*Вестник Томского государственного университета. 2019. № 448. С. 48–58.* DOI: 10.17223/15617793/448/6

УДК 81.32

*А.В. Колмогорова, А.А. Калинин, А.В. Маликова*

# ТИПОЛОГИЯ И КОМБИНАТОРИКА ВЕРБАЛЬНЫХ МАРКЕРОВ РАЗЛИЧНЫХ ЭМОЦИОНАЛЬНЫХ ТОНАЛЬНОСТЕЙ В ИНТЕРНЕТ-ТЕКСТАХ НА РУССКОМ ЯЗЫКЕ

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

Теоретически обосновывается понятие вербального маркера. Представлена типология вербальных маркеров восьми эмоций в русскоязычных интернет-текстах в соответствии с уровнем языковой системы, которому принадлежит маркирующая единица: лексические, морфологические, синтаксические, семантические, пунктуационные и текстовые. Приводятся примеры и статистика эффективности маркеров, рассматриваются их комбинаторные особенности.
**Ключевые слова:** эмоция; интернет-тексты; сентимент-анализ; вербальные маркеры; машинное обучение; когниция.

## Введение

Становление антропоцентрической парадигмы в языкознании способствовало изменению функционального статуса языковых данных в исследовательской деятельности. Так, если в системно-структурной лингвистике «показания» языкового материала служили для языковедов свидетельствами специфики устройства языковой системы, а исследователи-когнитивисты так называемой «первой волны» ставили своей целью объяснить различные языковые факты, опираясь на стоящие за ними ментальные феномены [1. С. 53], то сегодняшние лингвисты стремятся диагностировать скрытые от непосредственного наблюдения когнитивно-психологические процессы на основании установления корреляций между ними и употреблением в речи субъектом этих процессов тех или иных языковых единиц, структур.

Актуальность подобного подхода имеет под собой прагматические, теоретические и технологические основания.

Так, научные исследования, выполняемые при поддержке крупных корпораций, государственных и производственных структур, все более ориентированы на получение прагматически конкретных, ясных и применимых на практике результатов.

«Прорывной» с точки зрения теории явилась гипотеза о гипертекстуальном устройстве мозга, исходящей из постулата о том, что разум – это многослойная структура, сеть сетей нейронных сетей. Системной единицей функционирования такой гиперсети является ког – распределенная группа нейронов, сцепленная единым когнитивным опытом, в том числе – и опытом языковым [2]. Иными словами, если представить ког в виде снопа [Там же], то один или несколько его «колосьев», образующих в итоге единую вершину, представляют собой «следы» опыта взаимодействия со словом в некоторой когнитивно значимой для индивида ситуации. Следовательно, например, нейронные связи, сформированные в опыте переживания эмоции, отражают и опыт взаимодействия со словом в эмоционально маркированной ситуации таким образом, что, актуализировав первые, некий стимул актуализирует и вторые, и наоборот.

Наконец, технологический прорыв последних десятилетий дал возможность представителям всех наук, и в том числе лингвистам, использовать, во-первых, массивы «больших данных», а во-вторых, автоматизированные компьютерные приложения для обработки этих данных, в частности, в междисциплинарных исследованиях [3].

В лаборатории когнитивных исследований и прикладной лингвистики Института филологии и языковой коммуникации Сибирского федерального университета в течение нескольких лет ведется разработка теоретических оснований, сбор материала и создание технологического обеспечения для автоматического распознавания эмоциональной тональности интернет-текстов на русском языке.

Интернет-тексты в качестве объекта исследования имеют особую привлекательность, поскольку, во-первых, они формируют один из самых значимых сегментов речевой продукции носителей современного русского языка, а, во-вторых, разработка технологии автоматической оценки их эмоциональности имеет наиболее ясные коммерческие и социальные перспективы: она может быть использована для мониторинга субъективного эмоционального восприятия потребителями той или иной услуги, товара или события на основе оценки текстов форумов, чатов, пабликов, а также для мониторинга эмоциональной безопасности уязвимых социальных групп в социальных сетях (дети, подростки, пожилые люди).

Данная публикация ставит своей целью обобщить опыт установления корреляций между некоторым эмоциональным состоянием, переживаемым субъектом, и теми языковыми средствами, которые он использует в своей текстовой деятельности. В статье теоретически обосновывается понятие вербального маркера эмоций, приводится типология вербальных маркеров, делаются наблюдения о влиянии комбинации маркеров на точность работы классификатора, созданного в целях сентимент-анализа.

## Понятие вербального маркера

Анализ научной лингвистической литературы, в которой используется понятие маркера, позволил сде-

лать следующее обобщение: в лингвистике родовой термин «маркер» используется в ситуации, когда *некоторая единица или структура* наблюдается и фиксируется лингвистом на определенном отрезке речи / текста, продуцирование которого по времени совпадает с протеканием в сознании продуцента речи / текста недоступных прямому наблюдению коммуникативных, психических, когнитивных действий или процессов, косвенно указывая на них.

В зависимости от того, на каком уровне абстракции исследователь решается анализировать данные единицу и структуру, для их обозначения используются различные термины.

Так, анализируя интерпретативно-когнитивные процессы участников дискурсивного взаимодействия, оперируют терминами «дискурсивные маркеры» (discourse markers) [4, 5], «дискурсивные операторы» (discourse operators), «дискурсивные коннективы» (discourse connectives), а также «ключевые выражения» (cue phrases) [6].

В контексте средств реализации воздействующей функции языка указывается на важную роль прагматических маркеров, под которыми понимается перечень синтаксически разнородных лексем, обладающих свойствами индексальности, контекстной зависимости и многофункциональности, имеющих оценочную и метакоммуникативную функции, но лишенных концептуального значения [7].

Если исследователи сосредоточены на коммуникативно-жанровом аспекте речевого взаимодействия, то предпочтительным считают сочетание «речевой маркер». Например, в [8. С. 314] реплика «я не прокурор, чтоб с тобой по душам говорить» рассматривается как речевой маркер насмешливого отношения к жанру «разговор по душам», а описательное утверждение «я вчера разозлилась на тебя за то, что ты забыл про нашу годовщину» – как маркер установки на сотрудничество в ситуации бытового конфликта [9].

При рассмотрении корреляций между фонетическими и просодическими характеристиками речи (высотой основного тона, паузацией, акцентным выделением сегментов, ритмом речи, длительностью произнесения слов, звуков) и эмоциональным состоянием говорящего продуктивно используется терминологическое сочетание «речевые временные корреляты эмоций» [10], а для выявления специфической просодики кульминации в юмористических нарративах [11] – собственно «просодические маркеры».

Когда фокус внимания исследователя сосредоточен на лексических единицах, чаще всего в целях автоматической обработки и классификации текстов, то употребительной становится номинация «лексический маркер» [12].

В случаях, когда изучается совокупность единиц и структур разного уровня, совместная встречаемость которых указывает на некоторое неязыковое и недоступное наблюдению явление или феномен (правда и ложь, например [13]), используются обобщенные термины «лингвистические маркеры» или «вербальные маркеры». В работе [14] в качестве вербальных маркеров агрессии в рамках тактики возмущения рассматриваются и лексические (междометие *ну*, например), и синтаксические (неполные предложения), и пунктуационные (использование нескольких восклицательных знаков), и графические (запись всего предложения заглавными буквами) средства.

Представляется, что в качестве наиболее нейтральной родовой номинации следует использовать именно термин «вербальный маркер», поскольку, как показывает анализ лингвистической литературы, такая традиция уже сложилась – в случае, когда уровневая специфика маркеров не имеет значения, носит смешанный характер [15] или определяется уже post factum в результате исследования [16], авторы говорят об именно «лингвистических» или «вербальных маркерах».

