

Интеллектуальный анализ проявлений вербальной агрессивности в текстах сетевых сообществ¹

Аннотация. Обсуждаются вопросы, связанные с автоматическим выявлением в текстах социальных сетей проявлений вербальной агрессивности. Проблема выраженности вербальной агрессивности рассматривается в контексте формирования сетевых сообществ, предполагается, что ценностные приоритеты выступают в качестве основания группообразования и присоединения к сообществу. Проанализирована возможность применения средств интеллектуального анализа и методов оценки субъективного содержания текста (Opinion Mining). Разработан психолингвистический анализатор PLATIn, показаны его возможности в дифференцированном определении различных форм проявления вербальной агрессивности на массиве текстов нескольких сетевых сообществ. Выявлено своеобразие вербальной агрессивности в текстах сообществ, связанное с их ориентацией на определенные ценности.

Ключевые слова: ценности, вербальная агрессия, интеллектуальный анализ текстов, автоматическое определение тональности текста.

Введение

Благодаря развитию современных практик коммуникации, — в частности, общения, опосредованного интернетом, — возникают не только новые предметы (или их новые формы), но и средства исследования таких наук, как социология, психология, психолингвистика. Разносторонний анализ дискурса социальных сетей становится базовым инструментом для описания и моделирования содержания, структуры и динамики разнообразных явлений на уровне социума и составляющих его групп, процессов социализации индивида и формирования его «социального капитала» и т.п. [5, 15, 38, 47]. В частности, развитие компьютерных технологий создает качественно новые возможности получения информации о происходящих в виртуальном пространстве процессах возникновения и жизнедеятельности социальных общностей. Формы существования виртуальных сообществ отражают особенности опосредованного интернетом общения, которые активно исследуются начиная с момента перехода Сети в статус фактора жизненной практи-

ки большого количества людей [2, 7, 21]. Среди многочисленных привнесений в процесс взаимодействия с социальной реальностью, протекающего в Сети, мы выделяем в качестве предмета настоящего исследования специфические условия вхождения субъекта в виртуальное сообщество.

Как и любая социальная группа, которой оно является по своей сути, сетевое сообщество объединяется благодаря наличию у участников разделяемых целей и стоящих за целями общих ценностей. Именно ценности, как и в «реальных» группах, определяют содержание и критерии успешности совместной деятельности виртуальной группы, а также регулирующие ее нормы, иерархию и формы контроля, и, в конечном счете, замыкают те границы, которые разделяют мир на Мы и Они и создают коллективную идентичность.

Ценности, согласно концепции Д.А.Леонтьева, являются высшей формой мотивационных и смыслообразующих структур личности, которые самим субъектом-носителем переживаются как идеалы — конечные ориентиры желательного состояния дел. При этом такие идеалы для

¹ Работа выполнена при поддержке РФФИ (грант № 13-06-00641а)

субъекта представляются не воплощением его индивидуального желания, а обладающими статусом объективной желательности, желательность для всех вообще. Стабильность структуры личностных ценностей определяется тем, что отраженные в них отношения человека с миром перерабатываются и обобщаются в соответствии с совокупным опытом социальной группы. По обобщающей формулировке Д.А. Леонтьева, в ценностях человек всегда не один [25]. Итак, личностные ценности как по своему происхождению, так и по своему содержанию, связывают человека с социальной реальностью. Доступные человеку формы взаимодействия с различными социальными образованиями делают возможным и само существование личностных ценностей благодаря поддержке единомышленников, и их реализацию, которая подразумевает достижение «всеобщего блага», как его понимает или ощущает носитель. Эта сущностная «социальность» объясняет высокий мотивирующий потенциал личностных ценностей, которые требуют от своего носителя активности по поиску и/или организации поддерживающего социального окружения. Благодаря особенностям сетевого общения человек получает возможность, сохраняя личную безопасность, максимально открыто декларировать свои ценности и по собственному желанию отбирать партнеров для взаимодействия. Виртуальная среда, таким образом, становится пространством ценностной самореализации современного человека, протекающей, в частности, в виде присоединения к уже существующим сетевым сообществам с соответствующими ценностными приоритетами или, если таковые сообщества не обнаруживаются, образования новых.

В силу того, что для Сети основной формой совместной деятельности является коммуникация, ведущую роль в процессах группообразования играют характерные для сообщества особенности его дискурсивной практики. Наличие «своего» дискурса рассматривается как один из конституирующих признаков виртуальной социальной группы. Как показано в работах [28] и [29], большинство успешных сообществ возникает вокруг общей темы или интересующего всех членов группы вопроса. В процессе своего развития сообщество вырабатывает собственный дискурс, правила общения, разнообразные мемы, мифы, создает свой понятийный аппарат и групповой лексикон, в частности, собствен-

ные слова-маркеры, которые помогают членам сообщества с высокой вероятностью определять друг друга. Наличие такого рода коммуникативных особенностей может приводить к тому, что общение членов виртуального сообщества становится непонятно внешним наблюдателям. Для того, чтобы быть принятым в группу с уже сложившимися правилами и языком общения, ее новый член обязан следовать общегрупповой логике и «включиться» в дискурс сообщества. «Выключение» из общего дискурса автоматически ведет за собой перенесение данного члена группы на периферию.

Среди множества параметров, характеризующих дискурсивные практики сетевых сообществ, в качестве предмета дальнейшего рассмотрения мы выделяем такой показатель, как выраженность агрессивности в порождаемых сообществом текстах.

Агрессия вообще – явление функционально неоднозначное. Как на групповом, так и на индивидуальном уровне она может выполнять прямо противоположные по своим последствиям функции – разрушительную и созидательную. В первом случае речь идет о таком общем свойстве агрессивности, как совершаемое субъектом агрессии нарушение границ, несанкционированное вторжение на территорию, принадлежащую объекту агрессии. При этом границы могут быть как совершенно реальными: физическими, территориальными, – так и понимаемыми метафорически – например, «границы личности». Созидательная же функция заключается в том, что при известных условиях и на определенных этапах развития агрессия является фактором внутренней интеграции, определения и отстаивания границ, и, таким образом, способствует сохранению индивидуальности и идентичности – как на уровне образований социального масштаба, так и отдельной личности [3, 19, 30].

1. Психолингвистика вербальной агрессивности

Исследования вербальной агрессии показывают, что выполняемые ею коммуникативные функции также неоднозначны. С одной стороны, указывается, что речевая агрессия (обвинение, грубое требование, грубый отказ, угроза, злопожелание, брань, клевета, насмешка, кол-

кость, упрек) деформирует коммуникативное пространство, мешает полноценному общению, поглощая и коммуникативные ресурсы агрессора, и эмоциональные силы жертвы агрессии, вызывая у последней страх и ответную враждебность [23, 42]. С другой же стороны, выделяется так называемая «невраждебная» речевая агрессия как такой атрибут неофициального дружеского общения, особенно мужского, который служит для участников признаком «нормальности отношений». Инвективы, послылы с упоминанием «непристойных» частей тела, злопожелания, угрозы, и т.п., формально соответствующие признакам вербальной агрессии, в контексте дружеского общения не являются ею по сути [36]. Эти же средства вербальной агрессии, направленные вовне группы, по отношению к не принадлежащим к ней людям, призваны защитить на коммуникативном уровне границы общности.

