим, что по большей части – нет. Эта специфика, скорее, из разряда отрицательных: отсутствие дискриминантных черт, выявленных для ЭТ, как раз и характеризует в основном НТ. Однако выявлена и одна собственная специфическая черта данных текстов (для нейтрального подкорпуса текстов жанра интернет-откровение) – тесная связь между глаголом *быть* и местоимениями 3-го лица.

ЛИТЕРАТУРА

1. Кошкарлова Н.Н., Яковлева Е.М. Дискурс новой эмоциональности: коммуникативные практики цифровой реальности // Политическая лингвистика. 2019. № 5 (77). С. 147–152. DOI 10.26170/pl19-05-15.
2. Becker M., Philippe E. Critical Theory and Social Media: Alternatives and the New Sensibility: Honors Theses (PPE). 2018. 66 p.
3. Ловинк Г. Критическая теория Интернета. М.: Ад Маргинем Пресс, 2019. 304 с.
4. Пожидаева Е.В., Карамалак О.А. Хэштеги в социальных сетях: интенции и аффордансы (на примере группы сообщений на английском языке по теме «Food» (Пища / еда)) // Вестник Томского государственного университета. Филология. 2018. № 55. С. 106–118. DOI: 10.17223/19986645/55/8
5. Pang B., Lee L. Opinion Mining and Sentiment Analysis // Foundations and Trends in Information Retrieval. 2008. Vol. 2, № 1–2. P. 1–135.
6. Naakk N., Kirmani M., Mohd M. Emotion analysis: A survey // International Conference on Computer, Communications and Electronics (Comptelix). 2017. P. 397–402.
7. Lövhelm H. A New Three-dimensional Model for Emotions and Monoamine Neurotransmitters // Medical hypotheses. 2012. № 78. P. 341–348.
8. Колмогорова А.В., Калинин А.А., Маликова А.В. Типология и комбинаторика вербальных маркеров различных эмоциональных тоналностей в интернет-текстах на русском языке // Вестник ТГУ. 2019. № 448. С. 48–58. DOI: 10.17223/15617793/448/6
9. Дементьев В.В. «Лукоморье» vs «Википедия» = «Неформально» vs «Формально»? // Жанры речи. 2015. № 1 (11). С. 137–151.
10. Cimini N., Burr J. An Aesthetic for Deliberating Online: Thinking Through «Universal Pragmatics» and «Dialogism» with Reference to Wikipedia // The Information Society. 2012. Vol. 28, № 3. P. 151–160. DOI: 10.1080/01972243.2012.669448.
11. Колмогорова А.В. Использование текстов жанра «интернет-откровение» в контексте решения задач сентимент-анализа // Вестник Новосибирского государственного университета. Серия: Лингвистика и межкультурная коммуникация. 2019. Т. 17, № 3. С. 71–82. DOI 10.25205/1818-7935-2019-17-3-71-82
12. Мустайоки А., Копотев М.В. Современная корпусная русистика // Slavica Helsingiensia 34. Инструментарий русистики: корпусные подходы / под ред.: А. Мустайоки, М.В. Копотева, Л.А. Бирюлина, Е.Ю. Протасовой. Хельсинки, 2008. С. 7–24.
13. Николаев И.С., Митренина О.В., Ландо Т.М. Прикладная и компьютерная лингвистика. М.: ЛЕНАНД, 2016. 320 с.
14. Панов М.В. Труды по общему языкознанию и русскому языку. М.: Языки славянской культуры, 2007. Т. 2. 848 с.
15. Пиотровский Р.Г. Очерки по стилистике французского языка: морфология и синтаксис. Л.: Учпедгиз. Ленинградское отделение, 1960. 224 с.