С точки зрения своих онтологических характеристик, вербальные маркеры могут иметь качественное и количественное измерения. Так, для идентификации эмоционального состояния боязни, переживаемого персонажем художественного произведения, достаточным оказывается выявление в тексте такой качественной характеристики, как наличие лексем *шепот, прятаться, бледный* [17]. Но, например, для выявления и классификации текстов, написанных людьми, страдающими суицидальными мыслями, маркером является определенное статистически значимое количество так называемых «абсолютистских» слов (*everyone, completely, never,* etc.) [18]. А для исследования уровня манипулятивности политических текстов американских масс-медиа важным оказался не только статистически значимый порог частотности лексических маркеров манипуляции одного какого-то вида, но и степень полноты представленности в тексте различных маркеров из списка слов-предикторов [19]. Иными словами, можно говорить о трех измерениях маркеров: качественном, количественном интенсивном и количественном экстенсивном. В двух последних случаях маркер приобретает статус величины, которую можно измерить и, в соответствии с проведенным измерением, определить некую характеристику текста – маркер превращается в параметр [20. P. 3].

Таким образом, **вербальный маркер** – это единица или дискретная структура одного из уровней языковой системы, в ряде случаев поддающаяся параметризации, появление которой (а) само по себе или (б) в совокупности с другими релевантными единицами и / или (в) с определенной частотностью в сегменте речи / тексте косвенно указывает на некоторые сложные и недоступные непосредственному наблюдению процессы, феномены, фундированные в когнитивной деятельности, психической жизни продуцента речи / текста, способные при определенных условиях «запускать» аналогичные процессы во внутреннем мире реципиента речи / дискурса.

## Методология выявления вербальных маркеров русскоязычных интернет-текстов различных эмоциональных классов

Ведущей методологией реализуемого нами проекта в целом является методология так называемого сентимент-анализа текстов. Это выявление эмоциональной тональности текста при помощи методов NLP (автома-

тической обработки естественного языка), статистики, машинного обучения [21]. Впервые этот термин был использован в статьях S.R. Das и M.Y. Chen [22], B. Pang, L. Lee и Sh. Vaithyanathan [23].

Существующие алгоритмы сентимент-анализа варьируются по критерию количества классов эмоций, к которым будут отнесены тексты в результате прохождения через классификатор: бинарные классификаторы определяют тональность текста как позитивную / негативную [24] или объективную / субъективную [25]; тернарные – как позитивную / нейтральную / негативную (например, при анализе тональности твитов, опубликованных на протяжении Чемпионата мира по футболу 2014 г.) [26]; многоклассовые осуществляют атрибуцию текста к конкретному классу эмоций, в соответствии с выбранной классификацией эмоций. Такова попытка итальянских исследователей распределить новостные статьи между тринадцатью классами эмоций, доминирующих в них [27]. Для русскоязычного материала используются преимущественно тернарные классификаторы [28, 29]. В нашем проекте решается задача создания многоклассового классификатора для русскоязычных текстов. На сегодняшний момент значение F-меры (гармоническое среднее между точностью и полнотой классификации) в различных эмоциональных классах текстов, а их 8 (+1 нейтральный), варьируется от 30 до 50%. Данный результат несколько лучше аналогичных показателей [30], где значение F-меры для семи эмоциональных классов текстов составило 47%, а для класса нейтральных текстов – 70%, однако еще не достигает показателей, типичных для бинарных классификаторов с точностью 60,6% для текстов с позитивной эмоциональной тональностью и 72,8 – с негативной [31].

Для выделения классов текстов согласно критерию ведущей эмоции, вербализованной в них, мы использовали классификацию эмоций Гуго Лёвхейма, который установил, что, хотя сами по себе эмоциональные состояния, являясь функцией от адаптивных систем человеческого организма, порождаются в лимбической системе и миндалевидном теле головного мозга, дальнейший сигнал об эмоции активируется и распространяется на другие отделы головного мозга благодаря действию трех моноаминов: серотонина, допамина и норадреналина. Такая система моноаминных медиаторов служит своеобразным «эмоциопроводом» для передачи информации об эмоции всем остальным отделам мозга [32. P. 341]. Иными словами, изменение уровня того или иного моноамина является посланием для мозга об активируемой эмоции. Взяв за основу восьмичленную классификацию аффектов С. Томкинса, выявленную на основе анализа типов выражений лица, Г. Лёвхейм установил корреляцию каждой из восьми эмоций со специфической комбинацией уровней трех названных выше моноаминов. Исследователь визуализировал данную корреляцию в виде куба (рис. 1) на координатной плоскости с осями 5-HT (серотонин), NE (норадреналин), DA (дофамин). В зависимости от сочетания уровня данных гормонов в крови субъекта эмоции исследователь предложил восьмичленную классификацию эмоций, где первая номинация класса отражает наименее выраженную степень интенсивности эмоции-аффекта, а вторая – ее высшую точку (например, Злость – Гнев): Интерес / Возбуждение; Удовольствие / Радость; Удивление; Страдание / Тоска; Гнев / Ярость; Страх / Ужас; Презрение / Отвращение; Стыд / Унижение.

Данная модель представляется одной из наиболее объективных классификаций эмоций, существующих на сегодняшний день, поскольку она базируется на измеряемых физиологически фундированных параметрах, а количество выделяемых эмоций является удобным для дальнейшей работы с текстами.

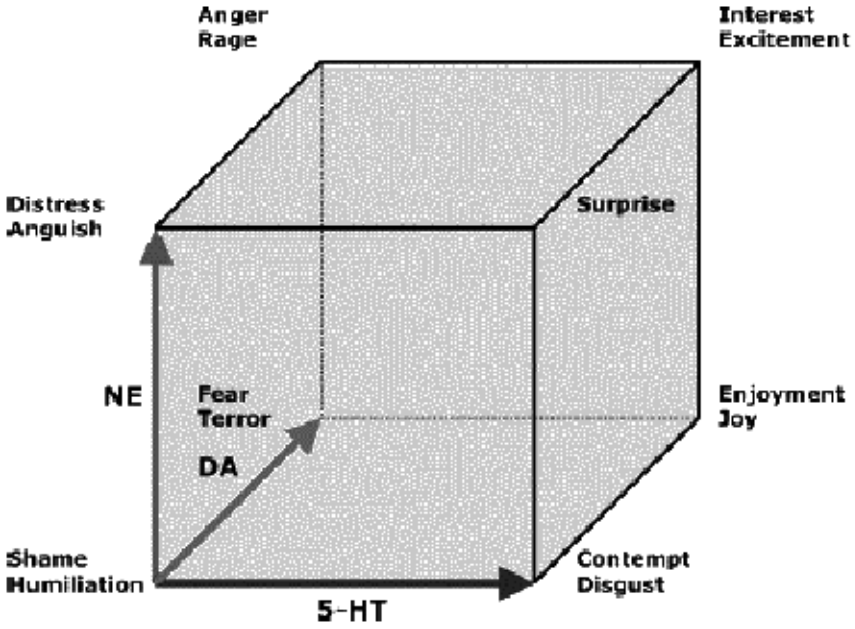


Рис. 1

Хотя существуют классификаторы для сентимент-анализа, построенные на использовании словарей тональностей, в качестве метода для нашего проекта мы выбрали технологию машинного обучения по преце-

дентам, основной принцип которой – по частным данным, представляющим набор пар «объект – ответ», выявить закономерности, присущие не только конкретной обучающей выборке, но и генеральной совокупности данных. Такой выбор обусловлен тем, что, как показывает анализ научных публикаций последних лет, методы машинного обучения существенно расширяют возможности обработки различного текстового материала. Существует два основных алгоритма машинного обучения − нейронные сети [33] и метод опорных векторов (SVM) [34]. В нашей работе использован SVM алгоритм.

