*На правах рукописи*

**КУЛИКОВ Сергей Юрьевич**

# АВТОМАТИЧЕСКОЕ ИЗВЛЕЧЕНИЕ МНЕНИЙ: ЛИНГВИСТИЧЕСКИЙ АСПЕКТ

Специальность 10.02.21 — прикладная
и математическая лингвистика,
филологические науки

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата филологических наук

Москва — 2016

Работа выполнена в секторе прикладного языкознания
Федерального государственного бюджетного учреждения науки
«Институт языкознания Российской академии наук»

**Научный руководитель:** доктор филологических наук, зав. сектором прикладного языкознания ФГБУН «Институт языкознания Российской академии наук»
**Рябцева Надежда Константиновна**

**Официальные оппоненты:** **Максименко Ольга Ивановна**
доктор филологических наук, профессор кафедры теоретической и прикладной лингвистики Института лингвистики и межкультурной коммуникации ГОУ ВО Московской области «Московский государственный областной университет»

**Ахренова Наталья Александровна**
кандидат филологических наук, доцент кафедры английского языка ГОУ ВО МО «Государственный социально-гуманитарный университет»

**Ведущая организация:** Федеральное государственное бюджетное учреждение науки **Институт научной информации по общественным наукам** Российской академии наук (отдел языкознания)

Защита состоится « *29* » *сентября* 201*6* года в *14* часов 00 минут на заседании диссертационного совета Д 002.006.03 при Институте языкознания РАН, по адресу: 125009, г. Москва, Б. Кисловский пер., д. 1, стр. 1.

С диссертацией можно ознакомиться в библиотеке и на сайте Института языкознания РАН по ссылке http://iling-ran.ru/theses/kulikov_full.pdf

Автореферат разослан «____» _____2016 года.

Ученый секретарь
диссертационного совета
кандидат филологических наук  А. В. Сидельцев

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Реферируемая диссертация посвящена автоматическому извлечению мнений, которое занимается проблемами определения отношения автора текста к описываемой в тексте проблеме, событию или продукту. Системы автоматического извлечения мнений применяются для задач оптимизации информационного поиска, бизнес-разведки, систем фильтрации спама, а также государственного мониторинга Интернета, направленного на предотвращение противоправной деятельности и анализ политической активности населения. На современном этапе наблюдается оторванность разработок в области автоматического извлечения мнений от лингвистических исследований по сходной тематике.

Актуальность исследования обуславливается важностью изучения общественного мнения и разного рода оценочных компонентов в сфере Интернет-коммуникации и необходимостью разработки принципов их автоматического извлечения из различных типов и видов текста.

Научная новизна исследования заключается в разработке лингвистических принципов автоматического извлечения мнений, призванных значительно повысить качество существующих технических средств идентификации субъективных компонентов контента, построенных (в основном) на основе статистических моделей без учета собственно лингвистических факторов. Также в рамках диссертационного исследования уточнена классификация оценочной лексики, в частности введено понятие однореферентных оценочных слов.

Теоретическая значимость исследования состоит в систематизации лингвистической информации, необходимой для задач автоматического извлечения мнений, а также в развитии понятийного аппарата рассматриваемой предметной области. Практическая значимость заключается в разработке принципов повышения качества действующих систем автоматического извлечения мнений за счет совершенствования лингвистического обеспечения. Материалы диссертации могут быть использованы при создании таких ресурсов по автоматическому извлечению мнений, как словари оценочной лексики, базы данных и размеченные корпусы текстов, а также при разработке комплексной системы автоматического извлечения мнений.

В качестве материала исследования выбраны тексты сети Интернет (блоги, сообщения информационных агентств и пользовательские отзывы на продукты и события) на русском языке: корпус ruTenTen (15,8 млрд. словоупотреблений); корпус русскоязычных СМИ и блогов (свыше 25,5 млн. словоупотреблений) и ряд других источников (около 150 тыс. словоупотреблений). Объектом исследования выбраны языковые и внеязыковые способы выражения оценок в электронных текстах. Предметом исследования являются способы формализации оценочных суждений.

Методология настоящего исследования сложилась в первую очередь на базе работ отечественных и зарубежных специалистов в области теории оценки

(Н.Д. Арутюнова, Е.М. Вольф, В.Н. Телия и др.), психолингвистики (И.Н. Горелов, А.А. Леонтьев, Ю.А. Сорокин и др.), теории автоматической обработки текста (Г.Г. Белоногов, Л.Н. Беляева, Ю.Н. Марчук, И.А. Мельчук, А.И. Новиков, Н.В. Лукашевич, R. Schank, Y. Wilks, W. Daelemans и др.), теоретической семантики и синтаксиса (Ю.Д. Апресян, Е.В. Падучева, Л.М. Васильев и др.) и практики автоматического извлечения мнений (А.Н. Соловьев, Т.Е. Загибалов, И.И. Четверкин, J.M. Wiebe, B. Liu, L. Lee, M. Klenner, S. Pulman, M.-T. Taboada и др.).

Специфика объекта изучения обусловила применение следующих методов исследования: классификации, корпусного, контекстуального и дискурсивного анализа, метода «черного ящика», различных статистических методов и метода моделирования. На некоторых этапах работы применялся компонентный анализ.

Целью диссертационного исследования является разработка принципов создания лингвистического обеспечения системы автоматического извлечения мнений для анализа текстов на русском языке.

Для достижения поставленной цели были поставлены следующие задачи:

1. Изучить особенности существующих систем автоматического извлечения мнений;
2. Проанализировать типы оценочной информации, моделируемые в существующих системах;
3. Определить классы оценочной лексики, необходимые для повышения качества автоматического извлечения мнений;
4. Уточнить определение понятия «мнение», принятого в практике автоматического извлечения мнений;
5. Разработать методы фильтрации априорно нейтрального контента;
6. Определить фрагменты этапов автоматического извлечения мнений, зависящие от аспекта задачи;
7. Разработать принципы автоматического и автоматизированного создания оценочных ресурсов для русского языка.

**Степень разработанности проблемы**. Автоматическое извлечение мнений является одной из наиболее динамично развивающихся областей компьютерной лингвистики. В последнее время особую актуальность приобретает разработка новых лингвистических ресурсов, принципов их создания, а также теоретическая обоснованность данных принципов. Необходимо отметить, что для анализа текстов на русском языке в настоящее время не выработано универсальных, общепринятых критериев для автоматического определения той или иной оценки.

**На защиту выносятся следующие положения:**

1. Современное автоматическое извлечение мнений основывается преимущественно на методах машинного обучения. Из методов машинного обучения наибольшее значение получили методы обучения «с учителем», ко-

торые не позволяют оперативно исправлять ошибки классификации текстов на оценочные классы (позитивные, негативные, нейтральные).

2. Оценочная классификация текстов на уровне объектов оценки наиболее полно отражает языковые свойства текстов на естественном языке. Для отдельных типов текста целесообразна иерархическая структура представления объекта, опирающаяся на его свойства. Для иерархической структуры объекта оптимальным представляется использование специализированных тезаурусов или онтологий.
3. При выделении объектов оценки из текстов необходимо проводить разграничение субъект-объектных и причинно-следственных отношений.
4. Для разных задач автоматического извлечения мнений требуется различная организация лингвистического обеспечения. Эти отличия заключаются в наличии или отсутствии дополнительных этапов автоматизированного анализа текстов. Наиболее сложной организацией лингвистического обеспечения обладают системы автоматической идентификации противоправного контента в сети Интернет.
5. При разработке лингвистических ресурсов для автоматического извлечения мнений требуется учитывать механизмы текстовой референции. К данным механизмам относятся деривационные модели имен прилагательных и принципы оценочного словосложения. Ключевым понятием оценочной референции также является ограничение на количество объектов-референтов у оценочных слов.

