

ЯЗЫКОЗНАНИЕ

Елена Георгиевна БРУНОВА¹

Юлия Владимировна БИДУЛЯ²

Александр Александрович ГОРБУНОВ³

УДК 81'322

АСПЕКТНО-ОРИЕНТИРОВАННЫЙ АНАЛИЗ ТОНАЛЬНОСТИ ПОЛИТИЧЕСКОГО ДИСКУРСА

¹ доктор филологических наук, профессор
кафедры английского языка,
Тюменский государственный университет
egbrunova@mail.ru; ORCID: 0000-0002-8493-5932

² кандидат филологических наук, доцент
кафедры информационных систем,
Тюменский государственный университет
bidulya@yandex.ru; ORCID: 0000-0003-1878-3114

³ студент магистратуры,
Тюменский государственный университет
alexandregorbunov@gmail.com

Аннотация

Существующие системы по успешному и точному определению тональности текста основаны на статистических и математических принципах. Однако ничуть не менее интересными и перспективными являются работы, которые посвящены изучению лингвистических особенностей выражения оценки. Результаты данной формализации возможно применить как в области аффективных вычислений для дальнейшего совершенствования автоматических систем, так и для лингвистики и смежных с ней наук.

Цитирование: Брунова Е. Г. Аспектно-ориентированный анализ тональности политического дискурса / Е. Г. Брунова, Ю. В. Бидуля, А. А. Горбунов // Вестник Тюменского государственного университета. Гуманитарные исследования. Humanitates. 2021. Том 7. № 3 (27). С. 6-22.

DOI: 10.21684/2411-197X-2021-7-3-6-22

Новизна данного исследования заключается, прежде всего, в составлении алгоритма, основанного на выявленных лингвистических правилах. Кроме того, материалом исследования стал политический дискурс, который еще недостаточно изучен специалистами аффективных вычислений. Актуальность работы обоснована растущей необходимостью категоризации публикуемой в сети Интернет информации.

Цель исследования заключается в создании системы машинного определения тональности англоязычного политического дискурса, а также в выявлении аспектов сообщаемой информации для последующего применения в разработке. В статье рассматриваются лингвистические особенности анализа тональности и предлагается классификация лингвистических единиц с оценочным потенциалом применительно к структурным уровням языка. Также приводятся результаты эксперимента по тестированию работы системы автоматического определения тональности, проведенного на 300 документах новостей и комментариев, взятых с сайта reddit.com, раздел r/politics. Точность работы системы составила 92%. Кроме того, в ходе исследования была выполнена ручная разметка отобранных 40 комментариев, в ходе которой эксперт выявил 25 аспектов. Помимо этого, были выявлены 3 формальные закономерности в распределении аспектных терминов, что является необходимым для создания автоматической системы. Первая закономерность заключается в том, что аспектные термины повторяются в двух последовательно идущих предложениях. Вторая — аспектные термины часто являются темой предложения. Наконец, третья — выявлена высокая частота распределения аспектных терминов в начале и конце текста (документа).

Ключевые слова

Прикладная лингвистика, обработка естественного языка, анализ тональности текста, аспект, оценочный лексикон, классификатор на основе правил, политический дискурс.

DOI: 10.21684/2411-197X-2021-7-3-6-22

Введение

Ежедневно пользователями сети Интернет генерируется огромное количество неструктурированных текстов, содержащих мнения. Следовательно, существует необходимость в структуризации и формализации таких массивов данных для последующего отбора, анализа и интерпретации с целью автоматизации процессов, предсказания поведения людей при изменениях рынка или политической ситуации и т. д.

Исследования в области анализа тональности (англ. sentiment analysis, opinion mining, emotion AI) сосредоточены, прежде всего, на формализации мнений по поводу продаваемых товаров и оказываемых услуг. Однако по мере достижения определенных успехов и совершенствования технологий, а также в связи с развитием социальных сетей и форумов, предоставляющих неисчерпаемый источник материала для исследования, данную группу методов начали применять на различных типах дискурса.

Расширение сферы исследуемого контента подсказало и углубление самой методики исследования: от первоначальной классификации мнений на позитивные и негативные до выявления их аспектов; иными словами, от заключения за (пра-

вится) или против (не нравится) до поиска ответов на вопросы: почему нравится или не нравится, или что именно нравится, а что не нравится в объекте оценки.

Целью нашего исследования является разработка алгоритма и создание системы анализа тональности, направленной на выявление аспектов информации, сообщаемой в англоязычных политических текстах через высказываемые по этому поводу мнения.

Политический дискурс был выбран как более сложный для анализа и интерпретации материал, чем отзывы потребителей товаров и услуг, поскольку цели политических высказываний могут варьироваться от простого сообщения фактов до пропаганды, что определяет актуальность создания инструментов для автоматического определения аспектов того или иного текста политического содержания.

Результаты и обсуждение

Предыдущие исследования

Изначально анализ тональности был создан для бинарной классификации (позитивный/негативный) отзывов и обзоров о товарах и услугах и применялся главным образом к текстам на английском языке [8, 12, 22]. С течением времени, с развитием сети Интернет и социальных сетей появились разработки для других языков, и область применения анализа тональности существенно расширилась [21, 24]. Помимо этого, у исследователей возник интерес к извлечению эмоций и настроений в других типах дискурса, например в политическом [14, 15, 23]. Многообещающими являются работы в области анализа мультимодальных данных, включающих в себя, помимо текстовых, аудио- и видеоматериалы [25].

