

ЛИНГВИСТИЧЕСКАЯ МОДЕЛЬ ДЛЯ КОМПЬЮТЕРНОГО АНАЛИЗА ТОНАЛЬНОСТИ ПУБЛИКАЦИЙ СМИ

Ермаков А.Е., Киселев С.Л.

ООО “Гарант-Парк-Интернет”, ООО “Белый ветер”, Москва

ermakov@metric.ru

Доклад освещает опыт практического решения задачи определения тональности текста по отношению к заданному объекту. Систематизируются средства, используемые автором текста для формирования тонально окрашенного образа объекта, и строится лингвистическая модель для выделения всех составляющих этого образа. Описывается схема оценки тональности “позитив/негатив” с учетом тех мест, которые занимают в составе пропозиций тональные и нейтральные слова, средства выражения отрицания и инверсии смысла.

Введение

Речь пойдет о задаче компьютерного анализа тональности текста (публикации СМИ) по отношению к заданному объекту – персоне или организации. Эта задача оказалась востребованной на рынке в связи с активным развитием политехнологий и технологий компьютерной разведки. Оперативное исследование субъективного образа объекта, естественно возникающего или намеренно формируемого в информационном поле СМИ, является важнейшей составляющей обеспечения эффективной политики и безопасности бизнеса, оценки эффективности PR-компаний и прочих видов манипуляций информационно-обусловленным общественным сознанием в современном мире.

В рассматриваемой здесь постановке под тональностью текста понимается позитивное или негативное отношение его автора к заданному объекту, фигурирующему в тексте.

Интуитивно понятно, что оценка тональности требует анализа смысла, заложенного автором в текст. С лингвистической точки зрения смысл текста (его субъективное содержание) характеризуется следующими группами факторов:

- введение лексико-грамматических средств, выражающих модальные характеристики ситуации, модусные смыслы и явное отношение автора к описываемой ситуации, в том числе выбор тонально окрашенного слова взамен нейтрального из синонимического ряда. Очевидно, что этих факторы должны прежде всего учитываться при оценке тональности текста.
- трансформации “нейтральной” структуры предложения (которая в норме является изоморфно изоморфной конструкцией), связанные с изменением порядка слов, осложнением, трансформацией залога, введением показателей смысловых отношений, экспликаторов и прочих элементов. Факторы этой группы определяют расстановку акцентов,

ракурс подачи ситуации, выбранные автором текста для достижения своих коммуникативных целей. Не ясно, как использовать эти факторы для оценки тональности текста.

Вся остальная информация в тексте, с точки зрения лингвистики, представляет его объективное содержание - совокупность семантических отношений между объектами (предметами, событиями и их свойствами) в описанном фрагменте внеязыковой действительности, которые автор пожелал отразить. При этом лингвистика не рассматривает возможность искажения описанной ситуации или ее подмены другой ситуацией, хотя в этих случаях именно содержание, переставшее быть объективным, отражает смысл, что сплошь и рядом наблюдается в текстах СМИ. В итоге, за рамками лингвистических моделей остаются глубинные факторы экстралингвистической природы, связанные со способом формирования самого содержания текста, отбором тех ситуаций для включения в текст, которые, внешне не получив тональной окраски и смысловой обработки, тем не менее формируют тональность текста благодаря своей семантике, набору вызываемых коннотаций. Сюда относятся фразы типа “Ельцинские реформы привели народ к потере своих сбережений”, “Между масками Ельцина и Путина выступает Смерть с косой - символ этих десяти лет”. Такие способы выражения тональности очень часто встречаются в тексте, особенно при выражении негатива. Ориентируясь на модель восприятия текста целевой аудиторией, автор связывает с объектом описание таких событий или метафорических образов, восприятие которых вызывает эмоциональную реакцию и формирует негативное или позитивное отношение к самому объекту.