16. Тинякова Е.А. Философский аспект рассмотрения языка. Москва; Берлин : Директ-медиа, 2015. 354 с.
17. Essays in semiotics / eds. J. Kristeva, J. Rey-Debove, D.J. Umiker. Paris : Mouton de Gruyter, 1971. 445 p.
18. Арутюнова Н.Д. Язык и мир человека. М., 1999. 896 с.
19. Апресян Ю.Д. Избранные труды. Т. II: Интегральное описание языка и системная лексикография. М. : Школа «Языки русской культуры», 1995. 167 с.
20. Арнольд И.В. Семантика. Стилистика. Интертекстуальность : сб. ст. / науч. ред. П.Е. Бухаркин. СПб. : Изд-во СПб. ун-та, 1999. 444 с.
21. Сепир Э. Избранные труды по языкознанию и культурологии: пер. с англ. / общ. ред. и вступ. ст. А.Е. Кибрика. М. : Прогресс-Универс, 1993. 656 с.
22. Шаховский В.И. Обоснование лингвистической теории эмоций // Вопросы психолингвистики. 2019. № 1 (39). С. 22–37.
23. Балли Ш. Французская стилистика. М. : УРСС, 2001. 392 с.
24. Adam J.-M. Les textes types et prototypes. Paris : Armand Colin, 2015. 290 p.
25. Koltsova O.Yu, Alexeeva S.V., Kolcov S.N. An Opinion Word Lexicon and a Training Dataset for Russian Sentiment Analysis of Social Media // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной международной конференции «Диалог» (Москва, 1–4 июля 2016 г.) / под общ. ред.: В. Селерей. М. : Изд-во РГГУ, 2016. Вып. 15. P. 277–287.
26. Negi S., Buitelaar P. Suggestion Mining From Opinionated Text // Sentiment Analysis in Social Networks / eds.: F.A. Pozzi, E. Fersini, E. Messina, B. Liu, M. Kaufmann. Elsevier, 2017. P. 129–139. DOI: <https://doi.org/10.1016/B978-0-12-804412-4.00008-5>.
27. Tiwari P., Yadav P., Agnihotri S., Mishra B., Nhu N., Gochhayat S., Singh J., Prasad M. Sentiment Analysis for Airlines Services Based on Twitter Dataset // Social Network Analytics. Computational Research Methods and Techniques. 2019. P. 149–162. DOI: 10.1016/B978-0-12-815458-8.00008-6.
28. Yadav A., Jha C.K., Sharan, A., Vaish, V. Sentiment analysis of financial news using unsupervised approach // Procedia Computer Science. 2020. Vol. 167. P. 589–598. DOI: <https://doi.org/10.1016/j.procs.2020.03.325>.
29. Воронина И.Е., Гончаров В.А. Анализ эмоциональной окраски сообщений в социальных сетях (на примере сети «ВКонтакте») // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. 2015. № 4. С. 148–155.
30. Арутюнова Н.Д. Высказывание в контексте диалога и чужой речи // Revue des études slaves. 1990. Tome 62, fascicule 1–2. L'énonciation dans les langues slaves [En hommage à René L'Hermite, sous la direction de Jean-Paul Sémon et Hélène Włodarczyk]. P. 15–30.
31. Падучева Е.В. Показатели чужой речи *мол* и *декабрь* // Известия РАН. Серия литературы и языка. 2011. Т. 70, № 3. С. 13–19.
32. Колмогорова А.В., Калинин А.А., Маликова А.В. Кто и о чем говорит в «радостных» и «грустных» текстах: в поисках дискриминантных черт текстов разных эмоциональных тональностей // Известия Уральского федерального университета. Серия 2: Гуманитарные науки. 2019. Т. 21, № 4 (193). С. 219–234. DOI 10.15826/izv2.2019.21.4.078
33. ИONOBA C.B. Эмотивность текста как лингвистическая проблема : дис. ... канд. филол. наук. Волгоград, 1998. 197 с.
34. Падучева Е.В. Глаголы *быть* и *бывать*: история и современность // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции Диалог (Москва, 27–30 мая 2015 г.). М. : Изд-во РГГУ, 2015. Т. 1, вып. 14 (21). С. 500–514.
35. Захаров В.П., Хохлова М.В. Анализ эффективности статистических методов выявления коллокаций в текстах на русском языке. Компьютерная лингвистика и интеллектуальные технологии. 2010. № 9 (16). С. 137–143. URL: <http://www.dialog-21.ru/dialog2010/materials/html/22.htm>

Статья представлена научной редакцией «Филология» 16 августа 2020 г.

The Specificity of Emotionally Neutral Internet Texts

Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal, 2020, 458, 22–31.