Кроме того, исследования в области анализа тональности критически необходимы для развития систем искусственного интеллекта, поскольку при решении возникающих проблем люди используют не только свои знания и опыт, но также эмоции и чувства [10, 20].

Благодаря успехам в области аффективных вычислений расширился также и инструментарий данного метода анализа текста: он стал применяться для определения уровня субъективности или объективности документа [27]. Помимо этого, современные системы по определению тональности не ограничиваются определением полярности, но также способны извлекать конкретные эмоции из текста, что стало возможным после составления подробных эмотивных лексиконов. Другим витком развития данного метода является его применение для определения объекта оценки, т. е. темы сообщения [17, 29]. Эта функция значительно упрощает процессы категоризации текстовых массивов данных.

В настоящее время представляется перспективным выявление аспектов объекта оценки, позволяющее провести более глубокий анализ мнения [9, 18, 28]. Знание отдельных аспектов объекта оценки может позволить скорректировать поведение, например, для совершенствования функционала товара или модификации политических решений. Стоит отметить, что аспекты не всегда можно извлечь поиском ключевых слов, поскольку они могут быть выражены имплицитно. Именно поэтому существует необходимость в формализации дистрибуции аспектных терминов.

Данная прикладная задача является междисциплинарной, находясь на стыке информационных технологий и лингвистики, а также других наук, в зависимости от исследуемого контента (психологии, маркетинга, политологии и т. д.). Однако среди авторов исследований в данной области редко встречаются лингвисты. Это представляется удивительным, поскольку материал для исследования — это тексты на естественных языках, имеющих свои законы и особенности, часть из которых универсальны и присущи всем языкам, или, по крайней мере, многим, а другие — специфичны и варьируются от языка к языку. Возможно, такую ситуацию можно объяснить тем, что некоторое время подобные исследования проводились исключительно на материале английского языка, и лингвистическая «непроработанность» стала заметна позже, когда попытки переноса готовых решений на другие языки не всегда оказывались легкими и успешными. Поэтому мы считаем целесообразным рассмотреть ряд лингвистических вопросов анализа тональности текста, причем сделать это на материале именно английского языка, давшего первый импульс в данной группе методов.

Лингвистические основы анализа тональности текста

Проблема определения тональности текста тесно связана с категорией оценки. Термин *оценка* традиционно используется в лингвистике при функционально-семантическом анализе языковых единиц — слов, высказываний и текста [3, с. 83; 4, с. 100; 5, с. 267]. Под оценкой понимается «непосредственная или опосредованная реакция говорящего (субъекта) на наблюдаемые, воображаемые, воспринимаемые органами чувств действия, признаки, признаки признаков реальных объектов, объектов внутреннего и внешнего мира говорящего» [5, с. 267].

В. А. Марьянчик предлагает дифференцировать *оценку* как реакцию субъекта и *оценочность* как способность лингвистических единиц «эксплицировать положительные или отрицательные свойства объекта» [4, с. 101]. Однако, по нашему мнению, рассматриваться должен потенциал не только слов и предложений, но и всех структурных уровней языка, где присутствует семантика: морфологический (морфема), лексический (лексема) и синтаксический (словосочетание, клаузма, предложение). В некотором смысле оценочный потенциал можно рассматривать также применительно к графическому уровню для письменной речи: цифры или количество символов (например, звезд для отелей или товаров), использование эмодзи для выражения эмоций, кавычек для выражения сарказма (в устной речи маркируется интонацией) или использование прописных букв (в устной речи маркируется повышением голоса) и т. д. Вовлечение графического уровня обусловлено поиском графических аналогов для выражения *невербальных* средств выражения оценки и эмоций, которые в анализируемом тексте должны учитываться как его непосредственные элементы, наряду с вербальными — словами, предложениями и т. д.

Соответственно, предлагаем к рассмотрению составленную авторами классификацию лингвистических единиц с оценочным потенциалом применительно к структурным уровням языка:

1. Графический уровень:

1.1. цифры (оценки успеваемости, товара или услуги);

- 1.2. количество символов (например, звезд для отелей или товаров);
- 1.3. кавычки для выражения сарказма, которые меняют полярность оценки на противоположную, например, “*comfortable*” *service*;
- 1.4. эмодзи для выражения эмоций;
- 1.5. прописные буквы как аналог крика.
2. Морфологический уровень:
 - 2.1. аффиксы с отрицательным значением, которые меняют полярность оценки на противоположную, например, *unhappy*, *disappointed*, *inadequate* и т. п.;
 - 2.2. аффиксы, выражающие отсутствие какого-то признака, которые меняют полярность оценки на противоположную, например, *hopeless* и т. п.;
 - 2.3. аффиксы и функциональные слова, выражающие степени сравнения прилагательных и наречий, например, *more optimistic*, *most comfortable*, *cleaner*, *the saddest*, *as ... as*, *so ... as* и т. п.
3. Лексический уровень:
 - 3.1. лексемы, семантика которых содержит положительную оценку или позитивную эмоцию (позитивный лексикон), например, *comfortable*, *clean*, *optimistic* и т. п.;
 - 3.2. лексемы, семантика которых содержит отрицательную оценку или негативную эмоцию (негативный лексикон), например, *sad*, *pessimistic*, *terrific* и т. п.;
 - 3.3. лексемы, выражающие положительную или отрицательную оценку в зависимости от предметной области и/или контекста (параметрический лексикон), например, *long waiting* (отрицательная оценка), но *long battery life* (положительная оценка);
 - 3.4. лексемы-усилители полярности, например, *very*, *absolutely* и т. п.;
 - 3.5. лексемы-ограничители полярности, например, *hardly*, *scarcely* и т. п.
4. Синтаксический уровень:
 - 4.1. функциональные слова, выражающие отрицание, которые меняют полярность оценки в предложении или клаузме на противоположную, например, *not*, *nobody*, *nothing*, *never* и т. п.;
 - 4.2. функциональные слова, выражающие отсутствие какого-либо признака, которые меняют полярность оценки на противоположную, например, *without*;
 - 4.3. функциональные слова, отменяющие смену полярности оценки на противоположную в отрицательных конструкциях, например, *so*, *such*, ср. *I was never disappointed* (слово из негативного лексикона, с отрицанием, оценка положительная), *I was never so disappointed* (слово из негативного лексикона с отрицанием, оценка отрицательная);
 - 4.4. восклицательные предложения как выражение эмоций;
 - 4.5. вопросительные высказывания как выражение сомнения, например, *Is it really comfortable?*;