Возвращаясь к проблеме ценностного самоопределения субъекта в виртуальном пространстве, мы можем высказать предположение, что тот или иной уровень выраженности вербальной агрессивности текстов может становиться фактором, определяющим выбор – присоединиться или нет к определенному сетевому сообществу. При этом воспринимаемый уровень агрессивности текстов не обязательно прямо соответствует содержанию тех целей и ценностей, которые служат объединяющим началом для данной группы. В ряде случаев может быть так, что сами по себе проповедуемые ценности связаны с насилием, разрушением и противостоянием кому-либо или чему-либо. Но возможна и такая ситуация, когда высокий уровень вербальной агрессии отражает сложившиеся в сообществе коммуникативные нормы, не совпадающие с конвенциональными, или такой этап ее становления, когда актуальной является задача замыкания собственных границ и активного противопоставления возникающего «мы» внешним «другим».

Для определения уровня вербальной агрессивности необходимо, прежде всего, выявить средства ее выражения. В работе Н.Е. Петровой и Л.В. Рацибурской [32] выделяется две основные группы использующихся при агрессивном речевом поведении языковых средств – лексические и дискурсивные.

1. В первую группу объединяются лексические средства русского языка, выражающие не-

гативную оценку. С помощью этих средств, как правило, выражается активная, открытая речевая агрессия. При этом лексические средства русского языка могут выражать негативную оценку сами по себе или под влиянием метафорического контекста:

1.1. Прямое, наиболее простое и наиболее распространенное орудие вербальной агрессии – оценочная, в том числе инвективная (бранная), обценная (непристойная) и стилистически сниженная (разговорная, просторечная) лексика.

1.2. Чрезвычайно экспрессивное и потому стилистически «опасное» средство выражения негативной оценки – жаргонная лексика, арготизмы.

1.3. Оказициональные слова – неологизмы, оценочность которых может быть связана как с мотивирующей (производящей) базой (например, собственные имена лиц, слова с негативным денотатом и т.д.), так и со словообразовательными средствами (например, маркированные суффиксы *-щина, -ость, -ация, -изм, -ист, -ец, -оид*, размерно-оценочные суффиксы, префиксы *а-, без-, анти-, контр-, де-, квази-, псевдо-, экс-* и т.д.).

1.4. Агрессивные сравнения и метафоры.

1.5. Немотивированное использование иноязычных элементов в целях агрессивного воздействия на читателя, провоцирующее возникновение у него чувства неполноценности из-за непонятности изложения; с другой стороны, заимствования создают интеллектуальную интервенцию в сфере национальной речемысли, происходит грубое вторжение в русскую этнокультурную картину мира чужеродной ментальности.

2. Вторую группу образуют дискурсивные средства вербальной агрессии, формирующиеся непосредственно в тексте-дискурсе. Их агрессивный потенциал заключен не столько в самих словах, сколько связан с принятыми нормами успешного общения, знаниями участников общения о мире, проблемами адекватного истолкования слов. Дискурсивные средства используются преимущественно для не прямой, скрытой речевой агрессии. К ним относятся:

2.1. Языковая демагогия (сознательное нарушение словесных пресуппозиций, постулатов успешного общения, использование речевых импликатур).

2.2. Тенденциозное использование негативной информации, прежде всего, перегруженность текста негативной информацией. Указывая на пороки, автор подчеркивает их всеохватность. Для этого широко используются языковые средства обобщения (например, безысключительная лексика: *каждый, все, никто* и пр.).

2.3. Интертекстуальность. Агрессия осуществляется с помощью цитирования, обращения к вербальным прецедентным феноменам, которые для разделяющих знание о них представителей социума связаны с определенными эмоциями и оценками: прецедентная ситуация, прецедентный текст, прецедентное имя. Основной целью цитирования в агрессивном тексте становится ироничность, насмешка, ерничаение. При этом базой для ироничного цитирования становятся культурно-исторические интертексты, которые по морально-нравственным критериям не должны подвергаться осмеянию.

Некоторые лексические и фразеологические единицы, нейтральные в своем основном значении, при определенных условиях могут становиться средствами вербальной агрессии. Реализация конфликтогенного потенциала слова, как показывают работы [6, 20, 37], зависит от таких переменных, как социальный статус коммуникантов или публичность ситуации коммуникации – и в ряде случаев конфликтогенным становится, например, упоминание в публичном тексте имени человека, его национальности и др.

Очевидно, что с точки зрения возможной формализации, перечисленные средства выражения вербальной агрессии оказываются весьма неравнозначными: для компьютерной лингвистики лексический анализ – направление, хорошо разработанное, а выявление и интерпретация, например, метафор или прецедентных текстов – все еще задача завтрашнего дня.

2. Методы выявления эмоционального окрашенности текстов

Как уже было сказано, основным направлением действия вербальной агрессии является эмоциональная сфера жертвы. Иными словами, одним из аспектов своего существования вербальная агрессия предстает как явление эмоцио-

нальное. В связи с этим был проведен анализ специальной литературы, посвященной существующим в настоящее время способам автоматического анализа текста, предназначенным для выявления субъективного содержания, или эмоциональной тональности, сетевых текстов.

В работе А.А. Давыдова [12] приведены основные характеристики активно развивающегося направления прикладных исследований, объединяемых названием Opinion Mining². Общая цель этих исследований – создание компьютерных систем для интеллектуального автоматического извлечения из текстов субъективной информации (мнений, оценочных суждений, аттитудов, эмоций, чувств, верований и т.д.). Созданные системы Opinion Mining используются для автоматической оценки в режиме реального времени материалов электронных СМИ, дискуссионных форумов, сообщений блогеров и т.д. Одним из основных методов анализа, который широко используется в Opinion Mining для практических приложений, является Sentiment Analysis. С его помощью осуществляется автоматический анализ представленного в текстовых сообщениях отношения к какому-либо объекту (персоне, сообщению СМИ, событию, организации и т.д.) в терминах позитивных, негативных, нейтральных оценок, благоприятных и неблагоприятных мнений, количественных индексов и т.д. Существуют различные классы Sentiment Analysis, например: Dynamics Sentiment Analysis, Visual Sentiment Analysis, Large-Scale Sentiment Analysis, Deeper Sentiment Analysis, Online Sentiment Analysis, Multilingual Sentiment Analysis и т.д.

Для реализации Sentiment Analysis в системах Opinion Mining используется следующие группы методов.

– Методы Natural Language Processing: (NLP) – Latent Semantic Analysis (LSA), в частности, Probabilistic Latent Semantic Analysis (PLSA), Deep language analysis, Semantic Orientation Method, Semantic Structure Networks, «bag of words» и т.д.

– Методы Computational Linguistics («вычислительной» лингвистики) – в частности, Pointwise Mutual Information method, Morphological analysis, Lexical relations analysis,

² Подробный обзор работ по данной тематике в [48]; современное положение освещено, в частности, в сборнике [46].

системы автоматического машинного перевода для Multilingual Sentiment Analysis и т.д.

– Методы Text Mining – анализ соответствий, ассоциаций и т.д.

– Методы Automatic Text Analysis – Automatic Semiotic Analysis, Semantic Spaces Analysis и т.д.

– Методы математической статистики – Pattern analysis в Corpus-based approaches и Dictionary-based approaches (словари синонимов и антонимов), Hidden Markov Models, Bootstrapping Method, Gloss Classification method и т.д.

– Методы Artificial Intelligence – Text Understanding, Soft Computing («мягкие вычисления»), Affective Computing («эмоциональные вычисления»), Meaning Based Computing (MBC), нейронные сети класса Support Vector Machines (SVM), Heuristic Rule-Based Reasoning и т.д. [12].