DOI: 10.17223/15617793/458/3

Anastasia V. Kolmogorova, Siberian Federal University (Krasnoyarsk, Russian Federation). E-mail: nastiakol@mail.ru

Keywords: emotion; Internet texts; sentiment analysis; verbal markers; neutral texts; social network.

The study is supported by the Russian Foundation for Basic Research, Project No. 19-012-00205.

The article aims to discuss the problem of verbal features distinguishing emotionally neutral texts from emotional ones. Such features are necessary for elaborating a computer classifier designed to assign Internet texts in Russian to different emotional classes of texts including the neutral class. The complex analysis of the notion of neutrality in stylistics, axiology, emotiology and sentiment analysis allowed the author to formulate the preliminary hypothesis that emotionally neutral texts are marked by (1) the absence of obscene words and evaluative adjectives; (2) the non-expression of the modal frame of the utterance, which influences the functioning of xenomarkers (markers of narrative retelling), of deictic words and frequent predicates; (3) the lack of emotional vocabulary. The sources of the data for the study were three subcorpora: a subcorpus of texts from Wikipedia (136,184 tokens)—“neutral texts a priori”; a subcorpus of texts from the Russian social network VKontakte assessed by 2,000 informants as neutral ones—“neutral texts a posteriori” (14,000 tokens), and, finally, 15,000 emotionally loaded fragments of 60 to 80 words (1,092,327 tokens) from VKontakte assessed by the informants as texts manifesting different emotions: from anger to joy. The choice of the two neutral subcorpora was due to the author’s intention to examine the category of neutrality taking under consideration the genre of Internet texts. To learn the features of neutral Internet texts, the author used the methodology of corpus linguistics largely supported by a number of tools offered by the Sketch Engine corpus manager platform. The results partially confirmed the hypothesis and provided new details. Thus, the analysis revealed that the genre of the Internet text plays a crucial role in how the neutrality category manifests in it: for instance, Wikipedia texts have no obscene words, VKontakte texts assessed as neutral have a number of such lexemes. A similar discrepancy between neutral texts of different genres is observed on the grammatical level: for VKontakte texts the collocation of the Russian verb *byt’* [be] and the pronouns of the third person *on, ona, oni* [he, she, they] is statistically significant, for Wikipedia texts it is not. There are still some features proper to the two subcorpora considered as neutral: the absence of so-called “xenomarkers” and emotional vocabulary (words naming emotions, their manifestations, and typical situations in which people usually feel them); the low frequency, in comparison with emotional texts, of verbs and deictic words indicating time or place. The author has made a conclusion that, in the internet communicative space, the category of neutrality is construed differently than in others spheres of communication. Internet texts are generally so saturated with emotions that, even if in another context a text would be interpreted as an emotional one, in the social network users assess it as a fragment with no emotion. Another significant result is that the author has found a correlation between the type of narrator—narrator-experiencer, narrator-observer and narrator-non observer—and the type of text in terms of the category of neutrality. In further research, this will facilitate the predictability of text class—emotional or neutral.