4.6. союз *and*, соединяющий прилагательные с однонаправленной полярностью, например, *nice and comfortable*;

4.7. союз *but*, соединяющий прилагательные с противоположной полярностью, например, *nice, but expensive*.

Что касается текста (или документа — в терминах информационных технологий), то он выражает мнение субъекта в целом, реализованное с помощью оценочного потенциала упомянутых выше лингвистических единиц. В отличие от языковых уровней, описанных выше, текст не имеет столь четкой структуры. Более того, структура документа, выражающего оценку, будет существенно различаться в зависимости от типа дискурса, например, структура отзыва о товаре и структура научной рецензии обладают своими специфическими особенностями. Мы можем рассматривать только часто встречающиеся элементы структуры такого текста, например, заключение о рекомендации/нерекомендации объекта оценки или выражение «обманутого ожидания», когда в первой и большей по объему части документа перечисляются позитивные свойства объекта, а затем следует негативное заключение, сопровождаемое союзом *but, however, yet* и т. п.

Лингвистические средства выражения оценки и эмоций могут быть учтены в процессе формализации различным образом. Средства лексического уровня используются для формирования двух основных классов оценочного лексикона — позитивного и негативного, а также для формирования служебных классов, усиливающих или, наоборот, смягчающих оценку. Средства других уровней (графического, морфологического и синтаксического) могут использоваться для создания правил, выделяющих типичные модели, влияющие на оценку (смена полярности, усиление или ослабление оценки, выражаемой на лексическом уровне).

Некоторые морфологические и синтаксические средства будут полезны для автоматического расширения оценочного лексикона, например, исходя из слов одного класса лексикона (*comfortable*) можно автоматически добавить слова противоположного класса с отрицательными аффиксами (*uncomfortable*). Для пополнения лексикона можно использовать также принцип совместимости мнений (*sentiment consistency*) для прилагательных и наречий, впервые описанный в работе [16]. Если известна полярность одного прилагательного, то полярность другого прилагательного, которое соединено с ним союзом *and*, будет такой же, а в случае союза *but* — противоположной. Например:

This i-phone is nice and easy to use.

This i-phone is nice, but expensive.

Политический дискурс как объект оценки

Политический дискурс рассматривается как особая подсистема национального языка, предназначенная для политической коммуникации: для пропаганды тех или иных идей, эмотивного воздействия на граждан страны и побуждения их к политическим действиям [6, с. 12]. Для него характерны типичные особенности текста, речи, интеракции и семиотических практик, связанные с определенными группами или организациями, которые имеют отношение к власти [2, с. 192-193]. К функциям политического дискурса относятся социальный контроль и

манипуляция общественным сознанием, легитимизация и воспроизводство власти, ориентация через формулирование целей и проблем и формирование картины политической реальности в сознании социума, функция социальной солидарности, социальная дифференциация, агональная функция (иницирование и разрешение социального конфликта, выражение несогласия и протеста против действий властей), акциональная функция (проведение политики через мобилизацию, состоящую в активизации и организации сторонников, а также через процесс умиротворения и отвлечения внимания, усыпления бдительности), распространение информации, определение повестки дня, проекция в будущее и прошлое [7, с. 50].

Характер политических текстов позволяет выявлять оценочные суждения, высказываемые по отношению к объектам оценки через аспектные термины. Присутствующая в политическом обращении, статье на политическую тематику или комментариях к ним эмотивная составляющая позволяет выявлять в политическом дискурсе аспекты и аспектные термины. Оценочный лексикон в данном случае используется как главный маркер, обернутый в своеобразные шаблоны выражения оценки, в состав которых попадают аспектные термины. Таким образом, анализ тональности является важнейшим этапом выявления аспектов.

Методика определения тональности текста

В качестве источника исследуемого материала нами был выбран англоязычный сайт [reddit.com](https://www.reddit.com) с пользовательским контентом и комментариями, раздел Politics [26]. В данном разделе администрация и пользователи могут публиковать политические новости и заметки из проверенных и одобренных источников. На сайте предоставляется возможность для ведения дискуссий в комментариях, а также оценка как самой популярной информации, так и мнений пользователей по соответствующей теме. Отбор материала выполнялся с применением поисковой системы [google.com](https://www.google.com) с использованием опции поиска по разделу сайта: [site:reddit.com/r/politics](https://www.reddit.com/r/politics).