**Апробация работы**. Материалы диссертации обсуждались на заседаниях сектора прикладного языкознания Института языкознания РАН в 2011—2014 гг.. Основные положения диссертации были изложены на следующих научных конференциях: Всероссийская студенческая научно-практическая конференция «Проблемы современной лингвистики и методики преподавания иностранных языков» (Коломна, 2010-2011), «II Межвузовский студенческий форум по прикладной лингвистике» (Жуковский, 2011), «Международная конференция студентов-филологов» (СПб, 2010-2011), «Актуальные задачи лингвистики, лингводидактики и межкультурной коммуникации» (Ульяновск, 2010, 2012), «Translation and Technology» (TRALOGY-2011) (Paris, 2011), Международная конференция «Язык. Культура. Общество» (Москва, 2011), Международная конференция студентов, аспирантов и молодых учёных «Ломоносов» (Москва, 2011-2014), школа-конференция молодых ученых ИЯз РАН (Москва, 2013-2015), V Международный конгресс исследователей русского языка «Русский язык: исторические судьбы и современность» (Москва, 2014), «Computational Linguistics in the Netherlands and the Flanders» (CLIN 24, CLIN 25) (Leiden, 2014, Antwerpen, 2015). Содержание работы отражено в 24 публикациях, из которых 3 опубликовано в изданиях, рекомендованных ВАК при Минобрнауки России. Некоторые решения, предложенные в работе, нашли применение в модуле лингвистической обработки текста системы «Аналитический Курьер».

**Структура диссертации.** Диссертация состоит из введения, трех глав, заключения, списка использованной литературы, включающего 178 отечественных и зарубежных работ, и четырех приложений, содержащих список мотивационных целей, фрагмент словаря первичной оценочной лексики, фрагмент размеченного корпуса текстов и фрагмент базы данных по этнофобонимам. Общий объем работы составляет 200 страниц: основное содержание изложено на 175 страницах.

## ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **Введении** обосновывается актуальность выбранной темы, ее научная новина, практическая и теоретическая значимость. Также определяются объект диссертационного исследования, задаются его цели, задачи и методы, описывается материал исследования и характеризуется структура работы.

В **Главе 1 «История и современное состояние автоматического извлечения мнений»** рассматриваются предпосылки зарождения области. Прослеживаются предшественники как со стороны искусственного интеллекта (Р. Шенк, Х. Карбонелл, Й. Уилкс), так и со стороны теоретической лингвистики (модель "Смысл-Текст", Н.Д. Арутюнова, Е.М. Вольф, Б.А. Успенский). Поворотным моментом в судьбе автоматического извлечения мнений стало появление Интернета в повседневной жизни (около 2000 г.). Интернет превратил направление с десятком модельных разработок в динамически развивающуюся отрасль. В это время в автоматическое извлечение мнений проникают два смежных течения — корпусная лингвистика (Hatzivassiloglou et al, 1997) и машинное обучение (Turney, 2002).

Массовое практическое внедрение автоматического извлечения мнений заставило исследователей обратить внимание на лингвистическую составляющую (Polanyi et al., 2004). В то же время основную массу исследований продолжают составлять работы, использующие машинное обучение. Современный этап развития автоматического извлечения мнений характеризуется появлением большого количества обзорных работ (например, Pang et al., 2008, Liu, 2012, Taboada, 2016). Также текущий этап знаменуется расширением анализируемых языков (английский, арабский, китайский, испанский, немецкий, итальянский, голландский и русский).

Технология извлечения мнений из текстов на русском языке зародилась в конце 1990-х гг. (Бессмертный и др., 1999), но активно стала развиваться только во второй половине 2000-х (Ермаков, Киселев, 2005). Именно в этот период появляются как теоретические работы (Zagibalov, 2010), так и промышленные системы (А.Е. Ермаков, А.Н. Соловьев). В настоящее время наблюдается значительная дифференциация задач, в которых используется технология автоматического извлечения мнений — мониторинг объектов на основе текстовой классификации (конференция "Диалог" 2012—2016), анализ обсуждений зако-

нов (Толкунов, 2014), мониторинг правонарушений в Интернете (Васечкин и др., 2014).

При этом большинство систем полностью игнорирует теоретические основания оценки, не различая оценочные суждения и интерпретируемые факты, различные нормы поведения и др. Во многом это вызвано поверхностным пониманием задачи извлечения мнений. Обилие альтернативных названий автоматического извлечения мнений (например, *анализ тональности*, *анализ мнений*, *контент-анализ мнений*, *сентимент-анализ*, *opinion mining*, *sentiment analysis*, *subjectivity analysis*) не позволяет оперативно следить за всеми новейшими тенденциями в области. Именно поэтому большую роль в лингвистической составляющей автоматического извлечения мнений играет унификация терминологии (п. 1.10).

В **Главе 2 «Структура лингвистического обеспечения в системах автоматического извлечения мнений»** рассматриваются этапы как непосредственно определения оценки, так и предварительного лингвистического анализа. Этап предварительного анализа текста должен не просто заниматься удалением html-тегов, но и извлекать значимую метатекстовую информацию, например, при помощи XPath. Другим этапом предварительного анализа текста является классификация текстов. Ее задачей является определение подъязыка документа, что способствует сокращению оценочной омонимии, а также уточнению объектов оценки. Например, слово «*спам*» во фразе фраза «*Как же мне надоел спам!*», несущее однозначно негативную оценку и для программистов, и для футбольных болельщиков, для первых будет обозначать оценку по отношению к разработчикам некачественной поисковой системы (или почтового сервера), а для вторых — к футбольному клубу «Спартак (Москва)» и к его болельщикам. Таким образом, указанная фраза в разных подъязыках соотносится с различными референтами, и, следовательно, фраза с форума болельщиков не должна попасть в выдачу по запросу о компьютерном спаме и наоборот.

В разделе 2.2 «Лингвистический анализ текста» рассматриваются особенности применения уровневой модели анализа к задаче автоматического извлечения мнений. Особое внимание при этом уделяется организации компонента автоматического морфологического анализа текста. Синтаксический компонент лингвистического обеспечения в системе автоматического извлечения мнений не играет существенной роли, т.к. может быть заменен набором оценочных правил. Напротив, большое значение в системе приобретает семантический компонент, результаты работы которого используются на последующих этапах фильтрации неоценочных объектов, а также определения объектов и субъектов мнения.