Из сказанного выше интуитивно понятно, что в общем случае никакими машинными методами невозможно разделить объективное и субъективное содержание текста – объективную констатацию фактов, пускай даже тонально окрашенных, и намеренное искажение действительности, в том

числе сознательное выведение в фокус внимания определенных ее сторон на фоне замалчивания других. К счастью для разработчиков, жизнь позволяет считать, что данной проблемы просто не существует, и заказной характер публикаций СМИ дает нам право отождествить со смыслом любое вызывающее эмоциональную реакцию содержание, исходя из принципа “раз это пишется – значит это кому-то нужно”.

Будем называть тонально-окрашенными или просто тональными такие элементы текста (слова, синтагмы, фразы), которые несут в себе оценочную семантику, например: *превосходный, бессовестно, робкие действия, медлительность, олигарх, вор, режим, бюрократизация, наглеть*.

Будем называть эмоционально-коннотативными такие элементы текста, семантика которых при восприятии вызывает эмоциональную реакцию вида “хорошо/плохо”. Строго говоря, такими элементами являются синтагмы, содержащие в себе событийный предикат вместе с его распространителем (*борьба с преступностью, повышение цен, пенсионеры голодают*), однако дальше мы будем называть эмоционально-коннотативными и отдельные элементы таких синтагм (*преступник, рост*), не забывая при этом, что окончательный эмоциональный коннотат “хорошо” или “плохо” они формируют только в сочетаниях друг с другом.

Окончательно, мы рассматриваем задачу оценки тональности в конструктивной постановке как процесс проведения следующих операций с текстом:

- 1) Распознавание всех упоминаний о целевом объекте в тексте, включая его полные, краткие, косвенные, местоименные и другие обозначения (необходимость в этом этапе отсутствует, если известно, что весь текст посвящен только целевому объекту);
- 2) Отсев и полный синтаксический разбор тех конструкций, в которых отражаются все ситуации (события и признаки), связанные с целевым объектом;
- 3) Выделение и классификация тех пропозиций, в которых явно выражается тональность, и тех пропозиций, которые описывают эмоционально-коннотативные ситуации.
- 4) Для каждой пропозиции принятие решения о тональности “позитив/негатив” с учетом тех мест, которые занимают в ее составе эмоционально-коннотативные, тональные и нейтральные слова, средства выражения отрицания.

Заключительный шаг - оценка общей тональности текста на основе тональности всех входящих в него пропозиций - является утилитарной задачей и зависит от целей, которые преследует конкретная прикладная система. Так, с точки зрения наших заказчиков, появление одного негатива в

тексте перечеркивает весь позитив. Тот факт, что под видом позитива в тексте может с иронией подаваться “махровый” негатив, распознать который машина в принципе не способна, подтверждает основательность данной точки зрения.

Проведение операций (1) и (2) требует наличия технических средств для проведения полного синтаксического анализа текста, выделения и отождествления наименований, которые к настоящему моменту разработаны в нашей компании [2,3]. Операции (3) и (4) относятся собственно к лингвистической модели оценки тональности и описываются далее.

Фильтрация и разбор пропозиций для оценки тональности

Технологически отбор пропозиций происходит как поиск в сети синтактико-семантических отношений (результат синтаксического анализа предложения) всех подсетей, изоморфных заданным лингвистическим схемам-графам [2]. При нахождении подсети, удовлетворяющей заданной схеме, происходит отбор тех ее узлов, которые соответствуют важным для оценки тональности участникам ситуации, а также связанных с ними узлов, содержащих дополнительную информацию (частицы, наречия, прилагательные, служебные глаголы). На основании класса, к которому отнесена пропозиция, лексического состава участников ситуации с учетом их ролей и семантики дополнительных слов принимается решение о тональности.

Отбор пропозиций, связанных с объектом мониторинга и используемых при оценке тональности, производится в соответствии с описанными ниже схемами.

Явная тональная характеристика

Ситуации класса “объект или инициированное им событие наделяется признаком, имеющим оценочную семантику”. Выделяются участники в следующих ролях:

Объект оценки – целевой объект;

Атрибут – существительное или именная группа, прилагательное, наречие, тонально окрашенный предикат-глагол.

Ниже перечислены типовые пропозиции, которыми выражаются ситуации этого класса, в которых курсивом выделены тонально окрашенные участники.