В целях решения задачи нами была разработана методика для определения тональности англоязычного текста, основанная на правилах (rule-based). Под правилами в данном случае понимаются некоторые условия и соответствующие им действия над единицами текста или служебными метасимволами, разметкой, тэгами. Результатом этих действий является расчет некоторой численной оценки, на основании которой делается вывод о тональности предложения либо всего текста.

Одним из важнейших инструментов применяемых правил является оценочный лексикон, составленный из эмоционально окрашенных слов — негативных и позитивных, представленных в виде лемм. В качестве оценочного лексикона использовалась библиотека NRCLExicon Марка Бэйли. Она содержит в себе 27 000 эмотивных слов и основана на аффективном лексиконе National Research Council Canada и наборах синонимов WordNet библиотеки NLTK [11, 13, 19]. В лексиконе представлены леммы таких частей речи, как имя существительное (*disagreement, accomplishment, rapture*), глагол (*bother, reassure, infect*), имя при-

лагательное (*gorgeous, culpable, apologetic*), наречие (*astonishingly, blindly, coldly*) и междометие (*amen, aye, damn, hurrah*).

Для получения численной оценки тональности текста мы воспользовались собственной методикой определения тональности, разработанной для русского языка [1], модифицировав ее с учетом синтаксических особенностей англоязычного текста. На вход поступает неформализованный текст новости или комментария и проходит предварительную обработку, включающую следующие операции:

1. Разбиение текста на предложения.
2. Разбиение сложных предложений на клауземы.
3. Разбиение предложений на слова, лемматизация.
4. Разметка всех слов и знаков препинания специальными тэгами в соответствии с типами, перечисленными в таблице 1.

Таблица 1

Тэги для типов слов и знаков препинания

Table 1

Tags for word types and punctuation marks

Тип	Примеры	Тэг
Оценочное слово: позитивная полярность	<i>Like, smart, fix, correct, help, easy</i>	POS
Оценочное слово: негативная полярность	<i>Opposite, impoverish, expensive, hurry, urgency</i>	NEG
Модификаторы полярности	<i>No, not, neither, never, none, nobody, nothing, nowhere</i>	ALT
	<i>Without</i>	WT
	? ?! (вопросительный знак, вопросительный знак с восклицательным знаком)	QM
	“ ” (кавычки)	Q
Усилители полярности	<i>Very, absolutely, completely, ever</i>	INC
	Слово, набранное прописными буквами	CAP
	! (восклицательный знак)	EM
Антимодификаторы полярности	<i>So, such</i>	TH
Прочие символы	Предложение	S
	Клаузема	F
	Союз <i>and / or</i>	CJ
	, ; – (запятая, двоеточие, точка с запятой, тире)	Z
	Слово, не входящее в лексикон	*

Далее все неразмеченные слова заменяются на символ *, а размеченные — на соответствующие тэги. Последовательности тэгов представляют не что иное как лексико-синтаксические шаблоны, к которым последовательно применяется набор правил для определения тональности предложения. Изначально словам оценочного лексикона присваивается вес (+1/-1) в зависимости от их позитивной или негативной тональности, затем в соответствии с правилами вычисляется оценка тональности всего предложения. Правила предписывают модификацию шаблонов с соответствующим пересчетом их весов.

Всего было разработано 15 правил модификации шаблонов. Приведем примеры некоторых правил вместе с их описанием и примерами применения.

Правило 1.1. $\langle \text{ALT} \rangle \langle \text{POS} \rangle \{n\} \rightarrow \langle \text{NEG} \rangle \{n\}$

Если в промежутке от начала предложения или знака препинания, или союза *and, or* до следующего знака препинания или союза *and, or* имеется модификатор полярности, то позитивная полярность всех слов, входящих в оценочный лексикон, в данном промежутке изменяется на негативную. Порядок указанных элементов (модификатор, слово с полярностью, любое другое слово) значения не имеет [1, с. 79-80].

'It's not good for our democracy': Calls grow for federal officers to shed military-style uniforms. Count = -1.

Правило 2.1. $\langle \text{INC} \rangle \langle \text{NEG} \rangle \rightarrow \langle \text{NEG} \rangle \langle \text{NEG} \rangle$

Если в промежутке от начала предложения или знака препинания до следующего знака препинания имеются слова, входящие в негативный лексикон, а также имеется усилитель, то каждый усилитель засчитывается за одно слово негативного лексикона. Порядок указанных элементов (усилитель, слово с полярностью, любое другое слово) значения не имеет [1, с. 80].

It was a very unprofessional memorandum. Count = -2.

В целях апробации методики был разработан программный инструмент для определения тональности с выводом разметки каждого предложения (рис. 1). Эксперимент, проведенный на 300 текстах новостей и комментариев, показал, что лишь в 24 случаях тональность, определенная системой, не совпадала с экспертной оценкой. При этом в 17 случаях негативный отзыв был ошибочно определен алгоритмом как позитивный, а в остальных случаях «пострадали» позитивные и гибридные отзывы. Вероятность ошибки составляет 8%, что соответствует 92% точности (Ассурасу) и является неплохим результатом.

Выявление аспектов оценки в политическом дискурсе

Термин *аспект* является общенаучным и междисциплинарным. В частности, в теоретической лингвистике его используют при рассмотрении видовременных форм глаголов. В компьютерной лингвистике использование данного термина носит узкоспециальный характер. При анализе тональности отзывов о товарах или об услугах, термин *аспект* обозначает конкретные характеристики объекта оценки, например, размер памяти смартфона, диагональ экрана, интерьер ресто-

рана или номера в отеле и т. д. Затем, после переноса методик данного анализа на другие виды дискурса, значение данного термина несколько расширилось.