Для подобной технологии важнейшим этапом является этап формирования обучающей выборки – коллекции текстов, где каждому из них уже приписан эмоциональный класс: к примеру, некоторый текст вербализует эмоцию радости, а другой отмечен страданием.

Источником данных для обучающей выборки в нашем случае послужил паблик Подслушано в социальной сети ВКонтакте – проект, в котором пользователи анонимно делятся каждый день своими откровениями и жизненными ситуациями. Посты в паблике имеют объем 60–80 слов. По своей жанровой характеристике это нарративы, в которых объектом эмоции является, как правило, некоторая жизненная ситуация, переживаемая самим автором-нарратором в данный момент или пережитая в прошлом, а также ситуации, происходящие с другими людьми, но проецируемые нарратором, в силу эмпатии, на себя.

Для извлечения данных эмоциональные классы по Лёвхейму были соотнесены с хештегами, под которыми размещаются посты пользователей, поскольку именно хештеги в случае коммуникации в социальных сетях передают основной образ, идею [35] (табл. 1):

Таблица 1

**Объем подкорпусов и их соотнесение с хештегами**

| Эмоциональный класс текстов (подкорпус) | Объем подкорпуса в токенах | Хештег в Подслушано |
|---|---|---|
| Страдание / Тоска | 56 470 | #Подслушано_одиночество |
| Интерес / Возбуждение | 184 074 | #Подслушано_успех |
| Удовольствие / Радость | 85 117 | #Подслушано_счастье |
| Страх / Ужас | 230 730 | #Подслушано_страшное |
| Брезгливость / Отвращение | 45 868 | #Подслушано_фууу |
| Злость / Гнев | 131 564 | #Подслушано_БЕСИТ |
| Стыд / Унижение | 70 232 | #Подслушано_стыдно |
| Удивление | 288 272 | #Подслушано_наблюдения<br>#Подслушано_странное |

Несмотря на то, что редакторы паблика самостоятельно категоризируют посты при помощи хештегов, тексты из каждого класса были рандомизированно оценены асессорами. В случае значительного разброса оценок асессоров данные подвергались повторной разметке информантами на одной из краудсорсинговых платформ. Так, например, произошло с корпусом текстов под хештегом #Подслушано_наблюдения#, в постах которого, по нашим оценкам, широко представлена эмоция удивления. Для проведения разметки асессорам в формате онлайн-опросника предлагалось поставить «галочку» напротив имени той эмоции, которую они чувствуют в данном текстовом фрагменте. В распоряжении информантов-асессоров было восемь двучленных имен эмоций и девятая номинация – нейтрально. Информанты не были ограничены в количестве приписываемых текстовому фрагменту эмоций. В итоге один и тот же текст оценивался тремя асессорами. Если в двух или трех ответах-оценках была указана эмоция удивления, то текст включался в класс текстов «Удивление», если подобная эмоция не была указана ни одним или всего одним информантом, он не включался в обучающую выборку.

Хотя классификатор способен самостоятельно строить статистические модели на основе закономерностей, устанавливаемых им на основе анализа обучающей выборки, но если «подать» ему на вход «лингвистические подсказки», т.е. языковые единицы, структуры, на присутствие которых или их частотность в текстах стоит «обратить особое внимание», то точность классификации значительно возрастает. Для того чтобы эти подсказки сформулировать, необходимо провести лингвистический анализ достаточно объемного массива текстов в каждой из категорий. Сделать это «вручную» чрезвычайно трудно, поэтому был использован инструментарий корпусной лингвистики, предлагаемый корпусным менеджером Sketch Engine. Корпусный менеджер – это специальная информационно-поисковая система, включающая программные средства для поиска данных в корпусе, получения статистической информации [36. С. 139]. В Sketch Engine мы загрузили восемь подкорпусов, каждый из которых соответствует одной из восьми эмоций, согласно Лёвхейму. Результаты, которые станут предметом обсуждения в следующей части статьи, были получены в ходе анализа данных восьми подкорпусов при помощи таких инструментов Sketch Engine, как анализ частотности лексемы / словоформы / синтаксемы, выявление комбинаторики лексем, создание конкордансов, тезаурусов для каждой лексемы из корпуса.

Приводимые далее данные об эффективности тех или иных вербальных маркеров для точности классификации получены в ходе работы с прототипом программы – классификатора текстов по их эмоциональной тональности, созданном на базе алгоритма Наивного байесовского классификатора с помощью языка программирования Python.

### Типы вербальных маркеров эмоций и эффективность их применения в качестве параметров для работы классификатора

*Лексические маркеры* эмоции могут иметь различную экстенсивность: для атрибуции текста к определенному эмоциональному классу значимой может быть частотность как одной лексемы, так и коллокации. Кроме того, функцию маркирования эмоции может выполнять число различных лексем-репрезентантов какого-либо лексико-семантического поля (далее – ЛСП) или лексико-семантической группы (далее – ЛСГ), встречающихся в тексте.

Например, сравнительный анализ (1) относительной частотности, (2) статистической значимости по результатам TF / IDF взвешивания[1] и (3) процентного соотношения текстов, содержащих хотя бы одну форму глагольной лексемы *говорить*, в восьми классах текстов из обучающей выборки показал, что в четырех эмоциональных классах текстов (Злость / Гнев, Удивление, Интерес / Возбуждение, Стыд / Унижение) значения всех трех параметров значительно выше, чем для оставшихся классов текстов. И наоборот, для класса текстов, передающих эмоциональную тональность Страдание / Тоска, лексема имеет самые низкие значения во всех трех типах измерений (столбцы 2–4, табл. 2).

Таблица 2

Относительная частотность и статистическая релевантность леммы *говорить* в восьми классах текстов

| № | Название | Относительная частотность / ранг | Ранг в первой двадцатке и значение по результатам TF / IDF взвешивания | % текстов в классе, содержащих *говорить* |
|---|---|---|---|---|
| 1 | **Злость / Гнев** | **2,082.64** pm / 6 | **5 7.549767342194897** | 9,811 |
| 2 | **Удивление** | **1,765** pm / 6 | **11 7.09326299833163** | **13,138** |
| 3 | **Интерес / Возбуждение** | **1,684** pm / 3 | **3 7.432211883185744** | **12,361** |
| 4 | **Стыд / Унижение** | **1,580** pm / 6 | **6 7.915890436201671** | **12,528** |
| 5 | Страх / Ужас | 1,512.69 pm / 8 | не входит в первые 20 | **13,078** |
| 6 | Удовольствие / Радость | 1,186.6 pm / 10 | не входит в первые 20 | 8,988 |
| 7 | Брезгливость / Отвращение | 1,046.48 pm / 13 | не входит в первые 20 | 7,342 |
| 8 | Страдание / Тоска | 1,027.09 pm / 14 | не входит в первые 20 | 5,030 |

Когда данный вербальный маркер был добавлен к уже показавшей свою эффективность базе параметров для автоматического анализа, точность классификации выросла на 1% для текстов из классов Брезгливость / Отвращение, Страх / Ужас и Удивление (см. табл. 4).

В ряду лексических маркеров эмоций важную роль играет частотность отдельных коллокаций. Так, коллокация *терпеть не могу* оказалась маркером для класса текстов, передающих эмоцию гнева, злости (табл. 3), ее подача на вход классификатора в качестве параметра позволила на 0,5% улучшить точность классификации в классе Злость / Гнев (см. табл. 4):

(1) *И я терпеть не могу, когда меня называют «человек с ограниченными возможностями»* (Злость / Гнев).