В следующем разделе реферируемого исследования рассматривается **компонент фильтрации объектов анализа**. Помимо подхода с заданием объектов анализа извне (реализованных в системах «Медиалогия» и «Крибрум») приводятся **структурные** (п. 2.3.2) и **семантические** (п. 2.3.3) критерии фильтрации объектов анализа. К **структурным** критериям относятся, прежде всего,

предложения, состоящие из одного слова, которое не входит в оценочный словарь. Таким образом, предложение «Вася» не будет подвергаться оценочному анализу, а предложение «Крымнашист» — будет. Другим типом фильтруемых предложений являются условные придаточные предложения, которые описывают лишь один из возможных сценариев развития ситуации. Последним из структурных фильтров является вопросительное предложение, оканчивающееся на вопросительный знак. Предложения вида *««Откуда он?» — спрашиваю я, указывая на флаг Новороссии.»* или «*С. КОРЗУН: В чем предмет спора и разногласий был основной с Фишером?*» не анализируются, в то время как предложения вида *«Какой отец?!»* анализируются по особой модели. К **семантическим** критериям фильтрации объектов относятся: комплексные именованные сущности и оценочные фрагменты бессвязочных предложений. В комплексных именованных сущностях не анализируются именованные части. Так, названия фильмов, законов и организаций оценочно нейтральны. Например, *Комитет по противодействию коррупции, Закон «О защите прав потребителей»* или «*В новелле «Убийца» наш герой добирается до самого центра степей, в столицу Мараху, затем посещает диких горцев, чтобы заручиться поддержкой в случае нападения на его королевство воинственной империи Мардинан.*». С другой стороны, при наличии модификаторов неименованной части, именные группы, соответствующие комплексным именованным сущностям, являются оценочными. Например, *Его главные произведения — это окопные дневники, а также, говоря по-простому, профашистская книга «Рабочий и гештальт»*. Оценочные компоненты бессвязочных предложений, в отличие от предыдущего случая, продолжают выражать оценку, но не могут становиться объектами анализа. Например, в предложении «*Павел Рязанцев - молодец, забил очень нужный гол*» слово *молодец* исключается из списка потенциальных объектов анализа. В то же время в составе именной группы слово *гений* является частью объекта, например, *Наш величайший гений Пушкин написал однажды…*

Раздел 2.4 (**Компонент автоматического извлечения мнений**) является центральным в реферируемой диссертации. Структурно он состоит из 3 частей, описывающих разные классы систем автоматического извлечения мнений: системы автоматического извлечения мнений общего назначения, специализированные системы и системы идентификации противоправного контента. Структура лингвистического обеспечения системы автоматического извлечения мнений общего назначения состоит из двух больших блоков предварительного компонента обработки, состоящего из этапов подготовки текста (фильтрации контента) и лингвистического анализа текста (Рисунок 1), и собственно модуля автоматического извлечения мнений (Рисунок 2).


**Рисунок 1. Схема взаимодействия модулей предобработки и лингвистического анализа текста**

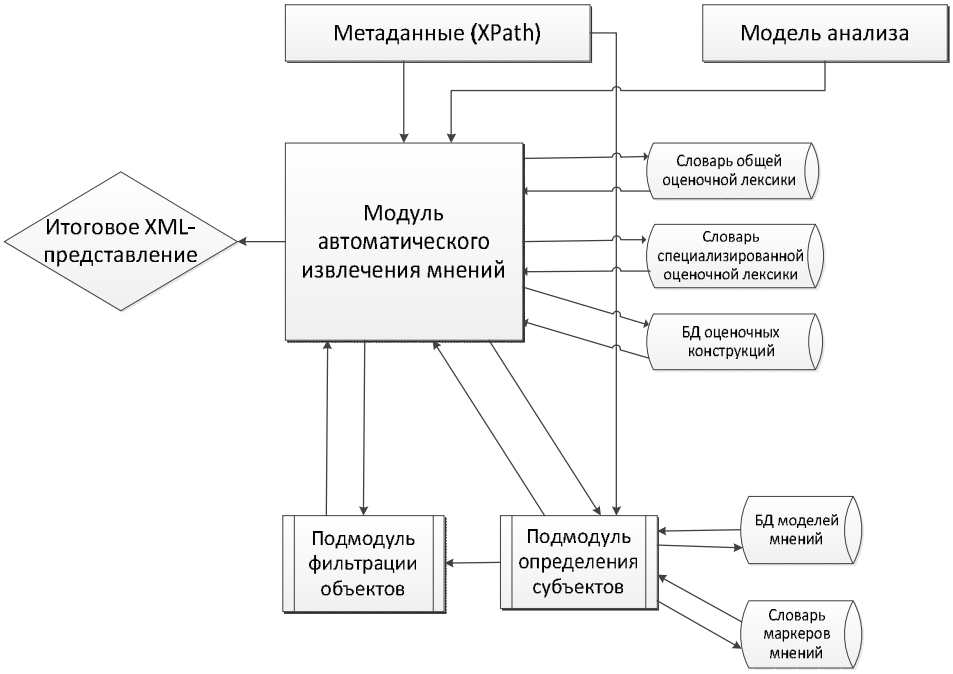


**Рисунок 2. Обобщенная схема модуля автоматического извлечения мнений**

В **подмодуле определения объекта** (п. 2.4.1.2) рассматриваются способы задания объектов через онтологии и при помощи эвристик. Сложным случаем для данного подхода являются сравнительные конструкции, особенно с имплицитным заданием объекта.[1] Например, для предложения из выборки *Маша самая красивая девушка на свете!* необходимо учитывать вхождение объекта

[1] В примере ниже объектом анализа является *участница конкурса красоты Юля*.

анализа *Юля* в онтологический класс *девушки*. Сложнее ситуация с текстами экономической сферы. Тексты данной предметной области имеют особую специфику, зачастую выражая оценочные суждения в числовой форме, например, *Газпром вошел в десятку крупнейших компаний мира*. Для правильной интерпретации предложения *Казахгаз вошел в тройку мировых лидеров добычи газа в 2011 году* относительно объекта *Газпром* важно знать место *Газпрома* в данном рейтинге за такой же период, что ведет к экспоненциальному рост базы знаний при падении скорости обработки. В результате можно констатировать, что применение онтологий при определении объектов должно быть ограничено узкими предметными областями, список объектов в которых конечен.

Как отмечалось выше, значительную сложность вызывают имплицитно задаваемые объекты. К ним относят сравнительные конструкции и отсылки к объекту через его свойства [Liu, 2012], анафорические упоминания [Ma & Wan, 2010], хештеги [Wan et al., 2013] и точку зрения (т.е. интерпретацию фактов) [Eidelman, 2012]. По нашему мнению, к имплицитным способам задания мнений также следует отнести совпадение в одном слове объекта оценки и маркера оценки, т.н. однореферентные оценочные слова, например, *антироссийский*[2].

В ряде предметных областей, например, сфере отзывов, представляется целесообразным помимо объектов определять и их свойства. Свойства объекта могут задаваться через онтологии, метаданные (при помощи технологии XPath), машинное обучение и эвристические правила.

**Подмодуль определения субъекта** (п. 2.1.4.3) непосредственно связан с моделью мнения. Процедура определения мнения представляет собой процесс декодирования сообщения в рамках электронной коммуникации. Таким образом, модель мнения должна в определенной мере отражать модель передачи сообщения. Существующие подходы к формальному заданию структуры мнения (Б. Лю, А.Н. Соловьев, А.А. Толкунов) игнорируют существующие наработки в области передачи информации. Наиболее адекватной из моделей передачи сообщений нам представляется *модель Шеннона-Уивера*. В соответствии с данной моделью помимо субъекта мнения целесообразно определять и *канал передачи сообщения*. Таким образом, модель мнения, по нашему представлению, имеет следующий вид:

$$Opinion = \{Obj; Sent; Subj; Chan^*; Source^*; Date^*; Feat^*\}, \quad (1)$$

где $Opinion$ — мнение, $Obj$ — объект анализа, $Sent$ — оценка объекта, $Subj$ — субъект мнения, $Chan$ — канал передачи сообщения, $Source$ — источник мнения, $Date$ — дата сообщения, $Feat$ — свойство (характеристика объекта), * — маркер факультативного элемента.