В настоящем исследовании под аспектом мы понимаем те или иные составляющие объекта оценки, который упоминается в тексте новости, комментария или политического выступления. Рассмотрим фрагмент комментария к статье на политическую тему:

By not paying their fair share of taxes[tax], and buying corrupt[corr] politicians to not allow an increase in the minimum wage[wage], they're only allowing the rot and everything that's the opposite of what they desire America to be, is actually only making things worse. If they paid their taxes[tax], there'd be money for more things like infrastructure, if people had more income[wage], they'd spend more, which would mean less impoverished[pov] neighborhoods [26].

В тексте комментария рассматривается целевой аспект *billionaire*, к которому в тексте можно отнести следующие аспектные термины: *taxes*, *corruption*, *wage*, *poverty*, которые отмечены нами в данном примере с помощью тэгов в квадратных скобках.

Аспекты могут быть выражены явно и неявно. Рассмотрим следующие примеры:

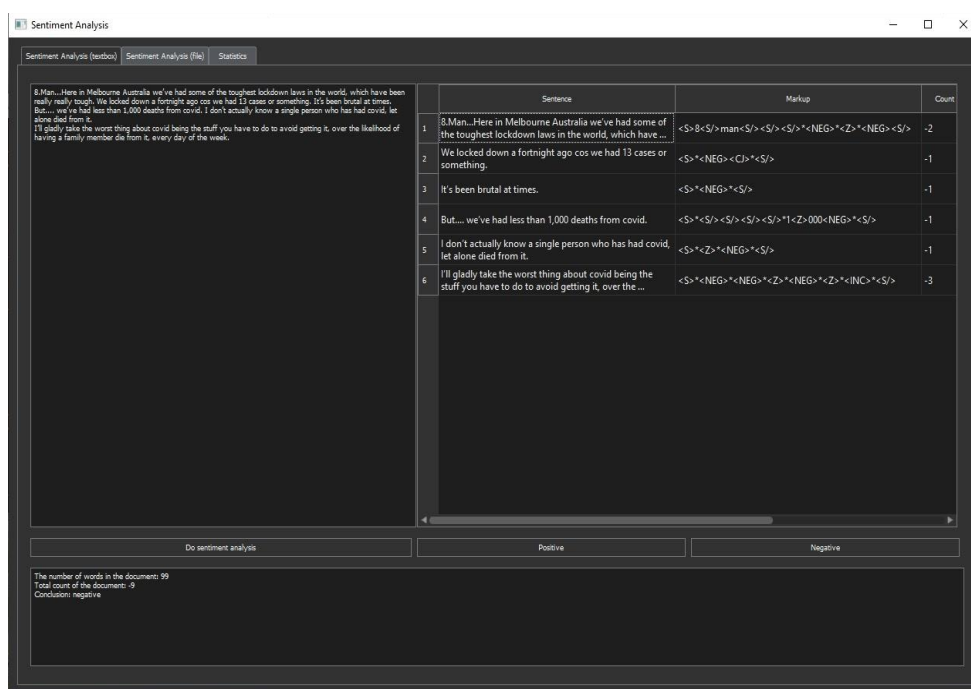


Рис. 1. Интерфейс программы автоматического определения тональности

Fig. 1. Sentiment Analysis Program Interface

- 1) *The protection our various government[gov] provide from inside and outside threats that they benefit from...* [26];
- 2) *While I would like nothing better than Biden to have free reign[gov] to fulfill his agenda[gov], packing the court[gov] will lead to instability and retaliation* [26].

В обоих примерах целевым аспектом является *government*, однако в первом случае аспектный термин явно указывает на него, в то время как во втором случае понимание об аспекте приходит через термины *reign, agenda, court*.

Поиск аспектов производился на описанном выше материале американской политической прессы. В сорока отобранных экспертом комментариях было выявлено 25 аспектов: *capitalism, conservatives, corruption, COVID pandemic, democrats, economy, education, employment, geography, government, healthcare, infrastructure, laws, media, military, mortality, politician, poverty, presidency, republicans, space exploration, taxes, voting, wage, wealth*.

Приведем пример комментария с разметкой и определением аспектов:

Look, I know he's not a perfect [**<ALT><POS> -> <NEG> -> -1**] man or a perfect [**<POS> -> +1**] leader[**pres**]. But even just seeing his face brings me a sense of peace [**<POS> -> +1**] and hope [**<POS> -> +1**] I haven't felt in over 4 years. He is doing a wonderful [**<POS> -> +1**] job. He is listening to all of the experts and letting them do their jobs. Hell, he's at least acknowledging [**<POS> -> +1**] this f* of a virus[**cov**] exists and poses a great threat [**<NEG> -> -1**] to our country and all of humanity. I am so glad [**<POS> -> +1**] he's in charge[**pres**] [26].

Count = +4 (positive)

[presidency; COVID pandemic]

В ходе анализа комментариев с выявлением аспектов были выявлены следующие закономерности:

1. Аспектные термины одной категории часто встречаются в двух стоящих рядом предложениях, например: *The "order" that government[gov] creates that enables them to even bring a company successfully to fruition. The protection our various government[gov] provide from inside and outside threats that they benefit from* [26].
2. Аспектные термины часто выступают в роли темы предложения (актуальное членение): *Government[gov] is the framework that should enable everyone to reach that minimum standard and if that fails, I can't be mad at a man for stealing bread because he's got to eat too. The benefits of space exploration[space] far exceeds the cost in the long term. Joe Biden's presidency[pres] so far has been pretty stellar* [26].
3. Высокая частота встречаемости аспектных терминов обнаружена в начале (в 70% текстов) и конце (60%) текста (документа).