В работе [31] описаны следующие характеристики автоматического определения тональности текста. Его цель заключается в выделении тех фрагментов текста, которые выражают позитивную или негативную эмоциональность по отношению к объекту эмоциональной оценки. Объект эмоциональной оценки может быть задан как один в целом для текста (с учетом его синонимических и анафорических употреблений), так и определяться в предложениях как любое имя собственное или нарицательное. Тональность текста определяется тремя факторами: 1) субъект тональности (автор текста или приведенной автором цитаты); 2) собственно тональная оценка (позитив/нейтрально/негатив); 3) объект тональности. Основные методы определения тональности текста:

1. Анализ текста методами векторного анализа (часто с применением n-граммных моделей), сравнение с ранее размеченным эталонным корпусом по выбранной мере близости и отнесение (классификация) текста к негативу или позитиву на основании полученного результата сравнения. Этот метод требует наличия предварительно размеченного эталонного корпуса, на основе которого происходит обучение алгоритма сравнения. Существенными недостатками такого подхода оказываются увеличение трудоемкости и ограничение разнообразности корпуса (т.е. неполнота лексического покрытия), что приводит к потере точности. К тому же данный метод не позволяет провести

глубокий анализ текста, то есть выявить и показать эмотивность на уровне предложения.

2. Поиск эмотивной лексики (лексической тональности) в тексте по заранее составленным тональным словарям (спискам паттернов) с применением лингвистического анализа. По совокупности найденной эмотивной лексики текст может быть оценен по шкале, отражающей количество негативной и позитивной лексики. Этот метод может использовать как списки паттернов, подставляемые в регулярные выражения, так и правила соединения тональной лексики внутри предложения. Данный метод трудоемок, поскольку требует составления тональных словарей (или получения списка тональных паттернов), но в сочетании с синтаксическим и морфологическим анализом позволяет не только показать цепочки тональной лексики, но и получить синтаксически корректные эмоциональные выражения. При хорошем наполнении тональных словарных списков этот метод позволяет достичь хорошей полноты (покрытия эмотивной лексики). Недостаток этого метода в том, что с помощью него сложно дать количественную оценку негативности-позитивности текста.

Чтобы избежать недостатков первого и второго метода, используют смешанный подход, частично включающий в себя два первых [31].

В качестве примера технологии распознавания эмоциональной окрашенности (сентимента) текста на русском языке можно рассмотреть продукт SentiScan [56]. Он реализован на основе сочетания методов, основанных на машинном обучении и статистике, и методов, основанных на правилах. Задача системы заключается в том, чтобы найти в тексте определенные объекты, выделить сентиментный контекст и распознать сам сентимент. Деятельность системы основана на словарях, состоящих из «полярных единиц» (прилагательных, существительных, глаголов, наречий), имеющих однозначное тональное значение — позитивное либо негативное. При составлении словарей применялись методы машинного обучения. В процессе синтаксического анализа производится наращивание информации и статистики о сентиментном потоке (его силе и полярном окрасе — позитив либо негатив) и его направленности на целевой объект. Накопив информацию об отдельных предложениях,

алгоритм переходит на уровень текста, на котором вычисляется финальная информация. В итоге алгоритм выносит вердикт по всему тексту: позитив либо негатив. Текст может быть также помечен и нейтральным флагом, если в тексте не было ни одной тонально окрашенной единицы либо синтаксического противопоставления или если в тексте был смешанный сентимент и неясно, что хотел сказать своим высказыванием автор [56].

Информация о русскоязычных системах аналогичной направленности содержится также в публикациях по итогам семинара РОМИП-2011 [17] и обзоре [26].

Важной тенденцией в развитии моделей определения эмоциональной тональности текста является отход от интегральной биполярной системы ее оценивания. В альтернативных подходах реализуются следующие идеи:

- Оценка выраженности сентимента, приписываемая значимому слову (при этом «вес» слова может принимать значения от 0 до ± 1) [8, 60, 62], или градуированная оценка общего сентимента текста (пример – пятибалльная шкала: 1 – отзыв об объекте отрицательный; 2 – в целом отрицательный, но указывает на отдельные положительные стороны объекта; 3 – указывает на значимые положительные и отрицательные стороны объекта; 4 – в целом положительный, но отмечены отдельные недостатки; 5 – положительный, недостатков у объекта практически нет [49]).

- Учет многомерности пространства эмоциональных состояний человека, создание вместо пары множества словарей, соответствующих базовым для культуры эмоциям (с примерами для разных языков можно ознакомиться, например, в работах [46]).

- Учет темы текста/высказывания. От обсуждаемой тематики зависит, прежде всего, состав словарей, используемых для определения эмоционального отношения [50], [64], [65]. Так, предлагаемый М. Thelwall и К. Buckley метод *Lexicon extension* подразумевает расширение общего эмоционального лексикона за счет дополнительных слов, отражающих конкретную тематику [59].

Задачи тематического анализа порождают интерес к новым средствам, которые позволили бы восстанавливать тему высказываний по самому интернет-контенту, а не в соответствии с

предварительно сформулированным запросом [44]. Это особенно актуально при анализе контента свободного сетевого общения, когда один текст может содержать мнения по нескольким темам [61].

Кроме того, ставятся задачи анализа структуры сентиментного сетевого контента: выборочное выделение в массиве текстов различных точек зрения на объект; определение таких качеств объекта, которые вызывают согласованные или рассогласованные оценки разных авторов; выявление сообществ, интересующихся объектом; оценка влиятельности мнений, высказываемых разными авторами [55]. Отмечаются также попытки научиться определять причины динамики выраженного в текстах отношения к определенному объекту [51, 53].

Для текстов на русском языке несколько лет назад была предложена модель системы аналитического аннотирования ценностного и оценочного содержания публикаций СМИ [14]. Ее основой послужили принципы и методы извлечения тезаурусных знаний из текста, разработанные Г.С. Осиповым и Б.Ю. Городецким. Лингвистическая концепция модели сочетает в себе идеи контент-анализа дискурса, тезаурусной систематизации понятий, компонентного анализа лексических значений, аксиологической семантики, прикладного словообразовательного анализа и лексикостатистики. Созданная компьютерная система включает в себя четыре модуля: (1) ведение опорных словарей; (2) прикладной морфологический анализ словоформ текста; (3) терминологический анализ текста; (4) собственно извлечение тезаурусных знаний из текста. Для обеспечения работы системы были созданы несколько словарей, основным из которых является Базовый Семантический Словарь (БАСС) – развиваемый словарь важнейших лексических единиц, существенных для интересующей предметной области. Каждая словарная статья описывает гиперлексему, то есть класс лексем, которые имеют общую квазиоснову и отождествляются в рамках данной понятийной системы. В отдельных зонах словарной статьи описаны аспекты тезаурусной характеристики гиперлексем. При анализе текстов системный БАСС использует, во-первых, 12 имплицитных обобщенных ценностных категорий, во-вторых, открытое множество эксплицитных ценностных категорий,

в-третьих, пометы об отрицательной или положительной оценке. Результатом анализа является Тезаурусный образ текста, формализующий ценностный и оценочный аспекты содержания [14].

Таким образом, современные средства анализа сетевого дискурса делают возможным исследование такой интегральной характеристики текста, как его эмоциональная тональность, и при этом не ограничиваются дихотомической квалификацией (положительный – отрицательный сентимент), но предлагают различные варианты дифференциации категорий оценки.

В основу нашего подхода положена модель отражения в формализуемых показателях текста эмоционального отношения его автора к различным аспектам окружающей социальной реальности.