- 1) Логическая пропозиция полная: Петров – *плохой* руководитель; власть становится *защитницей преступности*; политика власти – *дрянь*. Тональность выражается именной группой, образуемой существительным.
- 2) Логическая пропозиция свернутая с существительным: *гениальный авантюрист* Петров; *робкая политика* президента, *циничность* власти. Тональность выражается

именной группой, образуемой существительным.

- 3) Логическая пропозиция свернутая с прилагательным: *нерешительный* президент; президент *нерешителен*. Тональность выражается прилагательным.
- 4) Свернутая логическая пропозиция в составе событийной, отражающая оценку события, в котором целевой объект выступает в роли протагониста: Петров *бездумно* согласился; президент принял *авантюрное* решение; власть *обнаглела*. Тональность может выражаться наречием при глаголе, прилагательным при событийном существительном, самим глаголом.

Прямая эмоционально-коннотативная характеристика

Класс эмоционально-коннотативных ситуаций, отражающих отношение целевого объекта к эмоционально-коннотативным сущностям, их оценку целевым объектом, или наоборот – отношение этих сущностей к объекту, оценку объекта ими. Выражается событийными пропозициями: *президент борется с преступностью, народ выносит осуждение власти*. Выделяются участники в следующих ролях:

Субъект – активный участник, в приведенных примерах – *президент* и *народ*;

Объект – пассивный участник ситуации, в приведенных примерах – *преступность* и *политика власти*;

Предикат – глагол или существительное, выражающее отношение *Субъекта* к *Объекту* (в приведенных примерах – *борьба с* и *осуждение*).

Типовые пропозиции, которыми выражаются ситуации этого класса, следующие:

- 1) Событийная пропозиция полная или свернутая, в которой роль *Субъекта* занимает целевой объект: власть *борется с олигархами*; президент ведет *борьбу с коррупцией*, *борьба* президента за *права народа*. Тональность складывается из семантики именной группы в роли *Объекта* и семантики *Предиката* по принципу “положительное отношение к положительному - позитив” и наоборот. Если эмоциональный коннотат *Объекта* или *Предиката* не определен, тональность считается нейтральной (президент встретился с *олигархами*, Иванов *борется с сорняками* на даче).
- 2) Событийная пропозиция полная или свернутая, в которой роль *Объекта* занимает целевой объект: *олигархи испугались* президента, *страна* выражает *недоверие* к власти, *ненависть преступников* к власти. В случае, если семантика *Субъекта* имеет положительный эмоциональный коннотат, общая тональность складывается по тому же принципу, что и в (1). Если же семантика *Субъекта* имеет отрицательный коннотат, то общая тональность

не определена: *олигархи полюбили* президента (да мало ли за что?).

Ассоциированный эмоциональный коннотат

Класс эмоционально-коннотативных ситуаций, фигурирующих в одном предложении с целевым объектом, но не связанных с ним напрямую (в ряде случаев эту связь просто не удастся идентифицировать средствами автоматического анализа текста). Выражаются событийными пропозициями: *обнищание пенсионеров, повысить зарплату, рождаемость падает, пристраститься к пиву*. Выделяются участники в следующих ролях:

Участник – участник, на состояние которого влияет событие;

Предикат – событие, которое влияет на *Участника*.

Тональность складывается из семантики именной группы в роли *Участник* и семантики *Предиката* по принципу “хорошо для хорошего - позитив” и наоборот. Если эмоциональный коннотат *Участника* или *Предиката* не определен, тональность считается нейтральной.

Методика принятия решения о тональности

Способы оценки тональности отдельных пропозиций в зависимости от их типа качественно описаны выше, и при наличии всех слов и словосочетаний в словарях с оценкой тональности и эмоционального коннотата задачу можно считать решенной. Однако, поскольку число комбинаций слов практически бесконечно, необходим подход, позволяющий определить тональность участника ситуации по тональности составляющих его слов.

Для примера рассмотрим возможные комбинации тональности слов в составе именной группы “прилагательное-существительное” (Таблица 1).

