

УДК 81-114.2

DOI: 10.18384/2310-712X-2018-1-38-45

## **ДИХОТОМИЯ СУБЪЕКТИВНОСТЬ VS. ОБЪЕКТИВНОСТЬ И ТОНАЛЬНАЯ РЕЛЕВАНТНОСТЬ В ЗАДАЧАХ АНАЛИЗА ТОНАЛЬНОСТИ**

**Семина Т.А.**

*Московский государственный областной университет  
105005, г. Москва, ул. Радио, д. 10А, Российская Федерация*

**Аннотация.** Статья раскрывает содержание понятий «субъективность», «объективность» и «тональная релевантность» и сферу их применения в системах извлечения мнений. Автор прослеживает становление термина *тональная релевантность* и полисемическое использование этого термина в рамках данной научной проблемы. Автором проанализирован корпус рецензий на фильмы, в котором предложения размечены как релевантные или нерелевантные, и корпус новостных статей с размеченными отношениями между сущностями. Кроме того, был проведён эксперимент по автоматическому извлечению релевантных пар сущностей и полярности их оценки. Проведённый анализ показал целесообразность произошедшего перехода к тональной релевантности от дихотомии *СУБЪЕКТИВНОСТЬ VS. ОБЪЕКТИВНОСТЬ*.

**Ключевые слова:** анализ тональности, объективность, политический текст, субъективность, тональная релевантность.

## **SUBJECTIVITY VS. OBJECTIVITY DICHOTOMY AND SENTIMENT RELEVANCE IN SENTIMENT ANALYSIS TASKS**

**T. Semina**

*Moscow Region State University  
10A Radio ulitsa, Moscow 105005, Russian Federation*

**Abstract.** The paper describes the meaning of terms “subjectivity”, “objectivity” and “sentiment relevance” and the scope of their application in opinion mining systems. The author traces the formation of the term “*sentiment relevance*” and the polysemantic usage within the given scientific framework. The author analyzed a movie review corpora with sentences marked as relevant or non-relevant and news articles corpora with marked relations between entities. Moreover, an experiment on automatic extraction of relevant pairs of entities and the polarity of the relations was conducted. The analysis supported change from subjectivity vs. objectivity dichotomy to sentiment relevance.

**Key words:** sentiment analysis, objectivity, political text, subjectivity, sentiment relevance.

Анализ тональности, известный как сентимент-анализ или система извлечения мнений – это область изучения мнений, оценок, отношения и эмоций людей по отношению к таким объектам, как продукты, организации, личности, события, проблемы, и их атрибутам.

Сентимент (от англ. *sentiment* – чувство, мнение, настроение) – эмоциональная оценка, выраженная в тексте, также называемая тональностью текста [4].

При выявлении мнений мы говорим о субъекте тональности (далее – субъект), объекте тональности (далее – объект) и тональности высказывания.

Субъект – сущность, которой принадлежит мнение.

Объект – сущность, по отношению к которой высказывается мнение.

Тональность – положительная, отрицательная или нейтральная (возможно исключение нейтрального класса или добавление градации полярности мнения) оценка объекта субъектом.

Анализ тональности проводится на трёх уровнях:

- Уровень документа.

В данном случае мы исходим из допущения, что документ отражает мнение относительно одного объекта, и считаем общую тональность всего текста.

- Уровень предложения.

На этом уровне анализа классифицируется не текст целиком, а отдельные предложения. Данный уровень анализа связан с задачей определения субъективности.

- Аспектуальный уровень.

Задача анализа – понять не общую тональность текста или предложения или отношение автора к объекту, а что

именно нравится или не нравится автору.

Развитие сети Интернет привело к тому, что тексты, которые там находятся, стали самым распространённым материалом для анализа. Это связано с количеством текстов, их многообразием и тем, что они находятся в свободном доступе. С исследовательской точки зрения тексты из Интернета можно разделить на две группы: персонализированные и неперсонализированные тексты.

Персонализированные тексты – блоги, микроблоги, рецензии и отзывы. Тексты такого рода часто носят субъективный характер и имеют эксплицитно выраженное мнение, потому долгое время в задачах анализа тональности рассматривали только эти данные.