В реферируемой работе рассмотрены следующие ключевые способы представления субъекта мнения — прямая речь[3] (п. 2.4.1.3.1) и косвенная речь[4]

[2] негативно относительно России.
[3] «Где моя охрана?!» – пропищал 79-летний швейцарец

(п. 2.4.1.3.2). Кроме того, субъект может встречаться в конструкциях со свернутой косвенной речью[5] (п. 2.4.3.3). Все конструкции, кодирующие субъект, представляются в виде особой базы данных, в которой в отдельном поле указывается тип данной конструкции. Задача разграничения субъекта мнения от канала передачи мнения и источника мнения решается после определения первичного (формального) субъекта мнения. Одним из случаев, когда подобное разграничение является наиболее оправданным, можно считать двойное цитирование[6]. Двойное цитирование является, по-видимому, частным случаем n-арного цитирования, где n≤9, в соответствии с гипотезой В.Ингве. Так, в предложении *Активистки также приводят слова пресс-секретаря Путина Дмитрия Пескова, растиражированные российскими СМИ: «Это откровенное хулиганство.»* можно выделить три источника мнения, которые отличны от его субъекта (Путин). Таким образом, задачей подмодуля определения субъекта мнения является правильное определение всех компонентов следующей схемы (Рисунок 3).


**Рисунок 3. Обобщенная схема сообщения**

В ходе разработки правил определения субъекта оценки были выявлены некоторые **проблемы формального описания морфологии русского языка**. Во-первых, это проблема одушевленности собирательных существительных. Согласно классическим грамматическим описаниям [Клобуков, 1988] собирательные существительные в русском языке относятся к неодушевленным по формальным критериям (например, винительный падеж, сочетаемость с прилагательными во множественном числе). С другой стороны, категория одушевленности/неодушевленности для некоторых языков (например, английского) является скрытой [Виноградов, 1998 (1990)]. Если предположить, что для русского языка одушевленность собирательных существительных также является

[4] Роман Скорый считает, что в Нижнем Тагиле нужно развивать индустриальный, военно-технический и патриотический туризм.
[5] Глава Североатлантического альянса Йенс Столтенберг заявил об итогах экстренной встречи, на которой обсуждались угрозы Турции.
[6] Т.е. случай несовпадения источника мнения и канала передачи сообщения.

скрытой категорией, то это не только поможет решить сугубо техническую задачу определения субъектов мнений, но также поможет объяснить сочетаемость семантического класса «организация» с речемыслительными глаголами. Исходя из предположения, что собирательные существительные наследуют категорию одушевленности от составляющих их элементов, можно также вывести правила их сочетаемости с некоторыми оценочными глаголами, например, «*ненавидеть*». Сравним три предложения, в которых у глагола *«ненавидеть»* разные субъекты, а именно одушевленное существительное, собирательное существительное со значением «совокупность» людей и неодушевленное существительное.

*Путин патологически ненавидит Лукашенко.*

*Известный российский актер Владимир Епифанцев заявил, что российская молодежь ненавидит Русскую Православную церковь (РПЦ).*

*Альфа-банк ненавидит надёжных клиентов.*

В первых двух примерах субъекты глагола «*ненавидеть*» оценочно нейтральны, а в третьем — негативен. Таким образом, собирательные существительные, обозначающие совокупности людей, в оценочном плане ведут себя подобно другим одушевленным существительным. Ещё одним свидетельством в пользу одушевленности собирательных существительных, обозначающих группы людей, является их способность выступать в качестве субъекта при наиболее типичных представителях речемыслительных глаголов, таких как разговаривать, считать и думать. Продемонстрируем это на следующем примере:

*Часто наблюдаю, как молодежь выходит на дорогу, разговаривая по телефону, и даже не смотрит по сторонам*.

Во-вторых, это проблема **формального представления словоизменительных парадигм существительных-заимствований**. Данная проблема характерна для этнорелигиозной лексики. С точки зрения задач автоматического извлечения мнений проблема заключается в том, что не происходит отождествления одинаковых субъектов и/или объектов мнений. Этнорелигиозная лексика зачастую бывает представлена двумя морфологическими вариантами: 1) изменяющимся по законам языка-реципиента, 2) продолжающим изменяться по законам языка-донора. Именно слова второго типа вызывают сложности при автоматическом морфологическом анализе текста. Это вызвано тем, что существующие алгоритмы морфологического анализа моделируют морфологическую систему языка-реципиента [Белоногов и др., 1965, 1979; Зализняк, 1967; Voutilainen, 2003].

Существование этнонимов типа *рейхсдойч*, *ашкеназ*, имеющих три и две словоизменительных модели (*рейхсдойч — рейхсдойчи|рейхсдойче|—*; *ашкеназ — ашкеназы|ашкеназим*), и религионимов типа *моджахед*(*д*) и *муртад*(*д*), имеющих по две словоизменительные парадигмы (традиционная *-ы* и заимствуемая *-ин*), требует объединения в единой единице анализа всех существующих парадигм. В качестве выхода из данной ситуации можно использовать

комплексные «понятийные» леммы, представленные ниже (Рисунок 4). Леммы при таком подходе могут содержать различные основы, что позволяет описать специфику этнонимов с разными основами. В отличие от тезаурусного представления «понятийная» лемма является более простой структурой, что позволяет быстрее оперировать ее данными.

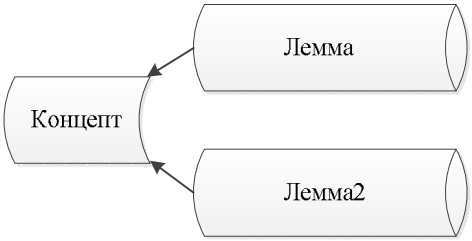


**Рисунок 4. Упрощенная структура представления «понятийных» лемм**

**Подмодуль определения причинно-следственных связей** (п. 2.1.4.4) не является обязательным компонентом системы автоматического извлечения мнений, однако в последнее время выявление причины оценки признается практически столь же значимым, как и само определение позитива/негатива в высказывании по отношению к продукту. К данному этапу из потенциальных причинно-следственных отношений уже отфильтрованы условные предложения. Остальные причинно-следственные отношения можно разделить на следующие 4 типа: интрасентенциальные (п. 2.4.1.4.2), интерсентенциальные (п. 2.4.1.4.3), внутриклаузные (п. 2.4.1.4.4) и лексикализованные (п. 2.4.1.4.5). Последовательность их обработки приведена ниже (Рисунок 5).

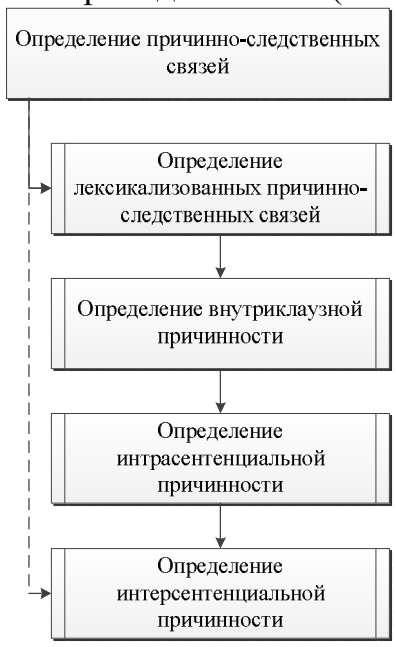


**Рисунок 5. Схема подмодуля определения причинно-следственных связей**

Интрасентенциальные, интерсентенциальные и внутриклаузные случаи причинно-следственных отношений определяются при помощи шаблонов, основанных как на технологии XPath, так и на основе синтаксических шаблонов.