REFERENCES

1. Koshkarova, N.N. & Yakovleva, E.M. (2019) Discourse of new emotionality: Communicative practices of digital reality. *Politicheskaya lingvistika – Political Linguistics*. 5 (77). pp. 147–152. (In Russian). DOI 10.26170/pl19-05-15
2. Becker, M. & Philippe, E. (2018) *Critical Theory and Social Media: Alternatives and the New Sensibility*. Honors Theses (PPE). Philadelphia.
3. Lovink, G. (2019) *Kriticheskaya teoriya Interneta* [Critical Theory of the Internet]. Translated from English. Moscow: Ad Marginem Press.
4. Pozhidaeva, E.V. & Karamalak, O.A. (2018) Hashtags in social networks: intentions and affordances (exemplified in the English language by message groups on the topic “Food”). *Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya – Tomsk State University Journal of Philology*. 55. pp. 106–118. (In Russian). DOI: 10.17223/19986645/55/8
5. Pang, B. & Lee, L. (2008) Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*. 1–2 (2). pp. 1–135.
6. Haakk, N., Kirmani, M. & Mohd, M. (2017) Emotion analysis: A survey. *Comptelix 2017*. Proceedings of the International Conference on Computer, Communications and Electronics. Jaipur. 1–2 July 2017. New York: IEEE. pp. 397–402. DOI: 10.1109/COMPTELIX.2017.8004002
7. Lövhelm, H. (2012) A New Three-dimensional Model for Emotions and Monoamine Neurotransmitters. *Medical Hypotheses*. 78. pp. 341–348.
8. Kolmogorova, A.V., Kalinin, A.A. & Malikova, A.V. (2019) The types and combinatorics of verbal markers of different emotional tonalities in Russian-language Internet texts. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 448. pp. 48–58. (In Russian). DOI: 10.17223/15617793/448/6
9. Dement'ev, V.V. (2015) “Lurkmore.ru” vs. “Wikipedia” = “informally” vs. “formally”? *Zhanry rechi – Speech Genres*. 1 (11). pp. 137–151. (In Russian).
10. Cimini, N. & Burr, J. (2012) An Aesthetic for Deliberating Online: Thinking Through “Universal Pragmatics” and “Dialogism” with Reference to Wikipedia. *The Information Society*. 3 (28). pp. 151–160. DOI: 10.1080/01972243.2012.669448
11. Kolmogorova, A.V. (2019) Texts of “internet confessions” as a source for training data set for the research on the sentiment-analysis field. *Vestnik Novosibirskogo gosudarstvennogo universiteta. Seriya: Lingvistika i mezhkul'turnaya kommunikatsiya – NSU Vestnik. Series: Linguistics and Intercultural Communication*. 3 (17). pp. 71–82. (In Russian). DOI: 10.25205/1818-7935-2019-17-3-71-82
12. Mustayoki, A. & Kopotev, M.V. (2008) Sovremennaya korpusnaya rusistika [Modern Corpus Russian Studies]. In: Mustayoki, A. et al. (eds) *Slavica Helsingiensia 34. Instrumentariy rusistiki: korpusnye podkhody* [Slavica Helsingiensia 34. Toolkit for Russian Studies: Corpus Approaches]. Helsinki: Helsinki State University. pp. 7–24.
13. Nikolaev, I.S., Mitrenina, O.V. & Lando, T.M. (2016) *Prikladnaya i komp'yuternaya lingvistika* [Applied and Computational Linguistics]. Moscow: LENAND.
14. Panov, M.V. (2007) *Trudy po obshchemu yazykoznaniyu i russkomu yazyku* [Proceedings on General Linguistics and Russian Language]. Vol. 2. Moscow: Yazyki slavyanskoy kul'tury.
15. Piotrovskiy, R.G. (1960) *Ocherki po stilistike frantsuzskogo yazyka: morfologiya i sintaksis* [Essays on the Stylistics of the French Language: Morphology and syntax]. Leningrad: Uchpedgiz.
16. Tinyakova, E.A. (2015) *Filosofskiy aspekt rassmotreniya yazyka* [Philosophical Aspect of Language Consideration]. Moscow-Berlin: Direkt-media.
17. Kristeva, J., Rey-Debove, J. & Umiker, D.J. (eds) (1971) *Essays in Semiotics*. Paris: Mouton de Gruyter.
18. Arutyunova, N.D. (1999) *Yazyk i mir cheloveka* [Human Language and Human World]. Moscow: Yazyki russkoy kul'tury.
19. Apresyan, Yu.D. (1995) *Izbrannye trudy* [Selected Works]. Vol. II. Moscow: Shkola “Yazyki russkoy kul'tury”.
20. Arnol'd, I.V. (1999) *Semantika. Stilistika. Intertekstual'nost'* [Semantics. Stylistics. Intertextuality]. Saint Petersburg: Saint Petersburg State University.