Таблица 3

Абсолютная и относительная частотность коллокации *терпеть не могу* в восьми классах текстов

| **Класс / коллокация** | **Страдание / Тоска** | **Интерес / Возбуждение** | **Удовольствие / Радость** | **Страх / Ужас** | **Брезгливость / Отвращение** | **Злость / Гнев** | **Стыд / Унижение** | **Удивление** |
|---|---|---|---|---|---|---|---|---|
| Терпеть не могу | 0 | 0 | 0 | 0 | 0 | 5/68.4 pm | 1 /14.24 pm | 1/1.73 pm |

Наконец, анализ процентного соотношения текстов, содержащих 2 и более лексем из ЛСП «Смерть» (*смерть, умирать, умереть, могила, похороны, кладбище, оплакивать, оплакать, скорбеть, хоронить, похоронить, скончаться, захоронить, погибнуть, погибать, кремировать, осиротеть* и т.д.), в каждом из восьми классов показал, что наибольший процент таких текстов отмечается в эмоциональном классе Страх / Ужас (пр. 2) (27,153%), а наименьший – Брезгливость / Отвращение (0,6%). В результате добавления данного вербального маркера в качестве дополнительного к базовой группе параметров показатель weighted average f1-score, представляющий собой взвешенное по доле каждого класса гармоническое среднее значений точности и полноты классификации, вырос на 3% (см. табл. 4). Кроме того, в качестве параметров для работы классификатора свою эффективность показали ЛСП «Болезнь» (пр. 4) (weighted average f1-score + 4% (см. табл. 4)), «Одиночество» (пр. 3) (weighted average f1-score + 4% (см. табл. 4)), ЛСГ абсолютистских слов (*нигде, никогда, ни с кем, всегда, все, везде* и т.д.):

(2) *У моей коллеги вчера во время <u>похорон</u> матери <u>умер</u> отец. Мать коллеги повезли на <u>кладбище</u>, а отца в <u>морг</u>. Врагу не пожелаешь такого* (Страх / Ужас);

(3) *Живу с молодым человеком. Он уехал в командировку на три недели. Я осталась в квартире <u>одна</u>. Думала, что это будет мой мини-отпуск, смогу больше времени посвятить уходу за собой и прочее.*

*Первые четыре дня было неплохо, но дальше стало сложнее. Завтракать одной, ложиться спать в пустой квартире... Я даже не представляла, как одиночество, пусть даже временное, может давить! В тот момент я задумалась о том, как люди живут совершенно одни. Особенно пожилые. Наверно, это крайне непросто, ведь человеку, как известно, нужен человек* (Страдание / Тоска);

(4) *У меня хронический гайморит, болею по два-три раза в год. Когда промываю нос морской водой и оттуда вываливаются большие зеленые сгустки…* (Брезгливость / Отвращение).

Среди *морфологических* маркеров эмоций особо стоит отметить частотность сравнительной и превосходной степеней сравнения у наречий, которая оказалась важной для вербализации эмоции стыда. Наибольший разброс значений наблюдался у двух категорий: наречия в превосходной или сравнительной степени встречаются в 28,8% «стыдных» текстов, а в «гневных» текстах это значение минимально – 15,8%. При добавлении данного параметра точность классификации выросла на 4% именно в первом классе текстов (см. табл. 4).

*Синтаксические маркеры* представлены как определенными синтаксическими структурами, например парцелляциями, так и доминированием у лексемы той или иной синтаксической функции или, наконец, специфической синтаксической комбинаторикой.

Наибольший процент текстов, содержащих парцелляцию, зафиксирован в эмоциональном классе Страх / Ужас (17,464%) (пр. 5), а наименьший – в классах Удивление (10,888%) и Унижение / Стыд (10,421%):

(5) *Меня не стало, просто как будто и не было. Моя старенькая мама просто забыла, что я у нее есть. Она узнает всех – моего брата, своих сестер, даже бывшую одноклассницу. А меня – нет. После смерти папы она живет у меня, никто из родственников ее не навещает, брат даже не звонит спрашивать о ней. Но именно меня она забыла. Первую. И так обидно. Ощущение, что я и не жила вовсе* (Страх / Ужас).

Использование частотности парцелляций в качестве одного из параметров для автоматической обработки текстов повысило точность атрибуции «удивительных» текстов на 1% (см. табл. 4).

Статистический анализ показал, что практически во всех классах текстов лексемы-соматизмы *рука*, *нога*, *глаз* входят в список из 20 лексем, имеющих наибольший статистический «вес» для текстов данного класса. Анализ синтаксических связей данных лексем выявил, что типичные синтаксические позиции данных соматизмов меняются в разных эмоциональных классах текстов. Например, среди всех классов текстов чаще всего (23% от всех синтаксических связей) соматизм *рука* занимает субъектную позицию в классах Злость / Гнев (например, *руки чешутся*, *тянутся*; *рука устает*, *отекает*) и Интерес / Возбуждение (19%) (*руки затряслись* (пр. 6), *опустились*, *рука попала*, *повисла*), а в классе текстов, маркированных эмоцией удивления, таких случаев вообще не зафиксировано, т.е. в «удивительных» текстах из нашего корпуса лексема *рука* никогда не играет роли агенса.

(6) *Муж “сделал предложение” 5 лет назад. Вечер. Вхожу – темно, дорожка из свечей. Ванна с лепестками роз и коробочка в виде сердца. И он стоит, улыбается. У меня руки затряслись, слезы, а он не понимает, почему такая реакция. Беру коробочку, открываю – никак. Я сильнее. И тут она крошится, падает и ШИПИТ в ванной! Оказывается, он просто решил сделать мне сюрприз-релакс, а это была бомба для ванны. Знатных тогда пистонов отхватил, не понимая за что. А через неделю сделал предложение. Напсиховала))* (Интерес / Возбуждение).

А вот соматизм *глаза* наиболее часто занимает субъектную позицию как раз в текстах из подкорпуса Удивление – 21% от числа всех синтаксических связей лексемы в данном подкорпусе на фоне 7,8% в среднем по всем другим корпусам: *глаза светятся*, *сияют*, *округляются*, *вылазят из орбит*, *радуются*, *выкатываются*, *дергаются*, *округляются* и т.д.

Учет синтаксической комбинаторики также позволяет формировать новые достаточно эффективные параметры для атрибуции текстов. Например, синтаксема $ADV_{интенсификатор}$ + ADJ оказалась характерна для текстов класса Удовольствие / Радость – 10,1% от всех текстов класса содержали данную синтаксему (пр. 7), а среднее значение по остальным классам не превышало 4,8%:

(7) *Однажды я потеряла память: обнаружила себя на отдыхе в бунгало рядом с любимым мужчиной (помнила я только знакомство и то, что я влюбилась по уши). Может, это и странно, но никогда я не была так счастлива, как тогда. Ведь при знакомстве я была уверена, что такой мужчина мне, простушке, не светит, а тут я обнаружила, что мы молодожёны, что он меня любит. Будто бы мгновенное исполнение желания…*

Синтаксема $ADV_{интенсификатор}$ + ADV присутствовала в 14,9% текстов из подкорпуса Унижение / Стыд (на фоне среднего значения по оставшимся подкорпусам 4,6%) (пр. 8):

(8) *Все время злился на жену за то, что она непонятно куда тратит огромные суммы денег. Пилил ее за то, что не понимал, куда можно деть 2 000 гривен за день, и так регулярно. А потом заглянул в ее ноут по мелочи, увидел незакрытую вкладку и так узнал, что она постоянно перечисляет деньги в благотворительные фонды и на операции людям. Как же стыдно за свои слова. Я никогда в средствах не нуждался, за все 30 лет жизни даже и не думал помогать людям. Теперь понял, что живу с ангелом.*

Совместное применение в качестве параметров для классификации вербальных маркеров, связанных с синтаксической позицией субъекта, у соматизмов и двух вышеописанных синтаксем дало увеличение гармонического среднего значений точности и полноты классификации на 1%, а по отдельным классам текстов – точность в классе Злость / Гнев увеличилась на 1%, а в классе Стыд / Унижение на 19% (см. табл. 4).