Субъектом в большинстве случаев является автор, что облегчает задачу поиска сущностей; при анализе текстов *Twitter* или рецензий, при анализе всего документа возможно присвоение статуса объекта по хештегу или названию (если мы говорим о рецензии или отзыве). В таком случае задача установления тональности состоит в поиске тональных высказываний – единиц, которые непосредственно выражают мнение [7]. Отметим, что ими не всегда будут лексические единицы. Это могут быть эмодзи, пунктуация, длина предложений или всего документа. При работе с короткими текстами хорошо себя показывают методы машинного обучения, что позволяет использовать статистическую информацию о вероятности отнесения документа к одному из классов (положительный, отрицательный или нейтральный) на основе указанных нами выше признаков. Возможно и применение иерар-

хии признаков для более качественной классификации.

Важнейшим понятием, связанным с сентиментом, является субъективность. Классификация по субъективности выявляет объективные и субъективные предложения.

Объективное предложение содержит фактическую информацию, в то время как субъективное выражает личные чувства, точку зрения и убеждения. Важно понять, что субъективность не тождественна сентименту, так как и в объективных предложениях может содержаться мнение, например, «В прошлом месяце мы купили автомобиль, и дворники на лобовом стекле уже не работают», и субъективные предложения не всегда содержат тональность, например, «Мне кажется, сейчас пойдёт дождь».

Долгое время в сфере анализа тональности широко применялось деление языковых единиц на субъективные и объективные. С 2013 г. высказано и экспериментально доказано предположение, что данные категории не эффективны при анализе тональности, вследствие чего возник термин «тональная релевантность».

Тональная релевантность (*Sentiment Relevance*) – понятие, позволяющее отличать информативное содержание для определения тональности документа от неинформативного. Оно контрастирует с обычным различием между субъективным и объективным содержанием [9]. Хотя между приведёнными двумя понятиями существует связь, они не эквивалентны.

Рассмотрим примеры:

(1) *Обвинённый в страшном преступлении, Джон Коффи оказывается в блоке смертников тюрьмы «Холодная гора».*

(2) *Фильм был номинирован на Оскар.*

Предложение (1) является субъективным, поскольку оценки, подобные страшному преступлению, субъективны для зрителя. Предложение (2) является объективным, так как возможно проверить истинность утверждения. Однако, несмотря на то, что предложение (1) имеет отрицательное субъективное содержание, оно не имеет отношения к тональности текста, поскольку оно связано с сюжетом фильма и появляется в положительных обзорах. И наоборот, предложение (2) указывает на положительное мнение, выраженное автором. Подобное деление языковых единиц можно представить при анализе рецензий на книги. В случае с Твиттером кажется не очевидным, что принимать за релевантную единицу, а что – нет, либо стоит применять понятие релевантности при фильтрации спама, который появляется в выборке при автоматической загрузке, либо стоит релевантным признавать твит, содержащий интересующий исследователя объект.

Как было сказано нами выше, субъективность и тональная релевантность – это две различные концепции, которые не подразумевают друг друга: в целом нейтральные и объективные предложения могут быть релевантными, в то время как субъективный контент – нерелевантным. Scheible C. и Schutze H. [9] проводили классификацию релевантности настроений в сфере кино. Для создания аннотированного корпуса, или SR-корпуса, они случайным образом отобрали 125 текстов из базы обзоров фильмов [8]. Всего в корпусе SR содержится 275 релевантных и 728 нерелевантных предложений, корпус находится в свободном доступе.

Рассмотрим фрагмент размеченного корпуса [9]:

“<SR> One of the last entries in the long-running Carry on series , Carry on Behind is very similar to Carry on Camping in that it involves a group of holiday-makers descending on a ‘caravan’ site. <SNR> Professors Anna Vrooshka ( Elke Sommer ) and Roland Crump ( Kenneth Williams ) and a group of archaeology students stay in the caravan site owned by Major Leep ( Kenneth Connor ) so that they can explore the nearby Roman settlement remains. <SNR> Anna has a little trouble understanding English and sometimes people get the wrong end of the stick: – for instance, when Anna is asking for ‘scrubbers for dirty caravan’ , she means that she wants a scrubbing brush to clean the caravan ! <SNR> Arthur Upmore ( Bernard Bresslaw ) and his wife Linda ( Patsy Rowlands ) take her mother Daphne Barnes ( Joan Sims ) and her minah bird on holiday with them. <SNR> Fred Ramsden ( Windsor Davies ) and Ernie Bragg ( Jack Douglas ) leave their wives ( Liz Fraser and Patricia Franklin ) behind for a fishing holiday. <SR> The story involves the disruption caused by the archaeological professors of the day-to-day running of the camp , the search for the minah bid and greyhound , the Major , Fred and Ernie ‘s desperate need of a woman , a misunderstanding leading to a striptease at the caravan park ‘s pub , and a shock in store for Daphne. <SR> The partnership between Sommer and Williams is very effective and amusing. <SR> This is what binds the movie together”

Как мы видим, к категории тонально нерелевантных единиц (<SNR>) были отнесены все предложения, описывающие сюжет фильма, к тонально релевантным (<SR>) – остальные. Раз-

метчики применяли и третью категорию – <uncertain> – в тех случаях, когда невозможно было однозначно приписать метку <SR> или <SNR>.