Лексикализованная причинность[7] выражается при помощи деадъективов или девербативов. Другим способом ее выражения являются перечисления и обобщения, например, *Плюсы: - высокое качество исполнения. - отличные точильные свойства*. В качестве наиболее точного способа обнаружения лексикализованной причинности, на наш взгляд, выступают шаблоны. По своей структуре такой способ выражения причинности характерен для вторичных текстов.

Центральным компонентом в системе автоматического извлечения мнений является **словарь оценочной лексики**, описанный в п. 2.4.1.5. В данном разделе реферируемого исследования проанализированы существующие лингвистические подходы к организации словарей оценочной лексики и их интерпретации (п. 2.4.1.5.1), способы представления силы оценки (п. 2.4.1.5.2), также рассмотрены оценочные классы лексики и их представление в словаре, а также в статистических системах (п. 2.4.1.5.3). Затем описаны возможности учета частеречной (п. 2.1.5.4) и фразеологической информации (п. 2.1.5.5). Далее приведены правила сочетания оценочных слов (п. 2.1.5.6). В заключительной части раздела проведен обзор недостатков подходов, основанных на машинном обучении (п. 2.1.5.7).

Обзор существующих лингвистических подходов к организации оценочных словарей демонстрирует, что в части основных тенденций сохраняется главенство словарного компонента, а также стремление к отказу от чисто эвристических или статистических методов в пользу гибридных. Тенденция к гибридным методам, по нашему мнению, объясняется невозможностью описать в рамках существующих лингвистических моделей оценки пограничные случаи сочетаемости (т.н. *polarity conflicts*), а также более высоким быстродействием статистических алгоритмов. К недостаткам существующих методов можно отнести игнорирование семантических классов слов, а также отсутствие этапа фильтрации неоценочных конструкций и инкорпорации неравноправно-оценочных глаголов. Отдельную проблему составляют имплицитные оценки, которые находятся вне рассмотрения существующих лингвистических подходов.

Классификация оценочной лексики представляет собой одну из наиболее сложных проблем в области автоматического извлечения мнений. Традиционное деление (Табоада, Кленнер) оценочной лексики на 6 классов (положительный, отрицательный, нейтральный, усилительный, уменьшительный и класс переключателей) не в полной мере отражают лингвистическую действительность. В реферируемой работе выделены 12 классов оценочной лексики:

1. положительные слова и словосочетания (*замечательный*, *талантище*);
2. отрицательные слова и словосочетания (*ужасный*, *алчность*);
3. нейтральные слова и словосочетания;[8]

[7] языковое явление, при котором языковые средства, выражающие следствие, одновременно выражают оценку, например, *По подбору игроков, по нынешней форме «Зенит» — явный фаворит*.

[8] не отражаются в словаре, за исключением особых случаев, описанных в п. 2.4.1.5.5.

4. усилители оценки (*более*, *гораздо*);
5. уменьшители оценки (*менее*, *еле*);
6. оценочные переключатели (шифтеры) (*отнюдь*, *не*);
7. положительные усилители (*влиятельный*);
8. отрицательные усилители (*злостный*);
9. субъектно-нейтральные слова и словосочетания (*оштрафовать*, *одобрить*);
10. субъект-позитивные глаголы (*победить*, *накостылять*);
11. субъект-негативные глаголы (*потворствовать*);
12. однореферентные слова (*тьфутболист*, *Мелкобритания*).

Данные классы оценочной лексики распределены по частям речи неравномерно, поэтому представляется важным учитывать информацию о части речи в словарном компоненте системы извлечения мнений. Существует два способа представления частеречной информации в оценочном словаре. Первый заключается в создании отдельных списков слов с указанием на часть речи (и соответствующий оценочный тип). Данное указание может быть выражено не эксплицитно, а при помощи настроек в программном коде. Второй подход заключается в кодировании указаний на часть речи в специальном поле базы данных. Первый способ обеспечивает возможность более быстрого пополнения словника, в то время как второй способствует минимизации ошибок при расширении словаря. Лексикографическое представление фразеологизмов и терминологических сочетаний должно учитывать их разделение на содержащие вариативную часть и нерасчленяемые. В словаре первый тип будет представлен в виде псевдорегулярного выражения, например, *как * на новые ворота*[9]. Нерасчленяемые фразеологизмы заносятся целиком в отдельный словарь.

Уровневая модель правил (п. 2.4.1.6) описывает правила оценочной навигации по дереву синтаксического разбора предложения[10]. Правила моделируют восходящий синтаксический анализ, связывая оценочными отношениями ближайшие узлы дерева. Оценочные конфликты решаются при помощи конфигурации, а при ее отсутствии по следующей схеме:

$$\exists\, Verb_{ev} \rightarrow Obj_{ev} = Verb_{ev} \;\&\; \nexists\, Verb_{ev} \rightarrow Obj_{ev} = (NP_{ev} | W_{ev}), \qquad (2)$$

где $Verb_{ev}$ — оценочный глагол; $Obj_{ev}$ — итоговая оценка объекта анализа; $NP_{ev}$ — оценка именной группы, содержащей объект анализа; $W_{ev}$ — лексикализованная оценка объекта анализа.

Особым случаем представления лексики в системе автоматического извлечения мнений являются лексикализованные правила (п. 2.4.1.6.2), в которых описаны правила сочетания низкочастотных оценочных классов, таких как уменьшители и переключатели оценки.[11] Дискурсивные правила (п. 2.4.1.6.3) описывают взаимодействие оценочного компонента с подмодулем определения

[9] *Как Барак на новые ворота. Турция: как Туран на новые ворота.*
[10] а при неполном синтаксическом анализе система сама строит подобное дерево.
[11] в совокупности составляющие 0,003 % всех оценочных слов.

причинно-следственных отношений, а также правила вывода общей оценки в случае нескольких вхождений объекта в анализируемый текст.

Подавляющее большинство работ в области автоматического извлечения мнений опираются на **методы машинного обучения** [библиография в Liu, 2012, Диалог-2015]. При использовании различных методов машинного обучения задача автоматического извлечения мнений рассматривается как частный случай классификации текста. В работах, которые оперируют объектами, при таком подходе считается достаточным вхождения объекта анализа в позитивный или негативный текст [Четверкин, 2012, Taboada, 2016]. Подобное упрощение оправдано для текстов, содержащих только один объект [ср. Liu, 2012]. Другой проблемой статистических методов считается адаптация к новой предметной области и/или сайту. Это связано как с оценочной спецификой различных предметных областей, так и со стилистическими особенностями текстов различных жанров. В результате применения классификатора к тексту, который существенно отличается от текстов в обучающей коллекции, нарушается принцип однородности анализируемых данных. В таких случаях статистические методы неприменимы к задачам анализа текста [Тулдава, 1987; Марчук, 2010]. Другой недостаток машинного обучения заключается в их «непредсказуемости» [Sonntag et al., 2014: 188], а также невозможности оперативного исправления ошибок алгоритма. Большинство **методов обучения с учителем**, применяемых для решения задачи оценочной классификации текстов, оперирует комбинацией метода «мешка слов» (bag-of-words) и оценочных шифтеров. В качестве обучающего корпуса при этом используется массив текстов, в котором каждый текст имеет общую разметку, а также содержит оценки для каждого слова в корпусе. На основе этих параметров строится векторное представление текста, которое подается на вход алгоритму классификации, например, опорных векторов. Применение метода опорных векторов объясняется его высокими показателями в плане точности классификации. Декодирование модели проводится при помощи модели косинусного подобия.