Перечисленные наблюдения открывают широкие возможности в применении методов интеллектуального анализа для выявления аспектов путем обучения классификаторов на размеченных текстах. После обработки текстов мы получили следующие признаки токенов:

1. Тип слова, определяемый на базе лексикона, описанного в таблице 1.
2. Тип шаблона, в рамках которого находится слово.
3. Порядковый номер слова в предложении, отнесенный к длине предложения.
4. Принадлежность слова аспектному термину.
5. Наличие аспектных терминов в предшествующем или последующем предложениях.

Выводы

Поиск аспектов в политическом дискурсе представляет интересную и практически значимую проблему. В данном исследовании мы разработали основанную на правилах методику выявления аспектов через их актуализацию в аспектных терминах, которые, в свою очередь, выявляются в выражении тональности и оценочных суждений в тексте политического содержания. Практическая реализация методики воплощена в программном инструменте, позволяющем определить тональность предложений, а также выделить лексико-синтаксические шаблоны, составляющие основу аспектных терминов в задаче аспектно-ориентированного анализа текста. Точность определения тональности в политических текстах при помощи этого инструмента составляет 92%.

Представленные в правилах лексико-синтаксические шаблоны являются также элементами признакового описания дискурса в модели машинного обучения наряду с семантическими и синтаксическими характеристиками текста, такими как принадлежность определенному лексикону, положение в предложении, порядок предложений.

СПИСОК ЛИТЕРАТУРЫ

1. Брунова Е. Г. Клиент всегда прав: анализ тональности текста в отзывах о качестве банковского обслуживания / Е. Г. Брунова, Ю. В. Бидуля // Вестник Тюменского государственного университета. Гуманитарные исследования. *Humanitates*. 2017. Том 3. № 1. С. 72-89.
2. Ван Дейк Т. А. Дискурс и власть: репрезентация доминирования в языке и коммуникации / Т. А. Ван Дейк; пер. с англ. М.: Книжный дом «ЛИБРОКОМ», 2013. 344 с.
3. Комиссарова Н. Г. Оценочность в дискурсе СМИ / Н. Г. Комиссарова, Д. Р. Якупова // Филологические науки. Вопросы теории и практики. 2018. № 2-1 (80). С. 83-87.
4. Марьянчик В. А. Оценка как категория текста / В. А. Марьянчик // Вестник Северного (Арктического) федерального университета. Серия: Гуманитарные и социальные науки. 2011. № 1. С. 100-103.
5. Папина А. Ф. Текст: его единицы и глобальные категории / А. Ф. Папина. М.: УРСС, 2002. 367 с.
6. Чудинов А. П. Россия в метафорическом зеркале: когнитивное исследование политической метафоры (1991-2000): монография / А. П. Чудинов. Екатеринбург: Урал. гос. пед. ун-т, 2001. 238 с.

7. Шейгал Е. И. Семиотика политического дискурса: дис. ... д-ра филол. наук / Е. И. Шейгал. М.: РГБ, 2005. 431 с.
8. Bhatt A. et al. Amazon review classification and sentiment analysis / A. Bhatt et al. // International Journal of Computer Science and Information Technologies. 2015. Vol. 6. No. 6. Pp. 5107-5110.
9. Blinov P. D. Semantic similarity for aspect-based sentiment analysis / P. D. Blinov, E. V. Kotelnikov // Computational Linguistics and Intellectual Technologies: Proceedings of the Annual International Conference "Dialogue" (Moscow, 27-30 May 2015). Moscow: RSUH, 2015. Iss. 14 / V. P. Selegey (ed.). Pp. 12-22.
10. Cambria E. Affective computing and sentiment analysis / E. Cambria // IEEE Intelligent Systems. 2016. Vol. 31. No. 2. Pp. 102-107.
11. Emotion classification using NRC Lexicon in Python, 2020.
URL: <https://www.geeksforgeeks.org/emotion-classification-using-nrc-lexicon-in-python/>
(дата обращения: 23.05.2021).
12. Fang X. Sentiment analysis using product review data / X. Fang, J. Zhan // Journal of Big Data. 2015. Vol. 2. No. 1. Pp. 1-14.
13. Fellbaum C. WordNet: An Electronic Lexical Database / C. Fellbaum. Cambridge, MA: MIT Press, 1998. 499 pp.
14. Gold D. et al. Corpus of aspect-based sentiment in political debates / D. Gold et al. // Proceedings of the 14th Conference on Natural Language Processing (KONVENS 2018). Vienna, 2018. Pp. 89-99.
15. Haselmayer M. Sentiment analysis of political communication: combining a dictionary approach with crowdcoding / M. Haselmayer, M. Jenny // Quality & Quantity. 2017. No. 51. Pp. 2623-2646.
16. Hatzivassiloglou V. Predicting the semantic orientation of adjectives / V. Hatzivassiloglou, K. McKeown // Proceedings of the 35th Annual Meeting of ACL. Madrid, 1997. Pp. 174-181.
17. Lin C. Joint sentiment/topic model for sentiment analysis / C. Lin, Y. He // Proceedings of the 18th ACM Conference on Information and Knowledge Management, 2009. Pp. 375-384.
18. Mayorov V. A high precision method for aspect extraction in Russian / V. Mayorov, I. Andrianov, N. Astrakhantsev, V. Avanesov, I. Kozlov, D. Turdakov // Computational Linguistics and Intellectual Technologies: Proceedings of the Annual International Conference "Dialogue" (Moscow, 27-30 May 2015). Moscow: RSUH, 2015. Iss. 14 / V. P. Selegey (ed.). Pp. 34-44.
19. Miller G. A. WordNet: A Lexical Database for English / G. A. Miller. 1995.
URL: <https://wordnet.princeton.edu/>
20. Minsky M. The emotion machine: commonsense thinking / M. Minsky // Artificial Intelligence, and the Future of the Human Mind. Simon & Schuster, 2006. 400 pp.
21. Mittal A. Stock prediction using twitter sentiment analysis / A. Mittal, A. Goel // CS229: Machine Learning. Stanford University Course. 2012. Vol. 15.
URL: <http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf> (дата обращения: 23.05.2021).
22. Mukherjee S. Feature specific sentiment analysis for product reviews / S. Mukherjee, P. Bhattacharyya // International Conference on Intelligent Text Processing and Computational Linguistics. Berlin; Heidelberg: Springer, 2012. Pp. 475-487.
23. Mullen T. A preliminary investigation into sentiment analysis of informal political discourse / T. Mullen, R. Malouf // AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs. 2006. Pp. 159-162.