Настоящее исследование направлено на выявление признаков эмоций определенного характера – а именно, связанных с агрессией, возникающей в отношении к различным аспектам социальной реальности. Поэтому в дальнейшем изложении будут описаны только некоторые особенности созданной системы анализа текстов, непосредственно имеющие отношение к диагностике уровня вербальной агрессивности сетевого дискурса.

Лингвистические аспекты выражения эмоциональной оценки исследовались в работах таких авторов, как Н.Д. Арутюнова, В.Ю. Апресян, Т.В. Булыгина, А. Вежбицка, Е.М. Вольф, В.Г. Гак, Г.Ф. Гибаева, Е.С. Кубрякова, В.А. Маслова, А.Ф. Папина, А.Б. Пеньковский, Ю.С. Степанов, В.Н. Телия, А.А. Уфимцева, А.Д. Шмелев и др. Благодаря многочисленным лингвистическим исследованиям были выявлены и описаны закономерности отражения эмоционального состояния автора в формализуемых показателях порождаемого им текста, например [1, 11, 13, 16, 18, 22, 35, 39].

3. Система психолингвистического анализа текстовой информации PLATIn

При разработке системы автоматической диагностики степени эмоциональной насыщенности текста представляется целесообразным задействовать такие показатели, значение которых в наименьшей степени зависит от контекста употребления. В этом отношении наиболее пер-

спективным представляется анализ средств выражения вербальной агрессии на лексическом уровне. Как уже упоминалось, лексические средства вербальной агрессии – это слова, выражающие негативную оценку [32].

На основе имеющихся в литературе представлений нами была разработана пилотажная версия списков лексических единиц («словарей»), предназначенных для использования в работе системы психолингвистического анализа текстовой информации PLATIn. Словари были составлены методом сплошной выборки из материала Русского орфографического словаря Российской академии наук [34]; кроме того, использованы материалы других словарей: [9, 10, 24, 27, 33, 40, 43]. Привлекался также лексический материал, опубликованный в Жаргон.ру, Slang [58] и Slang Dictionary [57].

Созданные списки включают в себя лексические единицы (слова, в некоторых случаях – словосочетания), семантика которых имеет отношение к следующим сферам: экспрессивности как таковой (без определения содержания отношения, которое испытывает автор), отрицательной эмоциональной оценки, отрицательной рациональной оценки, положительной эмоциональной оценки, положительной рациональной оценки и ментальных действий. Лексика, отнесенная к каждой из перечисленных сфер, содержательно распадается на более частные группы.

1. Лексика недифференцируемой по смыслу экспрессии.

- 1.1. Лексика мотивации, деятельности и напряжения (ок. 800 ед.), например: *решиительный, добиваться, кураж*.
- 1.2. Жаргонная лексика (ок. 8500 ед.), в том числе: «Молодежный» жаргон, «Компьютерный» жаргон, «Криминальный жаргон»: *оборзеть, бабло, передоз, зиповать*.
- 1.3. Обсценная лексика (ок. 800 ед.).
- 1.4. Безысклительная и усилительная лексика (ок. 2000 ед.): *все, никто, до фига, невозможно*.

2. Лексика отрицательной эмоциональной оценки

- 2.1. Инвективы (ок. 7000 ед.): общая и специальная отрицательная оценка субъектов и групп; аномальное речевое поведение; выражение социального порицания; телесные отправления и

секс; метафоры негативного характера: *безобразие, криворукий, харя, врать, чавкать*.

- 2.2. «Мягкие инвективы» (ок. 3000 ед.): порицаемые неразрушительные действия; отсутствие у предмета ценности / крайне низкая ценность предмета; процесс и результат утраты предметом признака или интенсивности признака; проявление неловкости в действиях субъекта: *развалюха, тыкаться, протухлый, истрепаться*.

- 2.3. Лексика разрушения и насилия (ок. 4000 ед.): разрушительные действия и их результаты; лексика катаклизмов и катастроф; «военная» лексика: *взрыв, раздирать, потонуть*.

- 2.4. Лексика страдания (ок. 3000 ед.): лексика смерти и болезни; лексика отрицательных телесных состояний; лексика отрицательных эмоциональных состояний астенического круга, в том числе, невербальных их проявлений: *вымирать, больной, тоска*.

- 2.5. Лексика стенических негативных эмоций (200 ед.): *злой, брезгливость, ревновать*.

- 2.6. Лексика социальной разобщенности (ок. 8000 ед.): лексика социальной изолированности и аномального социального положения; лексика преступлений и преступного мира (не жаргонная); лексика власти; лексика суда; лексика сокрытия, обмана и манипулирования; лексика зрелищ и публичности; лексика игры: *владыка, принуждать, беззащитный, репетировать, провокатор*.

- 2.7. Лексика протестного поведения (ок. 550 ед.): *голодовка, крамольный, пикетировать*.

3. Лексика отрицательной рациональной оценки (ок. 4000 ед.): неэффективные ментальные действия и их результаты; отсутствие у предмета необходимых качеств, утраты позитивных характеристик, возникновения негативных характеристик: *непонимание, отсталый, испортить*.

4. Лексика положительной эмоциональной оценки (ок. 5500 ед.): лексика позитивных эмоциональных состояний астенического и стенического характера, в том числе, их невербальных проявлений; лексика здоровья и вы-

здоровления; лексика зарождения, развития, роста; лексика улучшения, совершенствования; лексика доверия, искренности, симпатии и социального сплочения; лексика семейных ролей и отношений: *шутка, защищать, братский, надежда, обожать*.

5. Лексика положительной рациональной оценки и ментальных действий (ок. 9000 ед.): лексика ментальных действий (в т.ч., глаголы-делиберативы); позитивная оценка результата ментальных действий; описание наличия позитивных качеств, исправления негативных качеств, возникновения позитивных качеств у описываемых предметов: *стабильный, продумать, авторитет*.

Определение содержания в тексте или массиве текстов лексики, относящейся как к общим категориям, так и к отдельным подкатегориям, позволяет составлять своеобразный профиль субъективного отношения к обсуждаемым темам и, таким образом, описывать это отношение в соответствующих выделенным категориям терминах³.

Компьютерная реализация нашего подхода заключалась в создании системы психолингвистического анализа текстовой информации (PLATIn). Одной из характеристик разработанной системы является возможность автоматического анализа веб-страниц.

Система состоит из двух компонент – сервера и клиента. Сервер непосредственно осуществляет обработку страниц, клиент отображает исходные страницы и результаты анализа. Связь клиента и сервера осуществляется при помощи протокола HTTP.

Алгоритм анализа страниц состоит из нескольких шагов. Общая схема алгоритма представлена на Рис. 1 в виде диаграммы потоков.

На первом шаге алгоритма выполняется разбор HTML-кода обрабатываемой страницы сайта – выделяются текстовые данные. При этом сохраняется информация о структуре страницы для возможности последующего отображения размеченного текста.

На втором шаге с помощью процессора *Exactus Expert* [54] производится автоматический лингвистический анализ текста, определяющий синтаксические структуры предложений и морфологические свойства слов. Результаты лин-

³ В настоящей работе не будут обсуждаться данные для словарей, содержащих слова с положительной семантикой.