Специализированные системы автоматического извлечения мнений рассмотрены в п. 2.4.2. В отличие от систем автоматического извлечения мнений общего назначения, они направлены на обработку узкой предметной области или решение частной задачи. Как видно из схемы (Рисунок 6), задачи, решаемые в специализированных системах, разнообразны. Это реферирование мнений, определение спам-отзывов и идентификация социально-значимого контента.

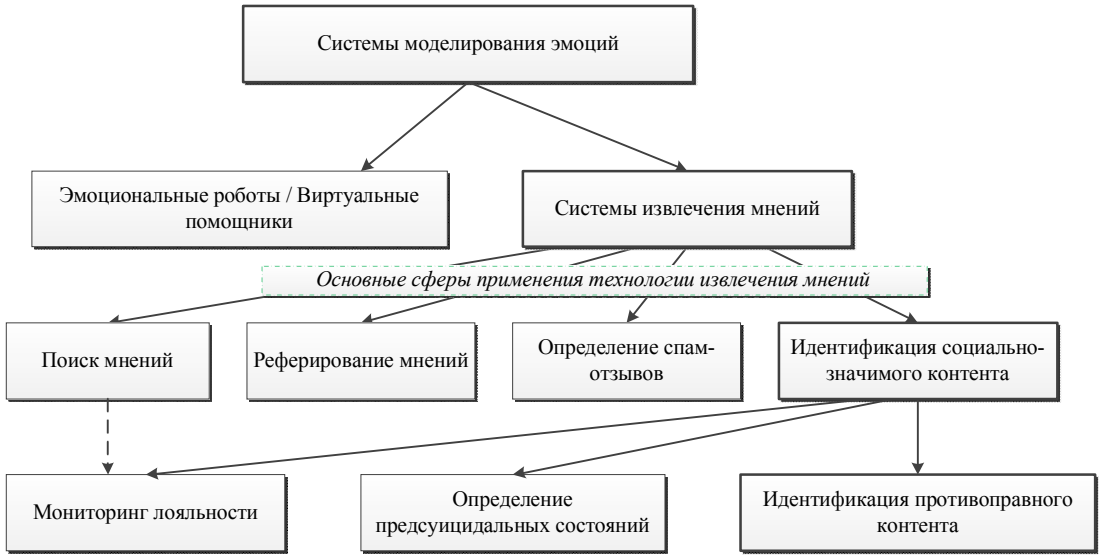


**Рисунок 6. Упрощенная классификация систем моделирования эмоций**

Одним из ключевых отличий специализированных систем автоматического извлечения мнений от систем общего назначения является организация их словарного компонента (Рисунок 7).

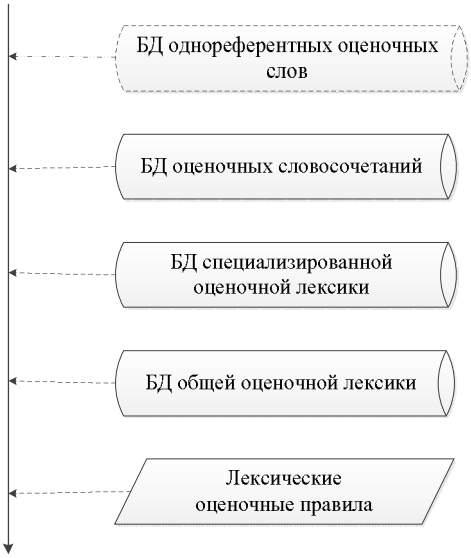


**Рисунок 7. Упрощенная схема взаимодействия словарных компонентов специализированной системы автоматического извлечения мнений**

За счет введения дополнительных словарных ресурсов достигается максимально точное моделирование предметной области. Представленная схема является гибкой, т.к. специализированные словари могут подключаться для любой предметной области. При этом также возможно использование нескольких словарей, последовательность обработки которых указывается в конфигурации системы. Известным ограничением специализированных систем извлечения мнений является невозможность комбинации различных моделей мира, что связано с детальной проработкой одной конкретной модели мира зачастую под конкретного заказчика.

Особой сферой применения технологии автоматического извлечения мнений является **информационная безопасность** (п. 2.4.3). Ключевыми направлениями в этой сфере следует признать онлайн-мониторинг, автоматическую модерацию комментариев пользователей и программные средства ограничения доступа к веб-контенту. Перечисленные направления имеют принципиально разных заказчиков. Самый широкий спектр заказчиков у систем онлайн-мониторинга, в то время как автоматической модерацией комментариев занимаются только онлайн-СМИ и социальные сети. Последний тип систем является исключительно прерогативой государства. Наибольшее значение при **онлайн-мониторинге** уделяется динамике объекта (упоминания, изменение лояльности, локализация). Из этих аспектов к автоматическому извлечению мнений относится только *лояльность*[12]. В некоторых системах мониторинга также строятся рейтинги лояльности по источникам (что важно при идентификации заказных кампаний, направленных на ухудшение репутации объекта). Для построения индекса лояльности используется *поиск мнений*. Он заключается в идентификации объекта анализа в заданном оценочном контексте. Существует три способа **модерации комментариев** и публичных сообщений в Интернете — ручная, автоматизированная и автоматическая. Ручная модерация представляется нецелесообразной при большом потоке данных. Автоматизированная модерация может осуществляться в двух режимах — «жадном» и «ленивом». В первом случае комментарии недоступны пользователю, а во втором скрыты от непосредственного наблюдения. Наиболее распространенным приемом автоматизированной модерации является проверка комментария на наличие ключевых слов (например, слова «революция», «протест» и нецензурная лексика). Ключевым недостатком простого словарного подхода является отсутствие средств «нейтрализации» оценки. К наиболее значимым нейтрализаторам можно отнести: цитаты из общепризнанных источников (Библия, Коран) или классиков литературы, а также употребление согласованных лексических маркеров, например, *слово*, *термин*, *понятие*. Автоматическая модерация комментариев отличается от автоматизированной только отсутствием подэтапа ручной проверки. У последних двух способов модерации комментариев есть один ключевой недостаток, сказывающийся на полноте идентификации негативно-оценочных случаев. Это языковая игра и оценочное словообразование, которые не имеют однозначного решения ни в компьютерной, ни в теоретической лингвистике. **Ограничение доступа к веб-контенту** является значимой опцией и для информационно-поисковых систем, и для веб-браузеров. В данных системах фильтрация контента проводится аналогично автоматической модерации комментариев. Другим случаем является автоматическое определение противоправного контента, как например, в системе Роскомнадзора [Васечкин и др., 2014]. Система представляет собой набор статистических классификаторов по параметрам, которые соответствуют определениям УК РФ (национальная, ре-

[12] соотношение положительных и отрицательных упоминаний объекта

лигиозная рознь, экстремизм и др.). Из-за невозможности изучения названной системы рассмотрим лингвистические принципы построения систем подобного класса. В общем виде автоматизированная система имеет следующий вид (Рисунок 8).


**Рисунок 8. Упрощенная схема автоматизации АРМ модерирования**

В отличие от систем общего назначения в данном классе систем извлечения мнений компонент фильтрации учитывает не отдельные семантические типы слов и словосочетаний, а тип высказывания (*факт* или *суждение*) и наличие слов-нейтрализаторов. При этом объем работы эксперта сокращается при помощи этапа однозначной идентификации слов и высказываний. Однозначная фильтрация может быть настраиваемой при помощи файла конфигурации.