24. Neidhardt J. Predicting happiness: user interactions and sentiment analysis in an online travel forum / J. Neidhardt, N. Rümmele, H. Werthner // *Information Technology & Tourism*. 2017. Vol. 17. No. 1. Pp. 101-119.
25. Poria S. Fusing audio, visual and textual clues for sentiment analysis from multimodal content / S. Poria, E. Cambria, N. Howard, G.-B. Huang, A. Hussain // *Neurocomputing*. 2016. No. 174. Pp. 50-59.
26. Reddit.com: website. URL: www.reddit.com (дата обращения: 23.05.2021).
27. Spencer J. Sentimentor: sentiment analysis of Twitter data / J. Spencer, G. Uchytig // *SDAD@ ECML/PKDD*. 2012. Pp. 56-66.
28. Vasilyev V. G. Aspect extraction and Twitter sentiment classification by fragment rules / V. G. Vasilyev, A. A. Denisenko, D. A. Solovyev // *Computational Linguistics and Intellectual Technologies: Proceedings of the Annual International Conference "Dialogue" (Moscow, 27-30 May 2015)*. Moscow: RSUH, 2015. Iss. 14 / V. P. Selegey (ed.). Pp. 76-88.
29. Wang X. et al. Topic sentiment analysis in Twitter: a graph-based hashtag sentiment classification approach / X. Wang et al. // *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*. 2011. Pp. 1031-1040.

Elena G. BRUNOVA¹

Yulia V. BIDULYA²

Alexander A. GORBUNOV³

UDC 81'322

ASPECT-BASED SENTIMENT ANALYSIS OF POLITICAL DISCOURSE

¹ Dr. Sci. (Philol.), Professor,
Department of English Language, University of Tyumen
egbrunova@mail.ru; ORCID: 0000-0002-8493-5932

² Cand. Sci. (Philol.), Associate Professor,
Department of Information Systems, University of Tyumen
bidulya@yandex.ru; ORCID: 0000-0003-1878-3114

³ Master's degree student, University of Tyumen
alexandregorbunov@gmail.com

Abstract

The existing systems for accurate sentiment analysis are mainly based on statistical and mathematical principles. However, more promising are the works that are devoted to the study of the linguistic features of the evaluation expression. The results of this formalization can be applied both in the field of affective computing for further improvement of automatic systems and for linguistics and related sciences.

The novelty of this study lies mainly in the development of an algorithm based on the identified linguistic rules. In addition, the research material is political discourse, which has not yet been studied enough by specialists of affective computing. The relevance of this work is justified by the growing need for categorization of information published on the Internet. The purpose of the study is to develop a system for machine sentiment analysis of English-language political texts, as well as to identify aspects and their distribution for subsequent use in enhancement. The article discusses the linguistic features of sentiment analysis and

Citation: Brunova E. G., Bidulya Yu. V., Gorbunov A. A. 2021. "Aspect-Based Sentiment Analysis of Political Discourse". Tyumen State University Herald. Humanities Research. Humanitates, vol. 7, no. 3 (27), pp. 6-22.

DOI: 10.21684/2411-197X-2021-7-3-6-22

suggests a classification of linguistic units with sentiment potential in relation to levels of language structure. The results of an experiment on testing the operation of the sentiment analysis system, conducted on 300 news articles and user comments taken from [reddit.com/r/politics](https://www.reddit.com/r/politics/), are also presented. The accuracy of the system is 92%. In addition, the selected 40 comments were manually marked up and tagged; during this process the expert identified 25 aspects. Furthermore, 3 formal patterns were identified in the distribution of aspect terms, which is necessary for creating an automatic system. The first peculiarity is that the aspect terms are repeated in two consecutive sentences. The second is that aspect terms are often the themes of sentences. Finally, the third — a high frequency of distribution of aspect terms at the beginning and end of the text (document) was revealed.

Keywords

Applied linguistics, natural language processing, sentiment analysis, aspect, sentiment lexicon, rule-based classifier, political discourse.