“<uncertain> Films adapted from comic books have had plenty of success , whether they ‘re about superheroes ( Batman , Superman , Spawn ) , or geared toward kids ( Casper ) or the arthouse crowd ( Ghost World ) , but there’s never really been a comic book like From Hell before.

<uncertain> For starters , it was created by Alan Moore ( and Eddie Campbell ) , who brought the medium to a whole new level in the mid ‘80s with a 12-part series called The Watchmen”

Несомненно, при анализе текстов рецензий встречаются ошибки, совершённые при разметке, но, так как первоначальная цель корпуса – быть материалом для задач обучения на прецедентах, это не критично, так как большой объём коллекции не позволит неточностям испортить результаты выдачи.

У термина «тональная релеванность» есть ещё одно значение. Тонально релевантными могут признаваться языковые единицы (синтагмы), кроме того, рассматривается релевантность объектов по отношению к тональным высказываниям.

В последнее время исследовательский интерес перешёл от персонализированных текстов (твиты, рецензии, отзывы) к неперсонализированным (новостные статьи, аналитические статьи). Следствием этого стало принятие во внимание не только оценочных высказываний (*opinion* в словаре тональных высказываний РуСентиЛекс), но и фактуальной информации, т. е. той, которая при ранней дихотомии «субь-

ективность vs. объективность», скорее, считалась бы объективной.

Существует корпус размеченных коротких новостных сообщений для английского языка [6], подобная же разметка сделана для корейского корпуса [10]. При анализе подобных текстов, в отличие от персонализированных, ищут не отношение автора к объекту или объектам, а отношения между сущностями. Рассмотрим пример (3), который мы взяли из корпуса MPQA:

(3) *Russia's Defence Ministry has denied Taleban assertions that Russian officers are serving in the Northern Alliance. "Formations of the anti-Taleban coalition do not include officers of the Russian armed forces," a high-ranking military official told Interfax on Friday. "Russia has provided only military-technical assistance to the Northern Alliance and continues to do so. Russian officers were not dispatched to Afghanistan either as unit leaders or as instructors," the official added.*

Извлечение мнений из таких текстов важно для многих прикладных задач. Благодаря этому можно проще создавать краткие новостные сводки, создавать системы мониторинга политической среды. Новостные тексты отличаются малым объёмом и сравнительно небольшим количеством упомянутых сущностей, что в некоторой степени упрощает анализ.

При автоматическом анализе такого текста мы хотим выявить отношения между сущностями, которые встретились в одном тексте. Мнение автора опускается, так как стилистика новостных сообщений не подразумевает субъективности журналиста по отношению к теме статьи. Примечательно то, что отношения между сущностями находятся не по оценоч-

ной лексике, а по фактуальной, метки *opinion* и *fact* в словаре оценочной лексики РуСентиЛекс [2].

Аналитические статьи являются более сложным материалом для исследования, так как содержат большее количество сущностей (т. е. потенциальных субъектов и объектов тональности).

В неперсонализированных текстах установление тональной релевантности сводится к тому, что подобранный лингвистом алгоритм должен находить тональные высказывания и подбирать к ним сущность-субъект и сущность-объект. Возможно и обратное направление анализа: поиск пар сущностей, потенциально являющихся элементами мнения (субъектом или объектом) и дальнейшая оценка тональности их отношения. Тональные высказывания могут находиться как при помощи методов машинного обучения, так и при помощи специальных словарей оценочной (тональной) лексики.

В качестве примера рассмотрим текст статьи с сайта *inosmi.ru*. Этот сайт размещает переводы статей из зарубежных изданий на русский язык. Конечно, нельзя считать перевод абсолютно эквивалентным с прагматической точки зрения [3], но качество тестирования и работы подобной системы анализа от этого не ухудшится.

Именованные сущности извлекались автоматически при помощи программы [1].