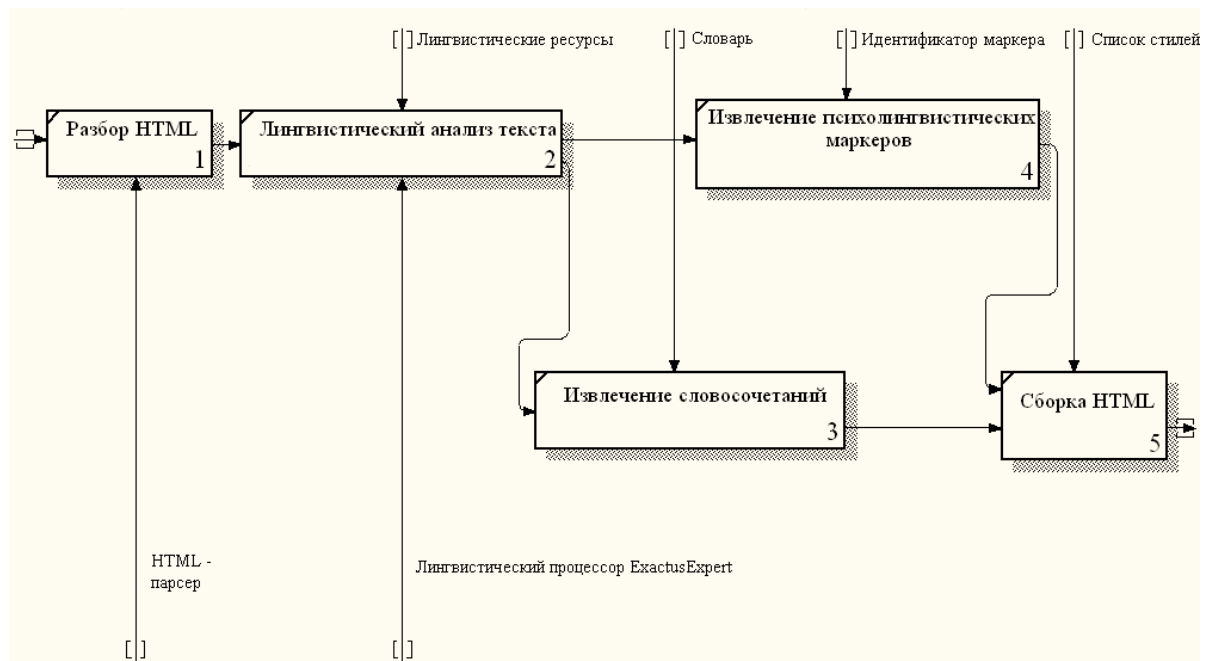


Рис. 1. Диаграмма потоков системы психолингвистического анализа текстовой информации (PLATIn)

гвистического анализа сохраняются в виде динамических структур данных (связных списков).

На третьем шаге выполняется выявление психолингвистических характеристик текста⁴ и определение процентного содержания слов и словосочетаний из заданных словарей. Для этого в результатах лингвистического анализа осуществляется поиск слов с заданными морфологическими признаками, на основании которых вычисляются значения коэффициентов в соответствии с их определением.

Извлечение словосочетаний по словарям проходит в два этапа. На первом этапе словарь слов и словосочетаний загружается в память в виде префиксного дерева trie [52]. Каждая вершина $k = 1..l$ такого дерева представляет собой двойку $\langle s_k, Q_k \rangle$, где Q_k – множество дочерних вершин для k , s_k – нормальная форма слова. Такое дерево заполняется по следующему алгоритму:

1. Каждое словосочетание из загружаемого словаря разбивается на отдельные слова.

2. Для каждого слова определяется нормальная форма.

3. Нормальная форма, соответствующая первому слову в словосочетании, добавляется в потомки Q_I корневой вершины дерева $\langle s_I, Q_I \rangle$, где s_I – пустая строка.

4. Нормальные формы следующих слов добавляются в соответствующие вложенные вершины дерева.

Для повышения эффективности программной реализации в системе психолингвистического анализа текстовой информации списки дочерних вершин реализуются в виде хеш-таблиц [63] (Рис. 2).

На втором этапе извлечения словосочетаний производится обход результатов лингвистического анализа текста и поиск нормальных форм слов в построенном дереве *trie*, что позволяет находить слова не только по точному совпадению, но и по их различным словоформам. Кроме того, поиск словосочетаний носит нечеткий характер, поскольку входящие в них слова могут следовать в тексте как в прямом, так и в обратном порядке, а также могут быть разбиты дополнительными словами.

Словари загружаются из реляционной базы данных, развернутой на сервере MySQL5. Был реализован веб-интерфейс редактирования словарей, с помощью которого можно создавать новые словари, а также изменять их содержимое.

4 Особенности текста, с помощью которых возможна диагностика состояния эмоционального напряжения автора на момент создания текста. Их описание можно видеть в работе [4]. Возможность определения психолингвистических показателей реализована в нашей системе, однако рассмотрение соответствующих характеристик сетевого дискурса выходит за рамки настоящего исследования.

На последнем шаге, на основании исходного кода страницы сайта и данных анализа формируется результат работы системы: подсвечиваются слова, определяющие психолингвистические характеристики текста, а также словарные слова, эти слова выводятся интерактивными списками, распределенными по активным словарям, и указывается число слов, относящихся к каждому из словарей. Имеется возможность сохранить результаты анализа в файл формата HTML, а также загрузить подобный файл с диска.

4. Эмпирическое исследование агрессивности блогов и сетевых СМИ с помощью системы PLATIn

Наше эмпирическое исследование было посвящено в первую очередь проверке возможностей созданного инструмента. Проверка проводилась в рамках работ по изучению связей между агрессивностью и ценностями. Были сформулированы две исследовательские задачи:

1. определение уровня вербальной агрессивности сетевого дискурса в группах, разделяющих различные ценности;
2. установление различий между уровнем агрессивности, характерным для текстового наполнения главных страниц интернет-версий ведущих российских газет и журналов различной направленности.

4.1. Вербальная агрессия сетевого дискурса в группах, разделяющих различные ценности

Группирование в сети, как уже отмечалось, также как и любое группирование, основано на принятии членами группы групповых целей и стоящих за целями общих ценностей. В качестве классификатора ценностей был выбран классификатор, предложенный Ш. Шварцем в 1992 г. в рамках его теории базовых ценностей и модифицированный им в работах последних лет, в частности, апробированный на российской выборке в 2011 г. [41]. Из списка 19 ценностных мотивационных типов было выбрано 6:

- стимуляция (стремление к возбуждению, новизне и переменам);
- гедонизм (стремление к удовольствию и чувственному удовлетворению);
- скромность (признание незначительности одного человека в круговороте жизни);

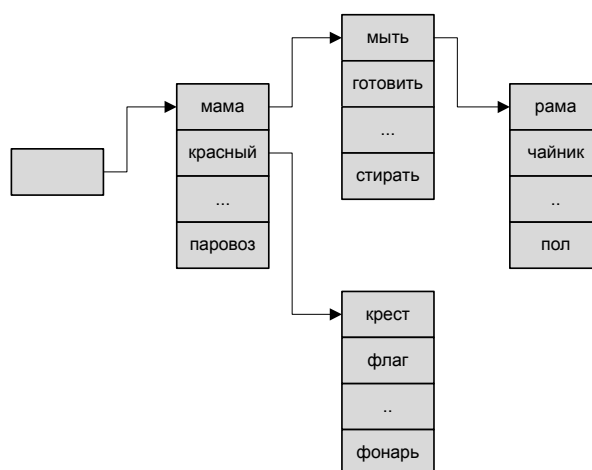


Рис. 2. Схема дерева хранения словосочетаний

- универсализм-забота о других (стремление к равенству, справедливости и защите всех людей);
- универсализм-забота о природе (сохранение природной среды);
- благожелательность-забота (преданность группе и благополучие ее членов).