В **Главе 3 «Разработка принципов автоматизированного составления ресурсов для автоматического извлечения мнений»** рассматривается методика автоматизации создания первичного словаря оценочных прилагательных (раздел 3.1). Обобщенное описание формата для корпуса текстов, предназначенного для тестирования систем автоматического извлечения мнений на уровне объектов, дается в разделе 3.2. В разделе 3.3 описана структура и принципы использования лексической базы данных по этнофобонимам. В разделе 3.4 приведен способ автоматического заполнения полей лексической базы данных по этнофобонимам с использованием технологии морфемного синтеза.

При создании **специализированного первичного словаря оценочной лексики** вначале на основе сверхбольшого корпуса текстов (объемом более 15 млрд. словоупотреблений) создается список слов, отражающих определенную предметную область. Формирование списка проводится при помощи технологии SketchEngine. На основе одного ключевого слова (*фильм*) строится квазисинонимический тезаурус, с порогом близости от 0,8. С целью расширения полученного списка слов процедура повторяется для наиболее статистически близких слов. На втором этапе создается список прилагательных, сочетающихся со словами из полученного тезауруса. После этого необходимо провести

фильтрацию низкочастотных прилагательных[13] с целью отсева ошибок предобработки и сокращения последующей ручной проверки на оценочность. Получившийся список прилагательных составил 1514 слов. Затем при помощи технологии морфемного синтеза[14] список был расширен до 2966. Данная методика отличается от альтернативных вариантов отсутствием привлечения готовых словарей прилагательных [De Smedt & Daelemans, 2012] или слов предметной области [Fahrni & Klenner, 2008].

При определении **формата размеченного корпуса**, предназначенного для тестирования качества систем извлечения мнений, необходимо опираться не на метатекстовую информацию (пользовательская оценка), а на структурную модель мнения. В качестве формата метаописания рассмотрен формат XML, который позволяет учитывать неограниченное количество характеристик.

**Тезаурусному описанию однореферентных оценочных слов** в реферируемой работе посвящен раздел 3.3. Подобные слова могут обозначать как положительную, так и отрицательную оценку. Пример отрицательной оценки приведен ниже.

*В своем интервью на Эхе воршебник Чуров прямо обвинил телевизионщиков в искажении фактов.*

Среди однореферентных слов особого внимания заслуживают социально значимые: ксенофобонимы[15] (включая этно-, социо- и религиофобонимы) и персонофобонимы[16]. Для идентификации слов в исследовательском корпусе текстов применялось регулярное выражение вида \w+(?>фоб)\w*[17]. Наиболее удобным форматом представления ксенофобонимов является тезаурус. В работе описан формат тезауруса, содержащий 6 полей (MAIN, ID, POS, SENTIMENT, INT, VALUE) и 5 типов отношений (синонимическое (SYN), агентивное (AGENT), общее (GENERIC), модификатор (MODIFIER)). Общее количество референтных слов в базе этнофобонимов составляет 85. В заключительном разделе главы описан способ **автоматического пополнения словаря этнофобонимов** при помощи технологии морфемного синтеза. Технология **морфемного синтеза** заключается в добавлении или замене определенных аффиксов (квазиаффиксов) в словах с целью получения определенного лексического или прагматического значения. Для ксенофобонимов были выбраны прототипические компоненты «-фоб» и «-ненавистник». Кроме того, были разработаны правила преобразования этих компонентов в формы женского рода (например, *-фобка*, *-ненавистница*) и отвлеченных существительных (*-фобия*, *-фобство*). Модели образования прилагательных были определены при помощи морфологического словаря ИПС Stocona [Огарок, 2008]. Алгоритм пополнения словаря приведен

[13] предварительно проведя повторную лемматизацию и пересчитав частоты.
[14] процедура описана в Приложении 2.
[15] слова, обозначающие негативную оценку относительно какого-либо дифференцирующего признака, отсутствующего у автора сообщения.
[16] слова, обозначающие негативную оценку конкретного человека.
[17] элемент -фоб- был выбран в качестве прототипического.

на схеме (Рисунок 9). У предлагаемого метода есть ряд недостатков, например, он не позволяет учесть оценочное словосложение (например, *Наглия* от *наглый* + *Англия*) и специализированные религиозные термины (например, *такфир*, *кяфир*, *муртад*).

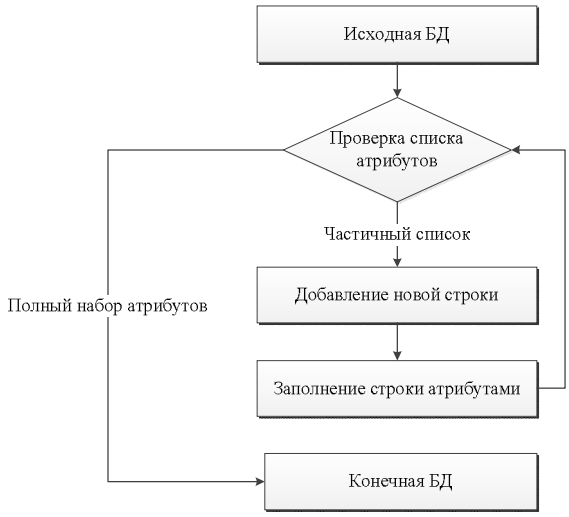


**Рисунок 9. Блок-схема пополнения словаря**

В **Заключении** подведены итоги проведенного исследования. Основные результаты работы состоят в следующем:

1. В современных системах автоматического извлечения мнений из текстов на русском языке применяются, преимущественно, статистические методы, а задача извлечения мнений рассматривается как частный случай классификации текстов.
2. Зародившись как междисциплинарное прикладное направление, находящееся на стыке компьютерной лингвистики и искусственного интеллекта, автоматическое извлечение мнений в настоящий момент не обладает устоявшейся терминологической системой.
3. В системах автоматической обработки текстов, ориентированных на язык Интернет-коммуникации, большое значение приобретает компонент предобработки текстов. В состав данного компонента необходимо включать инструменты анализа метатекстовой информации, которые позволяют учесть априорно заданную прагматическую информацию.
4. Обязательным компонентом лингвистического обеспечения систем автоматического извлечения мнений является фильтрация нерелевантных объектов анализа. Фильтрация данных объектов осуществляется на основе структурных и семантических критериев.
5. Формальная модель мнения, используемая в системе автоматического извлечения мнений, должна учитывать особенности функционирования коммуникации в Интернет-

пространстве. Поэтому она должна включать информацию об источнике мнения и канале передачи сообщения.

6. Наиболее адекватным способом описания объектов мы считаем объект-аспектную модель, которая позволяет учитывать свойства объекта мнения. При этом данная модель не является универсальной и не подходит для любого типа текстов.
7. Структура лингвистического обеспечения систем автоматического извлечения мнений в значительной степени зависит от конкретной задачи анализа текста. Наиболее сложной задачей мы считаем идентификацию противоправного контента, для решения которой требуется не только наличие особого набора специализированных оценочных словарей, но и значительно более жесткие требования к фильтрации входных данных.
8. При оценке качества систем автоматического анализа мнений мы считаем целесообразным использовать размеченные корпусы текстов, разметка которых полностью отражает модель мнения. Такая разметка позволяет определять качество работы не только системы в целом, но и отдельных компонентов.
9. Для автоматического определения некоторых случаев имплицитной оценки необходимо учитывать референтные свойства оценочных слов. Среди однореферентных оценочных слов особое социальное значение имеют ксенофобонимы, обозначающие отрицательную оценку объекта заданного класса (например, этнической, религиозной или социальной группы и отдельного человека). Наиболее адекватным инструментом для формального представления однореферентных оценочных слов мы считаем тезаурус.