Наибольшую трудность для обнаружения и дальнейшей параметризации представляют *семантические маркеры*. Например, одним из таких маркеров для выявления эмоциональной тональности является использование соматизмов в каритивных конструкциях в подкорпусе текстов Страх / Ужас:

(9) *Сестра очень много болеет. Не просто простудами, а потяжелее. Одной почки нет, легких 1,5 штуки, матку ей вырезали, кардиостимулятор на сердце, нет пальцев на правой ноге. Живет практически в больницах. На днях нашли лейкемию. А сегодня подошла ко мне и призналась, что рада. Рада, что больше не придется все это терпеть. Так больно мне ещё не было* (Страх / Ужас);

Причем маркирующий характер в данном случае заключается не в частотности каритивных конструкций, а в их наличии, поскольку для остальных подкорпусов каритивные конструкции вообще не характерны. Хотя данный маркер, будучи использован изолированно, не показал своей эффективности в качестве параметра для работы классификатора, мы полагаем, что его значимость обнаружится при сочетании с другими маркерами.

Пунктуационные маркеры являются, пожалуй, наиболее простыми для выявления и параметризации. Так, например, пунктуационные знаки «?» и «!», а также «?!» характерны в наибольшей степени для текстов из эмоционального класса Злость / Гнев (26,6%; 64,5% и 7,2% текстов из подкорпуса содержат данные знаки соответственно), а в наименьшей – для эмоциональных классов Брезгливость / Отвращение (6,5% текстов содержат «?») и Страдание / Тоска (10,2% содержат «!» и 0,4% включают знак «?!»). Добавление данных маркеров в качестве параметров к уже показавшей свою эффективность группе параметров дало увеличение гармонического среднего значений точности и полноты классификации на 6%, что значимо (табл. 4).

Наконец, такой сложный маркер, как способ передачи чужого слова. По сути дела, этот маркер уходит своими корнями в нарративно-текстовую категорию диегезиса: если нарратив ведется от первого лица (гомидиегезис), то доминировать будет передача чужой речи в форме косвенной, как правило, интегрированной в придаточное дополнительное предложение (*говорит, что…*); в случае гетеронарратива, где нарратор выдает себя за беспристрастного наблюдателя за действующими как бы по своей воле персонажами, доминирует передача слов другого в виде его прямой речи. При анализе текстов жанра интернет-откровений выяснилось, что тип диегезиса даже в текстах непрофессиональных авторов претерпевает влияние испытываемой эмоции: в текстах из подкорпуса Страдание / Тоска доминирует косвенная речь, а в текстах из подкорпусов Страх / Ужас и Удивление – прямая.

Таблица 4

**Динамика значений гармонического среднего значений точности и полноты классификации при включении различных вербальных маркеров в качестве параметров, подаваемых на вход классификатору**

| | Маркер | weighted average f1-score, % | % увеличения точности в отдельном классе |
|---|---|---|---|
| 1 | $ADV_{интенсификатор}$ + ADJ | +1 | +1 Злость / Гнев<br>+19 Стыд / Унижение |
| 2 | $ADV_{интенсификатор}$ + ADV | | |
| 3 | Соматизмы в субъектной позиции | | |
| 4 | ? | +6% | +9 Страдание / Тоска<br>+7 Страх / Ужас |
| 5 | ! | | +15 Злость / Гнев<br>+2 Страдание / Тоска |
| 6 | ?! | | - |
| 7 | Парцелл. | без изменений | +1 Удивление |
| 8 | ЛСП болезнь | +4 | – |
| 9 | ЛСП смерть | +3 | – |
| 10 | ЛСП одиночество | +4 | |
| 11 | *Терпеть не могу* | без изменений | +0,5 Злость / Гнев |
| 12 | Степени сравнения наречий | без изменений | +4 Стыд / Унижение |
| 13 | *Говорить* | без изменений | +1 Брезгливость / Отвращение, Страх / Ужас, Удивление |

Хотя данный маркер, будучи использован изолированно, не показал своей эффективности в качестве параметра для работы классификатора, мы полагаем, что его значимость будет ощутима при сочетании с другими параметрами, однако их набор еще предстоит установить.

Как видно из данных (табл. 4), наибольшую эффективность для увеличения гармонического среднего значений точности и полноты классификации имеют лексические и пунктуационные маркеры, но для отдельных классов (Стыд / Унижение) более эффективными оказываются, например, синтаксические и морфологические маркеры или пунктуационные (Страдание / Тоска).

### Комбинаторика маркеров

Как показал анализ, маркеры эмоций чувствительны к синтагматическим отношениям: они вступают в различного рода комбинации с другими маркерами, что оказывает влияние на их значимость в качестве параметров, подаваемых «на вход» классификатору в целях увеличения точности производимой классифи-

кации – значимость может увеличиваться или нивелироваться.

Нами отмечено 3 типа комбинаторных отношений вербальных маркеров: 1) взаимное дополнение, приводящее к усилению маркирующей функции; 2) взаимная нейтрализация, сводящая маркирующую функцию к нулю; 3) конкуренция, обусловливающая появление амбивалентности контекстов и противоречивость их интерпретации, что также приводит к снижению точности классификации.

Рассмотрим три примера, иллюстрирующие каждый из вышеупомянутых типов отношений.

Так, последовательное добавление к базовым параметрам, подаваемым «на вход» классификатора, вербальных маркеров (1) частотность «$ADV_{интенсификатор}$ + ADV», (2) присутствие ЛСГ «служебные слова со значением отрицания» (*не, нет, ни*), а затем (3) частотность пунктуационных знаков «!», «?» дало, соответственно, «прирост» гармонического среднего значений точности и полноты классификации на 1%, на 3% и на 3%, а по отдельным эмоциональным классам текстов – до 43%. Однако, когда к этой «успешной» комбинации был добавлен вербальный маркер ЛСГ «вопросительные наречия» (будучи задан изолированно, данный параметр показал значимый прирост), точность классификации в отдельных классах упала до 0 (Стыд / Унижение). Соответственно, именно первые три из упомянутых маркеров «работали» друг на друга, реализуя отношения дополнения. Четвертый же маркер полностью нивелировал маркирующую функцию первых трех. Иными словами, совокупный учет значений первых трех маркеров (параметров) позволил классификатору создать достаточно успешную статистическую модель, которая перестала работать с добавлением следующего.