(4) *По словам Пескова, выступившего перед репортёрами в Москве в пятницу, 27 октября, президент России Владимир Путин, который на прошлой неделе обвинил США в том, что «они возвращают нас в 1950-е годы», пристально следит за новостями о*

санкциях. **Путин** не захотел выступить с критикой в адрес лично **Трампа**, а **Песков**, сославшись на слова президента, отметил, что **Москва** продолжит стремиться к налаживанию двусторонних отношений настолько, «насколько к этому готовы и этого желают наши американские коллеги».

Мы видим ряд именованных сущностей и ряд тональных высказываний (т. е. языковых единиц, эксплицитно выражающих тональность). В ходе нашего эксперимента мы ставили перед собой задачу установления отношений между такими сущностями, иными словами, подбирали релевантные сущности для каждого тонального высказывания. Задача усложнилась тем, что в рамках одного абзаца могут быть как положительные, так и отрицательные отношения между сущностями, в том числе и между одной парой сущностей. Если представить текст статьи в виде списка мнений, который в общем виде будет выглядеть [*<Сущность А>*, *<Сущность В>*, тональность], где А – субъект тональности, В – объект тональности, тональность – полярность оценки (положительная или отрицательная), мы получим:

[Путин, США, neg], т. к. «обвинил США в том, что “они возвращают нас в 1950-е гг.”»;

[Путин, Трамп, pos], «**Путин** не захотел выступить с критикой в адрес лично **Трампа**»;

[Москва, США, pos], «**Москва** продолжит стремиться к налаживанию двусторонних отношений».

Понятие релевантности, как было показано выше, в современном анализе тональности изменило своё значение, от релевантности предложений как критерия отбора для обучения и тестирования исследователи перешли

к более узкой сфере – релевантности сущности для конкретного высказывания [5]. Иногда в работах употребление термина в этом значении называют тональной релевантностью сущностей (*entities sentiment relevance*), что кажется нам удачным термином. Терминологическое различие релевантности фрагментов текста и сущностей важно, так как, по сути, это разные подзадачи анализа тональности, каждая из которых требует применения особых алгоритмов анализа и создания обучающей или тестовой коллекции.

Переход от дихотомии *субъективность vs. объективность* к тональной релевантности был вынужденным шагом, так как старые термины не могли более «обслуживать» новые потребности исследователей, особенно после расширения интереса к автоматическому анализу семантики и анализу тональности в частности, что связано с рядом факторов. Во-первых, это заинтересованность коммерческих компаний в подобных исследованиях, во-вторых, повышающийся интерес к развитию систем искусственного интеллекта, что невозможно без решения задачи понимания текста. В-третьих, конечно, развитие систем векторного представления слов (*word embeddings*), таких как пакет *Word2Vec*, которые повысили качество классификации документов при применении методов машинного обучения.

Даже в ранних монографиях, посвящённых анализу тональности, обсуждался вопрос применения терминов *субъективность* и *объективность*, переход исследовательского интереса на другой материал способствовал изменению терминологии и отказу от старой дихотомии.

## ЛИТЕРАТУРА

1. Алексеев А.А., Лукашевич Н.В. Автоматическое извлечение сущностей на основе структуры новостного кластера // Искусственный интеллект и принятие решений. 2011. № 4. С. 95–103.
2. Лукашевич Н.В., Левчик А.В. Создание лексикона оценочных слов русского языка RuСентиЛекс // Открытые семантические технологии проектирования интеллектуальных систем (OSTIS–2016): материалы VI Международной научно-технической конференции / отв. ред. В.В. Голенков. Минск: Белорусский государственный университет информатики и радиоэлектроники, 2016. С. 377–382.
3. Марчук Ю.Н. Автоматизация перевода и типология текстов // Вестник Московского государственного областного университета. Серия: Лингвистика. 2016. № 2. С. 164–171.
4. Пазельская А.Г., Соловьев А.Н. Метод определения эмоций в текстах на русском языке // Компьютерная лингвистика и интеллектуальные технологии: «Диалог-2011». М., 2011. С. 510–522.
5. Deng L., Wiebe J. Joint Prediction for Entity/Event-Level Sentiment Analysis using Probabilistic Soft Logic Models // Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, 2015. P. 179–189.
6. Deng L., Wiebe J. MPQA 3.0: An Entity/Event-Level Sentiment Corpus // Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL. Denver, 2015. P. 1323–1328.
7. Greene S., Resnik P. More than Words: Syntactic Packaging and Implicit Sentiment // Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the ACL. Boulder, 2009. P. 503–511.
8. Pang B., Lee L. Opinion Mining and Sentiment Analysis // Foundations and Trends in Information Retrieval. 2008. № 2. P. 1–135.
9. Scheible C., Schutze H. Sentiment Relevance // Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics. Sofia, 2013. P. 954–963.
10. Shin H., Kim M., Jo Y. и др. Annotation Scheme for Constructing Sentiment Corpus in Korean // 26th Pacific Asia Conference on Language, Information and Computation pages. Bali, 2012. P. 181–190.