На основе экспертных оценок в сетях Фейсбук и ЖЖ были отобраны блоги, тексты которых отражают мотивированность их авторов одной из этих ценностей. Для ценности Стимуляция был выбран блог молодого яхтсмена, посвященный обсуждению морских походов и жизни на борту судна; для ценности Гедонизм – блог зрелого бизнесмена, владельца яхты, посвященный поездкам в Европу; для ценности Скромность – блог идеолога русского национализма Константина Крылова; для ценности Универсализм-забота о других – блог независимого муниципального депутата (г. Москва, Ломоносовский р-н); для ценности Универсализм-забота о природе – блог Гринпис в России; для ценности Благожелательность-забота – форум волонтеров приюта для животных (Москва, Кожухово) и блог руководителя добровольного сообщества помогающих больным детям («Конвертик для Бога»).

Для каждого сообщества, разделяющего определенную ценность, в текстах блогов было отобрано по два поста, к которым набралось не менее 50 комментариев (от 52 до 215). Тексты (пост + комментарии) были обработаны с помощью системы PLATIn. Результаты представлены на графиках (Рис. 3, Рис. 4).



Рис. 3. Абсолютная частотность агрессивной лексики на 100 комментариев



Рис. 4. Абсолютная частотность агрессивной лексики по группам на 1000 слов

Рассмотрим полученные результаты.

В группах, ориентированных на разные ценности, приняты различающиеся тактики ведения беседы: обсуждения защитников природы отличаются многословностью реплик (в среднем 122 слова на комментарий), а защитников прав человека – большим числом комментариев (215 комментариев); наиболее лаконичны комментарии в сообществе, поддерживающем ценность гедонизма (в среднем 20 слов на комментарий).

Защитники природы порождают тексты, резко отличающиеся представленностью лексики агрессии – практически по всем словарям (кроме словаря стенических негативных реакций) показатели этой группы выше, чем показатели остальных групп. Можно предположить, что эта особенность обусловлена объективными причинами – защитники обсуждают факты разрушения, истребления, порчи, готовят мероприятия противодействия, выражения несогласия, борьбы, мечтают об упразднении, судеб-



Рис. 5. Безысключительная лексика на 1000 слов: относительная и абсолютная частотность

ном преследовании, разорении провинившихся компаний. Однако для всех остальных эта эмоциональная напряженность текстов может служить психологическим барьером на пути принятия ценностей данной группы и присоединения к сообществу Гринпис.

Защитники прав человека чаще всего используют лексику социальной разобщенности. Более того, именно в текстах этой группы этот показатель достигает максимума – защитники природы и националисты тоже часто оперируют этой лексикой, но все же не так часто, как члены сообщества муниципального депутата.

Представляется, что данные по двум группам защитников демонстрируют адекватность применяемых словарей и всей процедуры в целом – защита как ценность, входящая в состав ценности универсализма, строится на выделении в мире образа врага – врага, разрушающего природу, и врага, нарушающего права людей и поступающего несправедливо. Если бы лексика разрушения и лексика разобщенности встречалась в текстах «универсалистов» не чаще, чем в текстах гедонистов, пришлось бы признать, что наш инструмент неэффективен.

Отметим, также, что те, кто проявляет заботу о выбранной группе подопечных (брошенных собаках или больных детях) часто пользуются инвективами, даже чаще, чем националисты. Природа этого – в использовании слов «собака» и «зверь» в прямом, а не в переносном смысле слова (в использованном посте благотворителей детских больниц обсуждался вопрос о добродушии собак – даже тех, что выглядят страшными зверюгами). Этот пример показывает, что оценка агрессивности должна строиться, во-первых, на основе данных многих словарей, а не только по частоте упот-

ребления инвектив, а, во-вторых, с учетом обсуждаемой тематики.

При обработке результатов мы столкнулись с проблемой различного объема словарей. Так, в словаре жаргонной лексики количество ЛЕ (лексических единиц) – 8500, а в словаре стенических негативных состояний – всего 200. Для того чтобы можно было учесть это различия, были вычислены «повышающие коэффициенты» для словарей (за единицу было принято количество ЛЕ в словаре жаргонной лексики). Смысл этой процедуры можно было бы сформулировать так: если бы во всех словарях было одинаковое количество ЛЕ, то, были бы получены результаты, отраженные на Рис. 5. Как можно видеть, при такой обработке различия между группами текстов становятся более выраженными.

Представляется, что, для того, чтобы оценить «субъективный образ» текста, то есть, то, каким представляется обсуждение в некотором сообществе стороннему наблюдателю, еще не принявшему для себя решения о присоединении к этому сообществу, имеет смысл использовать абсолютные показатели (Рис. 3 и Рис. 4). А для описания объективных тенденций в склонности к использованию то или иной лексики авторами можно использовать относительные показатели (Рис. 5).

4.2. Агрессивность текстового наполнения главных страниц интернет-СМИ

Предметом анализа второго исследования являлась агрессивность, характеризующая текстовое наполнение главных страниц сетевых СМИ и интернет-версий ведущих российских газет и журналов. Объекты исследования отби-

рались по ряду критериев: направленность СМИ (политическая / неполитическая), тематика (женские журналы, научно-полярные СМИ, новости, журналы для чтения, пропагандистские СМИ и т.п.), идеология (либеральная, социалистическая, националистическая). С помощью системы PLATIn анализировалось текстовое наполнение главной страницы сайта, включая заголовки, анонсы, цитаты, названия самих рубрик сайтов. Были исследованы 36 сайтов, различия между группами СМИ оценивались по критерию Манна-Уитни.

Полученные данные свидетельствуют, прежде всего, об оправданности выделения так называемых СМИ политической направленности – текстовое наполнение политических СМИ значительно отличается от текстов неполитических СМИ по таким показателям, как насыщенность лексикой социальной разобщенности и лексикой протестного поведения, а также значительно большим присутствием лексики разрушения и насилия. Интересно, что при этом различия внутри группы политизированных СМИ на уровне лексики агрессивности оказались не столь существенны. Видимо, различия между либералами, националистами и «левыми» проявляются на более глубоком, семантико-синтаксическом, уровне построения текста и должны исследоваться методами реляционно-ситуационного анализа.

Заключение

Для современной психологии, социологии и культурологии характерно обращение к реальности сетевой коммуникации. При этом одной из актуальных задач является задача изучения и мониторинга уровня агрессии и социальной напряженности в обществе, осуществляемых методами автоматического анализа сетевого контента, в котором агрессивность и напряженность находят своё выражение. Нас проблема выраженности вербальной агрессивности интересует в первую очередь как связанная с формированием сетевых сообществ и рекрутированием в них новых участников. Мы предполагаем, что ценностные приоритеты выступают в качестве основания группообразования и присоединения к сообществу, а выраженность враждебной позиции у членов сетевых дискуссий позволяет человеку на чувственном уровне, без глубокого анализа содержания тек-

стов, ориентироваться в ценностях, объединяющих членов сообществ и осуществлять таким образом выбор для себя группы принадлежности. В связи с этим особый интерес для нас представляют методы анализа эмоциональной тональности текстов. Были проанализированы возможности применения средств интеллектуального анализа и методов оценки субъективного содержания текста (Opinion Mining). Исследование вопроса о возможностях создания методов автоматического выявления агрессивности в текстах социальных сетей проявлений показало, что эффективно решать эту задачу можно с помощью процессора Exactus Expert, позволяющего производить автоматический лингвистический анализ текста и определять синтаксические структуры предложений и морфологические свойства слов.