Как пример конкуренции приведем, несмотря на то что в фокусе нашего внимания находятся вербальные маркеры, взаимодействие вербальных и невербальных маркеров – эмотиконов. Подобное логическое отступление обусловлено иллюстративной силой данных наблюдений. При использовании хорошо зарекомендовавшей себя базовой комбинации вербальных маркеров значение гармонического среднего значений точности и полноты классификации достигло 48%, но, когда к этим маркерам были добавлены эмотиконы, среднее значение упало до 46%, а для отдельных классов текстов «падение» составило около 5%. Анализ примеров таких текстов показал, что маркирующие функции эмотиконов вступают в конкурентные отношения с вербальными маркерами, что провоцирует противоречия, снижающие эффективность выстраиваемой классификатором статистической модели. Например:

(10) *Где-то прочитала, что 15 февраля – день одиноких людей. Кажется, у меня сегодня будет праааздник* 🎉🍷

В данном примере, иллюстрирующем феномен сарказма, мы имеем, с одной стороны, эффективный вербальный маркер эмоционального класса «Страдание / Тоска» – лексическую единицу-репрезентант ЛСП «Одиночество», а с другой – невербальный маркер. Это эмотиконы «праздничное конфетти» и «звон бокала», которые принадлежат семантическому полю «Праздник». Данный невербальный маркер получает и лексическую поддержку – *у меня сегодня будет праааздник*.

## Заключение

Промежуточные итоги проводимого исследования демонстрируют, что в условиях технологической поддержки корпусного инструментария понятие вербального маркера становится основой для проведения автоматической классификации текстов по критерию их эмоциональной тональности. Вербальные маркеры, наблюдаемые на большой коллекции размеченных по принципу «объект – ответ» (текст – тональность) текстов, базируются на единицах и структурах, принадлежащих разным уровням языковой системы: лексическому, синтаксическому, морфологическому, пунктуационно-графическому, а также уровню текстовой реализации. Анализ показывает, что при учете их комбинаторных особенностей вербальные маркеры становятся статистически значимыми и эффективными инструментами для установления корреляций между речевым поведением человека, с одной стороны, и его эмоциональными переживаниями, «запускаемыми» на физиологическом уровне тремя моноаминами, согласно Г. Лёвхейму, – с другой. Описание системы таких корреляций будет иметь не только практическое значение для создания технологии автоматической классификации интернет-текстов, но и теоретическую значимость, поскольку послужит доказательством единства когнитивной природы языка и эмоций, объединенных гиперсетью мозга.

## ПРИМЕЧАНИЕ

[1]Статистическая мера, используемая для оценки важности слова в контексте документа, являющегося частью коллекции документов или корпуса. Вес некоторого слова пропорционален количеству употребления этого слова в документе и обратно пропорционален частоте употребления слова в других документах коллекции.

## ЛИТЕРАТУРА

1. Жаботинская С.А. Имя как текст: концептуальная сеть лексического значения (анализ имени эмоции) // Когниция, коммуникация, дискурс. 2013. № 6. С. 47–76. URL: http://sites.google.com/site/cognitiondiscourse/ (дата обращения: 04.03.2019). DOI: 10.26565/2218-2926-2013-06-04
2. Анохин К.В. Когнитом – гиперсетевая модель мозга // Материалы XVII Всероссийской научно-технической конференции Нейроинформатика – 2015. URL: http://neuroinfo.mephi.ru/conf/Content/Presentations/Anokhin2015.pdf (дата обращения: 14.02.2019).
3. Масевич А.Ц., Захаров В.П. Методы корпусной лингвистики в исторических и культурологических исследованиях // Компьютерная

лингвистика и вычислительные онтологии : сб науч. ст. Труды XIX Междунар. объединённой науч. конф. «Интернет и современное общество» (IMS-2016). СПб. : Университет ИТМО, 2016. С. 24–43.
4. Подлесская В.И., Кибрик А.А. Дискурсивные маркеры в структуре устного рассказа: опыт корпусного исследования // Компьютерная лингвистика и интеллектуальные технологии : по материалам ежегодной Междунар. конф. «Диалог 2009». М. : РГГУ, 2009. Вып. 8 (15). С. 390–395.
5. Fraser B. What Are Discourse Markers? // Journal of Pragmatics. 1999. 31 (7). P. 931–952.
6. Knott A., Sanders T. The Classification of Coherence Relations and Their Linguistic Markers: An Exploration of Two Languages // Journal of Pragmatics. 1998. 30 (2). P. 135–175. URL: https://doi.org/10.1016/S0378-2166(98)00023-X
7. Furko P. The Pragmatic Marker – Discourse Marker Dichotomy Reconsidered: The Case of 'Well' and 'Of Course'. Debrecen : Debrecen University Press, 2007. 136 p.
8. Дементьев В.В. Теория речевых жанров. М. : Знак, 2010. 600 с.
9. Белова Е.В. Речевые маркеры бытового конфликта // Вестник ТвГУ. Сер. Филология. 2017. № 2. С. 157–161.
10. Потапова Р.К., Потапов В.В. Временные корреляты эмоции как специфические индивидуальные параметры идентификации говорящего в судебной фонетике // Акустика речи и прикладная лингвистика: Ежегодник Российского акустического общества / отв. ред. Р.К. Потапова. М., 2002. Вып. 3. С. 3–13.
11. Pickering L. et. al. Prosodic Markers of Saliency in Humorous Narratives // Discourse Processes. 2009. 46 (6). P. 516–540.
12. Зубова И.И. Автоматическая идентификация конфликтной речевой ситуации в письменном тексте // Инновации в науке и практике : сб. ст. по материалам VIII междунар. науч.-практ. конф. 2018. С. 35–42.
13. Arciuli J., Mallard D., Villar G. "Um, I can tell you're lying": Linguistic markers of deception versus truth-telling in speech // Applied Psycholinguistic. 2010. Vol. 31. P. 397–411.
14. Фомин А.Г., Якимова Н.С. Тактики и маркеры вербальной агрессии в коммуникативном поведении россиян и американцев (по материалам речеситуативного исследования) // Сибирский филологический журнал. 2012. № 2. С. 197–207.
15. Al-Mosaiwi M., Johnstone T. Linguistic markers of moderate and absolute natural language // Personality and Individual Differences. 2018. Vol. 134. P. 119–124. URL: https://doi.org/10.1016/j.paid.2018.06.004
16. Cohen K. et. al. Detecting Linguistic Markers for Radical Violence in Social Media // Terrorism and Political Violence. 2014. 26 (1). P. 246–256.
17. Колосов Я.В. Лингвистические корреляты эмоционального состояния «страх» в русской и английской речи: формирование базы данных : дис. … канд. филол. наук. М., 2004. 214 с.
18. Al-Mosaiwi M., Johnstone T. In an Absolute State: Elevated Use of Absolutist Words Is a Marker Specific to Anxiety, Depression, and Suicidal Ideation // Clinical Psychological Science. 2018. Vol. 6, is. 4. P. 529–542. URL: https://doi.org/10.1177/2167702617747074
19. Колмогорова А.В., Горностаева Ю.А., Калинин А.А. Разработка компьютерной программы автоматического анализа и классификации поляризованных политических текстов на английском языке по уровню их манипулятивного воздействия: практические результаты и обсуждение // Политическая лингвистика. 2017. № 4 (64). С. 67–75.
20. Raza M.S., Qamar U. Understanding and Using Rough Set Based Feature Selection: Concepts, Techniques and Applications. Singapore : Springer, 2017. 194 p.
21. Сарбасова А.Н. Исследование методов сентимент-анализа русскоязычных текстов // Молодой ученый. 2015. № 8. С. 143–146.
22. Das S., Chen M. Yahoo! for Amazon: Extracting market sentiment from stock message boards // Proceedings of the Asia Pacific Finance Association Annual Conference (APFA). 2001. P. 1–16.
23. Pang B., Lee L., Vaithyanathan Sh. Thumbs up? Sentiment classification using machine learning techniques // Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). 2002. P. 79–86.
24. Hogenboom A., Frasincar F., Jong F., Kaymak U. Polarity Classification Using Structure-Based Vector Representations of Text // Decison Support Systems. 2015. Vol. 74. P. 46–56.
25. Banea C., Mihalcea R., Wiebe J., Hassan S. Multilingual Subjectivity Analysis Using Machine Translation // Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing. 2008. P. 127–135.
26. Lucas G.M., Gratch J., Malandrakis N., Szablowski E., Fessler E., Nichols J. GOAALLL!: Using Sentiment in the World Cup to Explore Theories of Emotion // Image and Vision Computing. 2017. P. 58–65. doi:10.1016/j.imavis.2017.01.006
27. Staiano J., Guerini M. DepecheMood: A Lexicon for Emotion Analysis from Crowd-Annotated News // Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers). Baltimore, Maryland : Association for Computational Linguistics, 2014. P. 427–433.
28. Loukachevitch N.V., Blinov P.D., Kotelnikov E.V., Rubtsova Y.V., Ivanov V.V., Tutubalina E.V. SentiRuEval: Testing Object-Oriented Sentiment Analysis Systems in Russian // Computational Linguistics and Intellectual Technologies: Proceedings of the Annual International Conference "Dialogue 2015". Moscow, 2015. Vol. 14 (2). P. 3–15.
29. Loukachevitch N.V., Rubtsova Y.V. SentiRuEval-2016: Overcoming Time Gap and Data Sparsity in Tweet Sentiment Analysis // Computational Linguistics and Intellectual Technologies: Proceedings of the Annual International Conference "Dialogue 2016". Moscow, 2016. Vol. 15. P. 416–426.
30. Alm C.O., Rot D., Sproat R. Emotions from Text: Machine Learning for Text-based Emotion Prediction // Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing. Vancouver, 2005. P. 579–586.
31. Thelwall M., Buckley K., Paltoglou G., Cai D. Sentiment Strength Detection in Short Informal Text // Journal of the American Society for Information Science and Technology. 2010. Vol. 61 (12). P. 2544–2558.
32. Socher R., Perelygin A., Wu J.Y., Chuang J., Maning Ch. Ng A.Y., Potts Ch. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank // Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2013. P. 1631–1642.
33. Chaffar S., Inkpen D. Using a Heterogeneous Dataset for Emotion Analysis in Text // Canadian AI 2011: Advances in Artificial Intelligence. Lecture Notes in Computer Science. Berlin, Heidelberg : Springer, 2011. 6657. P. 62–67.
34. Lövheim H. A new three-dimensional model for emotions and monoamine neurotransmitters // Medical Hypotheses. 2012. № 78. P. 341–348.
35. 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
36. Захаров В.П., Богданова С.Ю. Корпусная лингвистика. Иркутск : ИГЛУ, 2011. 161 с.