## REFERENCES

1. Alekseev A.A., Lukashevich N.V. [Automatic extraction of entities based on the structure of news cluster] In: *Iskusstvennyi intellekt i prinyatie reshenii* [Artificial Intelligence and Decision Making], 2011, no. 4, pp. 95–103.
2. Lukashevich N.V., Levchik A.V. [Creating a lexicon of appraisal words Russian language RuSentiLeks]. In: *Golenkov V.V., ed. Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh sistem (OSTIS–2016): materialy VI mezhdunarodnoi nauchno-tekhnicheskoi konferentsii* [Open Semantic Technologies for Intelligent Systems (OSTIS–2016): materials of the VI International Scientific and Technical Conference]. Minsk, Belorussian State University of Informatics and Radioelectronics Publ., 2016, pp. 377–382.
3. Marchuk Yu.N. [Automation of the translation and typology of texts]. In: *Vestnik Moskovskogo gosudarstvennogo oblastnogo universiteta. Seriya: Lingvistika* [Bulletin of Moscow Region State University. Series: Linguistics], 2016, no. 2, pp. 164–171.
4. Pazel'skaya A.G., Solov'ev A.N. [The method of definition of emotions in Russian texts]. In: *Komp'yuternaya lingvistika i intellektual'nye tekhnologii: «Dialog-2011»* [Computational linguistics and intelligent technology: the “Dialogue-2011”]. Moscow, 2011, pp. 510–522.

5. Deng L., Wiebe J. Joint Prediction for the Entity/Event-Level Sentiment Analysis using Probabilistic Soft Logic Models. In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, 2015, pp. 179–189.
6. Deng L., Wiebe J. MPQA 3.0: An Entity/Event-Level Sentiment Corpus. In: *Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL*. Denver, 2015, pp. 1323–1328.
7. Greene S., Resnik P. More than Words: Syntactic Packaging and Implicit Sentiment. In: *Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the ACL*. 2009, pp. 503–511.
8. Pang B., Lee L. Opinion Mining and Sentiment Analysis. In: *Foundations and Trends in Information Retrieval*, 2008, no. 2, pp. 1–135.
9. Scheible S., Schutze H. [Sentiment Relevance]. In: *Proceedings of the 51<sup>st</sup> Annual Meeting of the Association for Computational Linguistics*. Sofia, 2013, pp. 954–963.
10. Shin H., Kim M., Jo Y., Jang H., Cattle A. [Annotation Scheme for Constructing Sentiment Corpus in Korean]. In: *26th Pacific Asia Conference on Language, Information and Computation pages*. 2012, pp. 181–190.

---

#### ИНФОРМАЦИЯ ОБ АВТОРЕ

Семина Татьяна Алексеевна – магистр лингвистики, аспирант кафедры теоретической и прикладной лингвистики Московского государственного областного университета;  
e-mail: taniasemina@gmail.com

#### INFORMATION ABOUT THE AUTHOR

Tatiana Semina – master of Philological sciences, postgraduate student at the Department of Theoretic and Applied Linguistics, Moscow Region State University;  
e-mail: taniasemina@gmail.com

---

#### ПРАВИЛЬНАЯ ССЫЛКА НА СТАТЬЮ

Семина Т.А. Дихотомия субъективность vs. объективность и тональная релевантность в задачах анализа тональности // Вестник Московского государственного областного университета. Серия: Лингвистика. 2018. № 1. С. 38–45.  
DOI: 10.18384/2310-712X-2018-1-38-45

#### FOR CITATION

Semina T.A. Subjectivity vs. Objectivity Dichotomy and Sentiment Relevance in Sentiment Analysis Tasks. In: *Bulletin of Moscow Region State University. Series: Linguistics*, 2018, no. 1, pp. 38–45.  
DOI: 10.18384/2310-712X-2018-1-38-45