Проведенное эмпирическое исследование показало возможность применения развиваемых методов автоматизированного анализа текстов в задаче поиска закономерностей средств использования вербальной агрессии людьми с разными типами ценностных приоритетов.

Литература

1. Апресян В.Ю. Имплицитная агрессия в языке // <http://www.dialog-21.ru/Archive/2003/Apresian.htm>
2. Белинская Е. К проблеме групповой динамики сетевого сообщества // 2-ая Российская конференция по экологической психологии. Тезисы. Москва, МГУ, 12-14 апреля 2000 г. М.: Экспосцентр РОСС. С. 249-251.
3. Белоконева А.С. Конструирование образа внешнего врага: исследование советских СМИ и официальных документов начала «холодной войны» (1946-1953 гг.). А/р. дисс. ... канд. социол. наук. М., 2004.
4. Белянин В.П. Психоллингвистика. М.: Флинта; МПСИ, 2009.
5. Бершадская Л., Биккулов А., Болгова Е., Чугунов А., Якушев А. Социальные сети и социометрические исследования: теоретические основания и практика использования автоматизированного инструментария изучения виртуальных сообществ // Информационные Ресурсы России. 2012. № 4 // http://www.aselibrary.ru/digital_resources/journal/irr3648/irr36483711/irr364837113734/irr3648371137343743/
6. Бикейкина Н.А. Имя собственное как фактор возникновения лингво-юридических конфликтов // Юрислингвистика-10: Лингвоконфликтология и юриспруденция: межвузовский сборник научных трудов / Под ред. Н.Д.Голева и Т.В.Чернышовой. Кемерово; Барнаул: Изд-во Алт. ун-та, 2010. С. 106-113.
7. Войскунский А.Е. Общение, опосредованное компьютером. Дисс. ...канд. психол. наук. М., 1990.

8. Воронова О.О., Соколов М.М., Петров С.А. Построение системы сентиментного анализа сообщений социальных сетей // <http://ispo.elit.sumdu.edu.ua/files/theses/Voronova%20O.Socolov%20M.M.Petrov%20S.A.pdf>
9. Гайдамак Н., Осипов Б., Имедадзе Н. Словарь современного русского города. М.: Русские словари, АСТ, 2003.
10. Голев Н.Д., Головачева О.В. Юрислингвистический словарь инвективной лексики русского языка // Юрислингвистика-6: Инвективное и манипулятивное функционирование языка: межвузовский сборник научных трудов / Под ред. Н. Д. Голева. Барнаул: Изд-во Алтайского университета, 2005. С. 121-128.
11. Голованова А.В. Анализ оценочной модальности при проведении лингвистической экспертизы // Юрислингвистика-7: Язык как феномен правовой коммуникации / Под ред. Н.Д.Голева. Барнаул: Изд-во Алт. ун-та, 2005.
12. Давыдов А.А. Системная социология: Opinion Mining // http://www.isras.ru/index.php?page_id=1024&printmode
13. Зайцева Е.Л. Выражение отрицательной оценки в политическом дискурсе: опыт сравнительно-сопоставительного исследования российских и французских печатных средств массовой информации. Дисс... канд.филол.наук. Чебоксары, 2006.
14. Зевахина Т.С., Олейникова Е.Е. Аналитическое аннотирование текстов СМИ: Ценности и оценки // <http://www.dialog-21.ru/digests/dialog2006/materials/pdf/ZevakhinaT.pdf>
15. Иванченко Д.А. Интерпретация интернет-пространства в дискурсе социологии // Ученые записки Российского государственного социального университета. 2009. №3 (66). С.72-78.
16. Колосов С.А. Конструирование социальной ненависти в дискурсе. Дисс... канд.филол.наук. Тверь, 2004.
17. Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 30 мая - 3 июня 2012 г.). Вып. 11 (18). М.: Изд-во РГГУ, 2012.
18. Котов А.А. Распознавание эмоционального состояния адресата в диалоге // Понимание в коммуникации. 2007. Язык. Человек. Концепция. Текст: Тезисы докладов Международной научной конференции (28 февраля – 1 марта 2007 г.). М.: НИВЦ МГУ, 2007 // http://www.srcc.msu.ru/unipersona/site/conf/conf_07/thesis-07.htm
19. Кравцов О.Г. Толерантность как единица исследования онтогенеза личности: на материале законопослушных и девиантных подростков и юношей. А/р. дисс. ...канд.психол.наук. М., 2008.
20. Краснянская Т.И. К обоснованию понятия *напряженность слова* в судебной коммуникации // Юрислингвистика-10: Лингвоконфликтология и юриспруденция: межвузовский сборник научных трудов / Под ред. Н.Д.Голева и Т.В.Чернышовой. Кемерово; Барнаул: Изд-во Алт. ун-та, 2010. С. 40-46.
21. Кузнецова Ю.М., Чудова Н.В. Психология жителей Интернета. М.: Издательство ЛКИ, 2011.
22. Лазариди М.И. Психические состояния в полевом описании: номинативно-функциональный аспект М.: Издательство Московского психолого-социального института; Воронеж: Издательство НПО «МОДЭК», 2010.
23. Лазебная И. Б. Особенности выражения речевой агрессии в современном английском языке. А/р. дисс. ...канд.филол.наук. Белгород, 2007.
24. Левикова С.И. Большой словарь молодежного сленга. М., 2003.
25. Леонтьев Д.А. Психология смысла: природа, строение и динамика смысловой реальности. М.: Смысл, 2007.
26. Меньшиков И.Л., Кудрявцев А.Г. Обзор систем анализа тональности текста на русском языке // Молодой ученый. 2012. № 12. С. 140-143.
27. Мокиенко В.М., Никитина Т.Г. Словарь русской брани (матизмы, обсценизмы, эвфемизмы). СПб.: "Норинт", 2003.
28. Мясникова М.Л. Конструирование социальных групп в киберпространстве: возникновение групповой мифологии // http://samlib.ru/m/mjasnikowa_marija_leonidowna/diplo m.shtml
29. Мясникова М.Л., Мясникова В.Л. Трансформация социальной среды: понятие сетевого сообщества // socio.myl.ru/_fr/1/7805249.doc
30. Никитин А.В. Психологические особенности агрессивного поведения в подростковом возрасте. А/р. дисс. ...канд.психол.наук. М., 2006.
31. Пазельская А.Г., Соловьев А.Н. Метод определения эмоций в текстах на русском языке // Компьютерная лингвистика и интеллектуальные технологии. Сб. научных статей. Вып. 10 (17). М.: Изд-во РГГУ, 2011. С. 510-522.
32. Петрова Н.Е., Рацибурская Л.В. Язык современных СМИ: средства речевой агрессии. М.: Флинта: Наука, 2011.
33. Плущер-Сарно А.Ю. Большой словарь мата. В 2 т. СПб.: Лимбус-Пресс, 2005.
34. Русский орфографический словарь / РАН. Ин-т рус.яз.им. В.В.Виноградова / Отв. Ред. В.В.Лопатин. Изд 3-е. М.: АСТ-ПРЕСС КНИГА. 2010.
35. Самарина А.В. Коммуникативные стратегии "создание круга чужих" и "создание круга своих" в политической коммуникации: прагматический аспект. Дисс... канд.филол.наук. Ростов-на-Дону, 2006.
36. Седов К.Ф. Агрессия и манипуляция в повседневной коммуникации // Юрислингвистика-6: Инвективное и манипулятивное функционирование языка: межвузовский сборник научных трудов / Под ред. Н.Д.Голева. Барнаул: Изд-во Алтайского университета, 2005. С. 87-103.
37. Тамразова И.Г. Функционально-прагматические характеристики эристического дискурса. Пятигорск, 2009.
38. Усланер Э. Социальный капитал и Сеть // Интернет в общественной жизни. М.: Идея-Пресс, 2006. С. 123-127.
39. Фаткуллина Ф.Г. О деструктивной лексике русского языка // Языковая семантика и образ мира: Тезисы международной научной конференции. Казань, 1997. Кн. 1. С. 64-66 // http://www.kcn.ru/tat_ru/science/news/lingv_97/n32.htm
40. Чудинов А. П. Россия в метафорическом зеркале: когнитивное исследование политической метафоры (1991-2000). Екатеринбург, 2001 // <http://helpforlinguist.narod.ru/linguistics2/chudinov-01.html>
41. Шварц Ш., Бутенко Т.П., Седова Д.С., Липатова А.С. Уточненная теория базовых индивидуальных ценностей:

- применение в России // Психология. Журнал Высшей школы экономики. 2012. Том 9. Вып. 2. С. 43–70.
42. Щербинина Ю.В. Речевая агрессия. Территория вражды. М.: ФОРУМ, 2013.
 43. Юнаковская А.А. Показатели конфликтного общения русских // Юрислингвистика-6: Инвективное и манипулятивное функционирование языка: межвузовский сборник научных трудов / Под ред. Н.Д.Голева. Барнаул: Изд-во Алтайского университета, 2005. С. 129–149.
 44. Atefeh, F., & Khreich, W. A Survey of Techniques for Event Detection in Twitter. Computational Intelligence. 2013 // <http://onlinelibrary.wiley.com/doi/10.1111/coin.12017/abstract>
 45. Francisco, V., & Gervás, P. EMOTAG: An Approach to Automated Markup of Emotions in Texts // Computational Intelligence, 2013, 29: 680–721.
 46. Gaber, M., Cocca, M., Weibelzahl, S., Menasalvas, E., & Labbé, C. (Eds.) First International Workshop on Sentiment Discovery from Affective Data (SDAD), Bristol, UK, September 28, 2012. Proceedings.
 47. Gil de Zúñiga, H., Jung, N., & Valenzuela, S. Social Media Use for News and Individuals' Social Capital, Civic Engagement and Political Participation // Journal of Computer-Mediated Communication, 2012, 17: 319–336.
 48. Houen, S. Opinion Mining with Semantic Analysis. 2011 // http://www.diku.dk/forskning/Publikationer/specialer/2011/specialerapport_final_Soren_Houen.pdf
 49. <http://romip.ru/ru/2011/track/sentiment-classification-5.html>
 50. Husaini, M., Koçyğit, A., Tapucu, D., Yanikoglu, B., & Saygin, Y. An Aspect-Lexicon Creation and Evaluation Tool for Sentiment Analysis Researchers. 2012 // <http://www.cs.bris.ac.uk/~flach/ECMLPKDD2012papers/1125807.pdf>
 51. Jiang, Y., Meng, W., & Yu, C. Topic Sentiment Change Analysis // P. Perner (Ed.): Machine Learning and Data Mining in Pattern Recognition - 7th International Conference, MLDM 2011, New York, NY, USA, August 30 - September 3, 2011. Proceedings, 443–457.
 52. Knuth D. E. The art of computer programming. Vol. 3. Sorting and Searching. Reading, Mass.: Addison-Wesley, 1973.
 53. Lee, S.Y.M., Chen, Y., Huang, C.-R. & Li, S. Detecting Emotion Causes with a Linguistic Rule-Based Approach // Computational Intelligence, 2013, 29: 390–416.
 54. Osipov, G., Smirnov, I., Tikhomirov, I., Shelmanov, A. Relational-situational method for intelligent search and analysis of scientific publications // Proceedings of the Workshop on Integrating IR technologies for Professional Search, vol. 968. CEUR Workshop Proceedings, Moscow, 2013.
 55. Pang B. & Lee L. Opinion Mining and Sentiment Analysis // Foundations and Trends in Information Retrieval, 2008, 2, 1-2: 1-135.
 56. SemanticAnalyzer Group Blog // <http://semanticanalyzer.info/blog/category/анализ-тональности/>
 57. Slang Dictionary <http://slangdictionary.ru/>
 58. Slanger // http://slanger.ru/?mode=library&r_id=6&page=24
 59. Thelwall, M., & Buckley, K. Topic-based Sentiment Analysis for the Social Web: The Role of Mood and Issue-Related Words // J. Am. Soc. Inf. Sci., 2013, 64: 1608–1617.
 60. Thelwall, M., Buckley, K. & Paltoglou, G. Sentiment Strength Detection for the Social Web // J. Am. Soc. Inf. Sci., 2012, 63: 163–173.
 61. Tsytarau, M., & Palpanas, T. Mining Subjective Data on the WEB // Technical Report # DISI-10-045. 2010 // <http://www.disi.unitn.it>
 62. Wilson, T., Wiebe, J., & Hwa, R. Recognizing Strong and Weak Opinion Clauses // Computational Intelligence, 2006, 22: 73–99.
 63. Wirth, N. Algorithms + Data Structures = Programs. Prentice-Hall, 1976.
 64. Yi, J., Nasukawa, T., Bunesco, R., & Niblack, W. Sentiment Analyzer: Extracting Sentiments about a Given Topic using Natural Language Processing Techniques // Proceedings of the 3rd IEEE international conference on data mining (ICDM 2003), Florida, USA, November 19–22, 2003: 427–434.
 65. Yu, N. Domain Adaptation for Opinion Classification: A Self-Training Approach // J. of infosci. theory and practice, 2013, 1(1): 10–26 // <http://dx.doi.org/10.1633/JISTaP.2013.1.1.1>

Девяткин Дмитрий Алексеевич. Инженер Института системного анализа РАН. Окончил Рыбинскую государственную авиационную технологическую академию в 2011 году. Автор 5 печатных работ. Область научных интересов: адаптивные методы дистанционного обучения, классификация и кластеризация текстов, интеллектуальный веб-краулинг, искусственный интеллект. E-mail: devyatkin@isa.ru.

Кузнецова Юлия Михайловна. Старший научный сотрудник Института системного анализа РАН. Окончила МГУ им. М.В.Ломоносова в 1991 году. Кандидат психологических наук (2006). Автор 70 печатных работ, в том числе 1 монографии. Область научных интересов: картина мира, психосемантика, психоллингвистика, психология агрессии.

Швец Александр Валерьевич. Инженер-исследователь Института системного анализа РАН. Окончил Сибирский федеральный университет в 2011 году. Автор 14 печатных работ. Область научных интересов: компьютерная лингвистика, математическое моделирование, методы оптимизации, искусственный интеллект. E-mail: shvets@isa.ru

Чудова Наталья Владимировна. Старший научный сотрудник Института системного анализа РАН. Окончила МГУ им. М.В.Ломоносова в 1982 году. Кандидат психологических наук. Автор более 70 печатных работ, в т.ч. 1 монографии. Область научных интересов: когнитивная психология, картина мира, психология интернета, психология агрессии. E-mail: nchudova@gmail.com