Статья представлена научной редакцией «Филология» 25 сентября 2019 г.

**The Types and Combinatorics of Verbal Markers of Different Emotional Tonalities in Russian-Language Internet Texts**

*Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*, 2019, 448, 48–58.
DOI: 10.17223/15617793/448/6

**Anastasia V. Kolmogorova,** Siberian Federal University (Krasnoyarsk, Russian Federation). E-mail: nastiakol@mail.ru
**Alexander A. Kalinin,** Siberian Federal University (Krasnoyarsk, Russian Federation). E-mail: verbalab@yandex.ru
**Alina V. Malikova,** Siberian Federal University (Krasnoyarsk, Russian Federation). E-mail: malikovaav1304@gmail.com

**Keywords:** emotion; Internet texts; sentiment analysis; verbal markers; machine learning; cognition.

The article aims to present theoretical grounds for the concept of the verbal marker, proposes a typology of such markers and summarizes observations about the impact of verbal marker combinations on the accuracy of the computer classifier designed to assign Internet texts in Russian to different emotional classes of texts. As a result of the complex analysis of the up-to-date information based on the international scholarship, the authors of the article give a definition of the term "verbal marker". The latter is a unit or structure belonging to the one of the linguistic system levels, available to parametrization and appearing in the text as an indicator of processes, covert from direct observation, occurring in human cognitive system. According to the level of the linguistic system in which the unit or the structure with the marking function is localized, the authors propose to distinguish the following types of verbal markers relevant for the analysis of written texts: lexical markers, morphological markers, syntactical markers, semantic markers, punctuation markers and, finally, textual markers. To prove the practical viability of the conception, the authors applied it in their project conducted in the field of sentiment analysis and supposed to resolve the problem of attributing an Internet text in Russian to a particular class of emotions. The authors are deeply interested in the emotional tonality of Internet texts because they became one of the most common forms of texts in Russian, and the technology of their automatic assessment has the clearest commercial and social prospects. The concept of the classifier is based on eight emotions detected by Swedish neuroscientist H. Lövheim in relation to some specific combinations of the levels of monoamines in the limbic system of human brain. To build the classifier, the authors used the method of supervised machine learning which demands the sample selection and the extraction of features. As the data, the authors took 15,000 emotionally rich fragments of 60–80 words selected from the Russian social network VK public Podslushano [Overheard]. For sample extraction, firstly, the authors mapped eight emotional classes of Lövheim's model to a range of hashtags used by public group editors to publish users' posts. Secondly, each text from the sample was assessed by three informants on the crowdsourcing platform. After that, the preliminary classified data went through the expert linguistic analysis made by using multiple tools offered by the linguistic corpus manager Sketch Engine. This analysis led the authors to the extraction of a feature set for the SVM algorithm-based classifier. The analysis of eight texts classes by methods of corpus linguistics and the use of prototype of the classifier showed the dynamics of the weighted average f1-score while incorporating different verbal markers as the classifier features. Thus, the results of the research showed the greatest efficiency of lexical and punctuation markers. However, syntactical and morphological markers also proved to be effective for some classes of emotions. In addition, the authors stress the relevance of marker combinations for accuracy of the statistical models created by the classifier. At present, the f1-score of the classifier in different emotional classes of texts varies from 30% to 50%, which is comparable with the results showed by classifiers built for other languages.