DOI: 10.21684/2411-197X-2021-7-3-6-22

REFERENCES

1. Brunova E. G., Bidulya Yu. V. 2017. "The customer is always right: sentiment analysis for bank service quality". Tyumen State University Herald. Humanities Research. Humanitates, vol. 3, no. 1, pp. 72-89. [In Russian]
2. Van Dijk T. A. 2013. *Discourse and Power: Representation of Domination in Language and Communication*. Moscow: LIBROKOM. 344 p. [In Russian]
3. Komissarova N. G., Jakupova D. R. 2018. "Evaluation in the media discourse". *Philological Sciences. Questions of Theory and Practice*, no. 2-1 (80), pp. 83-87. [In Russian]
4. Maryanchik V. A. 2011. "Evaluativity as a text category". *Northern (Arctic) Federal University Herald. Series: Humanities and Social Sciences*, no. 1, pp. 100-103. [In Russian]
5. Papina A. F. 2002. *Text: Its Units and Global Categories*. Moscow: URSS. 367 p. [In Russian]
6. Chudinov A. P. 2001. *Russia in a Metaphorical Mirror: Cognitive Research of Political Metaphor (1991-2000): monography*. 238 p. [In Russian]
7. Shejgal E. I. 2005. *Semiotics of Political Discourse*. Moscow: RGB. 431 p. [In Russian]
8. Bhatt A. et al. 2015. "Amazon review classification and sentiment analysis". *International Journal of Computer Science and Information Technologies*, vol. 6, no. 6, pp. 5107-5110.
9. Blinov P. D., Kotelnikov E. V. 2015. *Semantic Similarity for Aspect-Based Sentiment Analysis*. *Computational Linguistics and Intellectual Technologies: Proceedings of the Annual International Conference "Dialogue"*. Edited by V. P. Selegey. Moscow: RSUH. Iss. 14, pp. 12-22.
10. Cambria E. 2016. "Affective computing and sentiment analysis". *IEEE Intelligent Systems*, vol. 31, no. 2, pp. 102-107.
11. Emotion classification using NRC Lexicon in Python. 2020. Accessed on 23 May 2021. <https://www.geeksforgeeks.org/emotion-classification-using-nrc-lexicon-in-python/>
12. Fang X., Zhan J. 2015. "Sentiment analysis using product review data". *Journal of Big Data*, vol. 2, no. 1, pp. 1-14.

13. Fellbaum C. 1998. WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press. 499 p.
14. Gold D. et al. 2018. "Corpus of aspect-based sentiment in political debates". Proceedings of the 14th Conference on Natural Language Processing (KONVENS 2018). Vienne. Pp. 89-99.
15. Haselmayer M., Jenny M. 2017. "Sentiment analysis of political communication: combining a dictionary approach with crowdcoding". *Quality and Quantity*, no. 51, pp. 2623-2646.
16. Hatzivassiloglou V., McKeown K. 1997. "Predicting the semantic orientation of adjectives". Proceedings of the 35th Annual Meeting of ACL. Madrid. Pp. 174-181.
17. Lin C., He Y. 2009. "Joint sentiment/topic model for sentiment analysis". Proceedings of the 18th ACM Conference on Information and Knowledge Management. Pp. 375-384.
18. Mayorov V., Andrianov I., Astrakhantsev N., Avanesov V., Kozlov I., Turdakov D. (ed.). 2015. "A high precision method for aspect extraction in Russian". *Computational Linguistics and Intellectual Technologies: Proceedings of the Annual International Conference "Dialogue"*. Edited by V. P. Selegey. Moscow: RSUH. Iss. 14, pp. 34-44.
19. Miller G. A. 1995. WordNet: A Lexical Database for English. <https://wordnet.princeton.edu/>
20. Minsky M. 2006. "The emotion machine: commonsense thinking". *Artificial Intelligence, and the Future of the Human Mind*. Simon and Schuster. 400 p.
21. Mittal A., Goel A. 2012. "Stock prediction using twitter sentiment analysis". CS229: Machine Learning. Stanford University Course, vol. 15. Accessed on 23 May 2021. <http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf>.
22. Mukherjee S., Bhattacharyya P. 2012. "Feature specific sentiment analysis for product reviews". *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer, Berlin, Heidelberg. Pp. 475-487.
23. Mullen T., Malouf R. 2006. "A Preliminary Investigation into Sentiment Analysis of Informal Political Discourse". *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*. Pp. 159-162.
24. Neidhardt J., Rümmele N., Werthner H. 2017. "Predicting happiness: user interactions and sentiment analysis in an online travel forum". *Information Technology and Tourism*, vol. 17, no. 1, pp. 101-119.
25. Poria S., Cambria E., Howard N., Huang G.-B., Hussain A. 2016. "Fusing audio, visual and textual clues for sentiment analysis from multimodal content". *Neurocomputing*, no. 174, pp. 50-59.
26. Reddit.com. Accessed on 23 May 2021. www.reddit.com
27. Spencer J., Uchyigit G. 2012. "Sentimentor: sentiment analysis of twitter data". *SDAD@ ECML/PKDD*. Pp. 56-66.
28. Vasilyev V. G., Denisenko A. A., Solovyev D. A. 2015. Aspect Extraction and Twitter Sentiment Classification by Fragment Rules. *Computational Linguistics and Intellectual Technologies: Proceedings of the Annual International Conference "Dialogue"*. Edited by V. P. Selegey. Moscow: RSUH. Iss. 14, pp. 76-88.
29. Wang X. et al. 2011. "Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach". Proceedings of the 20th ACM International Conference on Information and Knowledge Management. Pp. 1031-1040.