21. Sapir, E. (1993) *Izbrannye trudy po yazykoznaniyu i kul'turologii* [Selected Works on Linguistics and Cultural Studies]. Translated from English by A.E. Kibrik. Moscow: Progress-Univers.
22. Shakhovskiy, V.I. (2019) Linguistic theory of emotions and its foundations. *Voprosy psikholingvistiki – Journal of Psycholinguistics*. 1 (39). pp. 22–37. (In Russian). DOI: 10.30982/2077-5911-2019-39-1-22-37
23. Bally, Ch. (2001) *Frantsuzskaya stilistika* [French Stylistics]. Translated from French by K.A. Dolinin. Moscow: URSS.
24. Adam, J.-M. (2015) *Les textes types et prototypes*. Paris, Armand Colin.
25. Koltsova, O.Yu., Alexeeva, S.V. & Koltsov, S.N. (2016) An Opinion Word Lexicon and a Training Dataset for Russian Sentiment Analysis of Social Media. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii* [Computational Linguistics and Intellectual Technologies]. Proceedings of the International Conference “Dialog 2016” [“Dialogue 2016”]. 15. Moscow. 1–4 July 2016. Moscow: Russian State University for the Humanities. pp. 277–287.
26. Negi, S. & Buitelaar, P. (2017) Suggestion Mining From Opinionated Text. In: Pozzi, F.A. et al. (eds) *Sentiment Analysis in Social Networks*. Elsevier. pp. 129–139. DOI: 10.1016/B978-0-12-804412-4.00008-5
27. Tiwari, P. et al. (2019) Sentiment Analysis for Airlines Services Based on Twitter Dataset. In: Dey, N. et al. (eds) *Social Network Analytics. Computational Research Methods and Techniques*. Elsevier. pp. 149–162. DOI: 10.1016/B978-0-12-815458-8.00008-6
28. Yadav, A. et al. (2020) Sentiment analysis of financial news using unsupervised approach. *Procedia Computer Science*. 167. pp. 589–598. DOI: 10.1016/j.procs.2020.03.325
29. Voronina, I.E. & Goncharov, V.A. (2015) Sentiment analysis of messages in social media (on example of “Vkontakte”). *Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriya Sistemnyy analiz i informatsionnye tekhnologii – Proceedings of Voronezh State University. Series: Systems analysis and information technologies*. 4. pp. 148–155. (In Russian).
30. Arutyunova, N.D. (1990) Vyskazyvanie v kontekste dialoga i chuzhoi rechi [Utterance in the context of dialogue and other's speech]. *Revue des études slaves*. 1–2 (62). pp. 15–30.
31. Paducheva, E.V. (2011) Particles MOL and DESKAT as markers of somebody else's speech. *Izvestiya RAN. Seriya literatury i yazyka – Bulletin of the Russian Academy of Sciences: Studies in Literature and Language*. 3 (70). pp. 13–19. (In Russian).
32. Kolmogorova, A.V., Kalinin, A.A. & Malikova, A.V. (2019) Who and about what speaks in “cheerful” and “sad” texts: in search of discrimination features in texts of different emotional tonalities. *Izvestiya Ural'skogo federal'nogo universiteta. Ser. 2: Gumanitarnye nauki – Izvestia. Ural Federal University Journal. Series 2. Humanities and Arts*. 4:21 (193). pp. 219–234. (In Russian). DOI: 10.15826/izv2.2019.21.4.078
33. Ionova, S.V. (1998) *Emotivnost' teksta kak lingvisticheskaya problema* [Text emotiveness as a linguistic problem]. Philology Cand. Diss. Volgograd.
34. Paducheva, E.V. (2015) [Verbs Byt' and Byvat': Contemporary State and History]. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii* [Computational Linguistics and Intellectual Technologies]. Proceedings of the International Conference “Dialog 2015” [“Dialogue 2015”]. 14 (21). Vol. 1. Moscow. 27–30 May 2015. Moscow: Russian State University for the Humanities. pp. 500–514. (In Russian).
35. Zakharov, V.P. & Khokhlova, M.V. (2010) [Study of effectiveness of statistical measures for collocation extraction on Russian texts]. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii* [Computational Linguistics and Intellectual Technologies]. Proceedings of the International Conference “Dialog 2010” [“Dialogue 2010”]. 9 (16). Moscow. 26–30 May 2010. Moscow: Russian State University for the Humanities. pp. 137–143. [Online] Available from: <http://www.dialog-21.ru/dialog2010/materials/html/22.htm>. (In Russian).

Received: 16 August 2020