**REFERENCES**


1. Zhabotinskaya, S.A. (2013) The name as a text: conceptual network of lexical meaning (analysis of the name of emotion). *Kognitsiya, kommunikatsiya, diskurs – Cognition, communication, discourse*. 6. pp. 47–76. [Online]. Available from: http://sites.google.com/site/cognitiondiscourse/. (Accessed: 04.03.2019). (In Russian). DOI: 10.26565/2218-2926-2013-06-04
2. Anokhin, K.V. (2015) [Cognitom, a hypernetwork model of the brain]. *Neyroinformatika – 2015* [Neuroinformatics – 2015]. Proceedings of the 17th All-Russian Conference. [Online]. Available from: http://neuroinfo.mephi.ru/conf/Content/Presentations/Anokhin2015.pdf. (Accessed: 14.02.2019). (In Russian).
3. Masevich, A.Ts. & Zakharov, V.P. (2016) Metody korpusnoy lingvistiki v istoricheskikh i kul'turologicheskikh issledovaniyakh [Methods of corpus linguistics in historical and cultural studies]. In: *Komp'yuternaya lingvistika i vychislitel'nye ontologii* [Computer linguistics and computational ontologies]. St. Petersburg: Universitet ITMO. pp. 24–43.
4. Podlesskaya, V.I. & Kibrik, A.A. (2009) Diskursivnye markery v strukture ustnogo rasskaza: opyt korpusnogo issledovaniya [Discursive markers in the structure of an oral narrative: the experience of corpus research]. In: *Komp'yuternaya lingvistika i intellektual'nye tekhnologii* [Computer Linguistics and Intelligent Technologies]. 8 (15). Moscow: Russian State University for the Humanities. pp. 390–395.
5. Fraser, B. (1999) What are Discourse Markers? *Journal of Pragmatics*. 31 (7). pp. 931–952.
6. Knott, A. & Sanders, T. (1998) The Classification of Coherence Relations and Their Linguistic Markers: An Exploration of Two Languages. *Journal of Pragmatics*. 30 (2). pp. 135–175. DOI: 10.1016/S0378-2166(98)00023-X
7. Furko, P. (2007) *The Pragmatic Marker – Discourse Marker Dichotomy Reconsidered: The Case of 'Well' and 'Of Course'*. Debrecen: Debrecen University Press.
8. Dement'ev, V.V. (2010) *Teoriya rechevykh zhanrov* [Theory of Speech Genres]. Moscow: Znak.
9. Belova, E.V. (2017) Verbal markers of conflict. *Vestnik TvGU. Seriya Filologiya*. 2. pp. 157–161. (In Russian).
10. Potapova, R.K. & Potapov, V.V. (2002) Vremennye korrelyaty emotsii kak spetsificheskie individual'nye parametry identifikatsii govoryashchego v sudebnoy fonetike [Temporal correlates of emotions as specific individual parameters of speaker identification in judicial phonetics]. In: Potapova, R.K. (ed.) *Akustika rechi i prikladnaya lingvistika* [Speech Acoustics and Applied Linguistics]. Vol. 3. Moscow: Moscow State Linguistic University. pp. 3–13.
11. Pickering, L. et. al. (2009) Prosodic Markers of Saliency in Humorous Narratives. *Discourse Processes*. 46 (6). pp. 516–540.
12. Zubova, I.I. (2018) [Automatic identification of a conflicting speech situation in a written text]. *Innovatsii v nauke i praktike* [Innovations in science and practice]. Proceedings of the 8th International Conference. Barnaul: Izdatel'stvo "Dendra". pp. 35–42.
13. Arciuli, J., Mallard, D. & Villar, G. (2010) "Um, I can tell you're lying": Linguistic markers of deception versus truth-telling in speech. *Applied Psycholinguistic*. Vol. 31. pp. 397–411.
14. Fomin, A.G. & Yakimova, N.S. (2012) The tactics and markers of verbal aggression in the communicative behavior of Russians and Americans. *Sibirskiy filologicheskiy zhurnal – Siberian Journal of Philology*. 2. pp. 197–207. (In Russian).
15. Al-Mosaiwi, M. & Johnstone, T. (2018) Linguistic markers of moderate and absolute natural language. *Personality and Individual Differences*. 134. pp. 119–124. DOI: 10.1016/j.paid.2018.06.004
16. Cohen, K. et. al. (2014) Detecting Linguistic Markers for Radical Violence in Social Media. *Terrorism and Political Violence*. 26 (1). pp. 246–256.
17. Kolosov, Ya.V. (2004) *Lingvisticheskie korrelyaty emotsional'nogo sostoyaniya "strakh" v russkoy i angliyskoy rechi: formirovanie bazy dannykh* [Linguistic correlates of the emotional state of "fear" in Russian and English speech: forming a database]. Philology Cand. Diss. Moscow.
18. Al-Mosaiwi, M. & Johnstone, T. (2018) In an Absolute State: Elevated Use of Absolutist Words Is a Marker Specific to Anxiety, Depression, and Suicidal Ideation. *Clinical Psychological Science*. 6 (4). pp. 529–542. DOI: 10.1177/2167702617747074
19. Kolmogorova, A.V., Gornostaeva, Yu.A. & Kalinin, A.A. (2017) Computer program design for classifying English polarized political texts by their manipulative impact: results and discussion. *Politicheskaya lingvistika – Political Linguistics*. 4 (64). pp. 67–75. (In Russian).
20. Raza, M.S. & Qamar, U. (2017) *Understanding and Using Rough Set Based Feature Selection: Concepts, Techniques and Applications*. Singapore: Springer.
21. Sarbasova, A.N. (2015) Issledovanie metodov sentiment-analiza russkoyazychnykh tekstov [The study of the methods of sentiment analysis of Russian-language texts]. *Molodoy uchenyy – Young Scientist*. 8. pp. 143–146.

22. Das, S. & Chen, M. (2001) Yahoo! for Amazon: Extracting market sentiment from stock message boards. *Proceedings of the Asia Pacific Finance Association Annual Conference (APFA).* Bangkok. 4 April 2001. Bangkok: [s.n.]. pp. 1–16.
23. Pang, B., Lee, L. & Vaithyanathan, Sh. (2002) Thumbs up? Sentiment classification using machine learning techniques. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP).* Philadelphia, PA. 6–7 July 2002. Stroudsburg, PA: Association for Computational Linguistics. pp. 79–86.
24. Hogenboom, A. et al. (2015) Polarity Classification Using Structure-Based Vector Representations of Text. *Decision Support Systems*. 74. pp. 46–56.
25. Banea, C. et al. (2008) Multilingual Subjectivity Analysis Using Machine Translation. *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing.* Honolulu, Hawaii. 25–27 October 2008. Stroudsburg, PA: Association for Computational Linguistics. pp. 127–135.
26. Lucas, G.M. et al. (2017) GOAALLL!: Using Sentiment in the World Cup to Explore Theories of Emotion. In: Soleymani, M., Schuller, B. & Chang, Sh.-F. (eds) *Image and Vision Computing*. pp. 58–65. DOI: 10.1016/j.imavis.2017.01.006
27. Staiano, J. & Guerini, M. (2014) DepecheMood: A Lexicon for Emotion Analysis from Crowd-Annotated News. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers).* Baltimore, Maryland: Association for Computational Linguistics. pp. 427–433.
28. Loukachevitch, N.V. et al. (2015) SentiRuEval: Testing Object-Oriented Sentiment Analysis Systems in Russian. Computational Linguistics and Intellectual Technologies. *Proceedings of the Annual International Conference "Dialogue 2015*". 14 (2). Moscow. 27–30 May 2015. Moscow: Russian State University for the Humanities. 14 (2). pp. 3–15.
29. Loukachevitch, N.V. & Rubtsova, Y.V. (2016) SentiRuEval-2016: Overcoming Time Gap and Data Sparsity in Tweet Sentiment Analysis. Computational Linguistics and Intellectual Technologies. *Proceedings of the Annual International Conference "Dialogue 2016"*. Moscow. Vol. 15. pp. 416–426.
30. Alm, C.O., Rot, D. & Sproat R. (2005) Emotions from Text: Machine Learning for Text-based Emotion Prediction. *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing.* Vancouver, British Columbia, Canada. 6–8 October 2005. Vancouver: Association for Computational Linguistics. pp. 579–586.
31. Thelwall, M. et al. (2010) Sentiment Strength Detection in Short Informal Text. *Journal of the American Society for Information Science and Technology*. 61 (12). pp. 2544–2558.
32. Socher R. et al. (2013) Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. *Proceedings of the Conference on Empirical Methods in Natural Language Processing.* Seattle, Washington, USA. 18–21 October 2013. Stroudsburg, PA: Association for Computational Linguistics pp. 1631–1642.
33. Chaffar, S. & Inkpen, D. (2011) Using a Heterogeneous Dataset for Emotion Analysis in Text. *Canadian AI 2011: Advances in Artificial Intelligence. Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer. pp. 62–67.
34. Lövheim, H. (2012) A new three-dimensional model for emotions and monoamine neurotransmitters. *Medical Hypotheses.* 78. pp. 341–348.
35. 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
36. Zakharov, V.P. & Bogdanova, S.Yu. (2011) *Korpusnaya lingvistika* [Corpus Linguistics]. Irkutsk: Irkutsk State Linguistic University.

Received: 25 September 2019