**ПУБЛИКАЦИИ**

По теме диссертации автором опубликованы следующие работы, в том числе в ведущих рецензируемых изданиях ВАК при Минобрнауки РФ:

1. **Куликов С.Ю. Автоматизация составления оценочного словаря широкой предметной области [Текст] : (опыт использования неспециализированного корпуса текстов) / С. Ю. Куликов // Вестник Иркутского Государственного Технического Университета. — 2014. -— № 8. — С. 240—243.**
2. **Куликов С.Ю. Однореферентные оценочные слова и их формализация [Текст] (на примере этнофобонимов) / С. Ю. Куликов // Вопросы психолингвистики. — 2014. — № 4 (22). С. 166—173.**
3. **Куликов С.Ю. Морфемный синтез как способ автоматического пополнения словаря оценочной лексики тезаурусного типа (на материале ксенофобонимов) / С. Ю. Куликов // Вопросы филологии. — 2015. — № 2 (50). С. 91—98.**
4. Куликов С.Ю. Оценочные усилители в тексте // Перевод и когнитология в XXI веке. Сборник студенческих научных статей (по материалам конференции 13 апреля 2010 г.). — М., 2010. — С. 68—73.
5. Куликов С.Ю. Оценочная шкала в системах автоматического извлечения мнений // XIII межвузовская научная конференция студентов-филологов. Тезисы. Часть 3. — СПб., 2010. — С. 35—36.
6. Куликов С.Ю. Некоторые вопросы описания положительных оценок именных групп для задач автоматического извлечения мнений // Актуальные проблемы теоретической и прикладной лингвистики: сборник научных трудов. — Ульяновск, 2010. — С. 118—122.
7. Куликов С.Ю. Типология оценочных шифтеров при автоматическом извлечении мнений // Проблемы современной лингвистики и методики преподавания иностранных языков: сб. тезисов науч.-практич. конф. для студентов. — Коломна, 2010. — С. 78—81.
8. Куликов С.Ю. Автоматическое извлечение мнений как новая информационная технология // Теоретические и практические вопросы межкультурной коммуникации. Материалы научно-методической конференции преподавателей, аспирантов и студентов МГОУ. Студенческие работы. — М. 2010. — С. 69—71.
9. Куликов С.Ю. Проблема однородной выборки при отборе ресурсов для задач автоматического извлечения мнений // Прикладная лингвистика сегодня и завтра: актуальные проблемы: Материалы Межвузовского студенческого форума по прикладной лингвистике, 18 февраля 2011 г.: Вып.1. — Жуковский, 2011. — С. 47—49.
10. Куликов С.Ю. Формальные свойства текста как фактор субъективности // Проблемы современной лингвистики и методики преподавания ино-

странных языков: сб. тезисов науч.-практич. конф. для студентов. — Коломна, 2011. — С. 99—102.

11. Kulikov S. What is web-based machine translation up to? // Proceedings of TRALOGY-2011 «Métiers et technologies de la traduction : quelles convergences pour l'avenir?» Paris, France, 2011 Electronic publication: http://lodel.irevues.inist.fr/tralogy/index.php?id=118

12. Куликов С.Ю. Контекстологическое определение оценки именной группы // XIV Международная научная конференция студентов-филологов. Тезисы. Часть 4. — СПб., 2011. — С. 26—27.

13. Куликов С.Ю. Морфемная структура оценочных слов в английском языке (корпусное исследование) // Международная конференция «Ломоносов-2011». Секция «Филология». — М., 2011. — С. 593—594.

14. Куликов С.Ю. К вопросу о месте автоматического извлечения мнений в рамках компьютерной лингвистики // VI Международная научная конференция «Язык, культура, общество». Тезисы докладов. Том 1. — М., 2011. — С. 94.

15. Куликов С.Ю. Общение в Интернете и новые вызовы компьютерной лингвистике // Материалы Международного молодежного научного форума «ЛОМОНОСОВ-2012» / Отв. ред. А.И. Андреев, А.В. Андриянов, Е.А. Антипов, К.К. Андреев, М.В. Чистякова. [Электронный ресурс] — М.: МАКС Пресс, 2012. http://lomonosov-msu.ru/archive/Lomonosov_2012/1911/32293_74bc.doc

16. Куликов С.Ю. Использование моделей глагольного управления для задач автоматического извлечения мнений // Актуальные задачи лингвистики, лингводидактики и межкультурной коммуникации. Материалы 5-й Международной научно-практической конференции (20–21 сентября 2012): сборник научных трудов / отв. ред. Н.С. Шарафутдинова.— Ульяновск: УлГТУ, 2012.

17. Куликов С.Ю. Теория подъязыков и автоматическое извлечение мнений // Сборник научных трудов, посвященный юбилею проф. Марчука Ю.Н. — М.: МГОУ, 2012. — С. 69—73.

18. Куликов, С.Ю. Определение автора высказывания при двойном цитировании // Проблемы языка: Сборник научных статей по материалам Второй конференции-школы «Проблемы языка: взгляд молодых ученых». — М.: Институт языкознания РАН, 2013. С. 209—215.

19. Куликов С.Ю. Об одном аспекте определения мнений на уровне объектов // Материалы Международного молодежного научного форума «ЛОМОНОСОВ-2013» / Отв. ред. А.И. Андреев, А.В. Андриянов, Е.А. Антипов, К.К. Андреев, М.В. Чистякова. [Электронный ресурс] — М.: МАКС Пресс, 2013. http://lomonosov-msu.ru/archive/Lomonosov_2013/2309/32293_8be6.doc

20. Куликов С.Ю. Способы выражения причин оценки в языке Интернета // Материалы Международного молодежного научного форума «ЛОМОНОСОВ-2014» / Отв. ред. А.И. Андреев, А.В. Андриянов, Е.А. Антипов. [Элек-

тронный ресурс] — М.: МАКС Пресс, 2014. http://lomonosov-msu.ru/archive/Lomonosov_2014/2713/2200_32293_2abfde.doc

21. Куликов, С.Ю. Распознавание именованных сущностей как первый этап прагматического анализа текста // Русский язык: исторические судьбы и современность: V Международный конгресс исследователей русского языка (Москва, МГУ имени М. В. Ломоносова, филологический факультет, 18-21 марта 2014 г.): Труды и материалы / Составители М. Л. Ремнёва, А. А. Поликарпов, О. В. Кукушкина. — М.: Изд-во Моск. ун-та, 2014. — С. 574.

22. Kulikov S. Opinion Lexicon Organisation in a rule-based Sentiment-analysis System // In Abstracts of Clin24, Leiden, the Netherlands. 2014. p. 78.

23. Kulikov S. Detecting Implicit Opinions with a Target-specific Thesaurus // In Abstracts of Clin25, Antwerpen, Belgium. 2015. p. 24.

24. Куликов С.Ю. Способы выражения причин оценки в языке Интернета и их автоматическая идентификация // Актуальные проблемы филологической науки: взгляд нового поколения: Материалы XX–XXI Международных конференций студентов, аспирантов и молодых ученых «Ломоносов»: Секция «Филология» / Ред.-сост. А. Е. Беликов. — М.: Издательство Московского университета, 2015. — Выпуск 6. — 627—630.