Все прочие примеры, в том числе сочетания “глагол-наречие” укладываются в подобную схему, которая может быть обобщена так: “При наличии хотя бы одного слова с негативной тональностью общая тональность участника негативна; в противном случае общая тональность позитивна, если присутствует хотя бы одно слово с позитивной тональностью”.

Пример	Тональность существительного	Тональность прилагательного	Общая тональность
Плохой руководитель, Хороший руководитель	Нейтрально	Любая	Тональность прилагательного
Новый олигарх, новый победитель	Любая	Нейтрально	Тональность существительного
Плохой защитник демократии	Позитив	Негатив	Негатив
Отличный защитник демократии	Позитив	Позитив	Позитив
Слабый диктатор, Сильный диктатор	Негатив	Любой	Негатив

Таблица 1. Общая оценка тональности именной группы в зависимости от комбинаций тональностей составляющих ее слов

Дополнительно при оценке тональности каждого участника ситуации необходимо учесть наличие при нем слов, которые инвертируют тональность – частиц, наречий, прилагательных и некоторых глаголов. Показатели инверсии часто могут употребляться совместно, причем четное количество инверсий эквивалентно отсутствию таковой, а нечетное есть инверсия (*якобы мало борется = борется, якобы не безуспешно борется = якобы борется = не борется = безуспешно борется*). В ситуациях, выраженных предикативной конструкцией, показатели тональности и отрицание могут быть выражены не при главном, а при дополнительном предикате (который и сам по себе может выражать отрицание), например: *власть слишком медленно пытается укрепиться, президент начал бы активно борьбу с олигархами, президент якобы не отказался повысить пенсии*.

Влияние показателей отрицания на тональность отдельных слов необходимо просчитывать до применения указанного выше принципа получения результирующей тональности участников факта. Исключением является случай, когда показатель инверсии стоит на самом целевом объекте: *“не президент отвечает за...”*, что инвертирует общую окончательную оценку тональности ситуации.

Общая схема тональной оценки каждой пропозиции выглядит следующим образом:

- Определение тональности всех словосочетаний, относящихся к участникам ситуации, по словарям (для каждой роли необходимы свои словари), включая учет показателей инверсий по словарям;
- Полная оценка каждого участника ситуации на основании тональности относящихся к нему слов;
- Оценка тональности всей пропозиции в соответствии с ее схемой и ролями участников;
- При наличии показателя инверсии на целевом объекте – инверсия тональности всей пропозиции.

Заключение

Несмотря на большой интерес, проявляемый к задаче оценки тональности, нам известна только одна практическая разработка в данной проблемной области – “Проект ВААЛ” [1]. Постановка задачи, решаемой ВААЛом, наиболее точно, на наш взгляд, отражена парой следующих цитат с сайта [1]: “... оценивать неосознаваемое эмоциональное воздействие фонетической структуры текстов и отдельных слов на подсознание человека” и “производить эмоционально-лексический анализ текстов”. Применяемые алгоритмы – это превращение текста в частотный словарь и отнесение некоторых слов к определенным психолингвистическим категориям, которые позволяют, к примеру, определить, что текст отражает “истероидное” или “депрессивное” настроение автора (категории “Психиатрического анализа”). Об анализе семантики текста, который невозможен без учета синтаксических связей слов, речь, очевидно, не идет. Это принципиально ограничивает применимость прикладных решений, которые могли бы быть разработаны на базе одной только технологии ВААЛ, однако позволяет рассматривать проведенные авторами исследования в области лексической фоно- и психосемантики как важный и, возможно, уже достаточно глубоко проработанный элемент полной многофакторной модели оценки тональности текста. Весьма вероятно, что разработанные авторами проекта ВААЛ словари и методики категоризации могли бы успешно использоваться на последних этапах описанной здесь схемы анализа текста, позволяя производить более тонкую оценку его содержания по различным шкалам.

Описанная нами лингвистическая модель является в основном полной. Как показали эксперименты, ошибки оценки тональности на выходе всей технологической схемы в большей степени определяются ошибками синтаксического разбора текста, нежели неучтенными в рамках модели факторами.

Следует отметить, что с высокой достоверностью удастся выделить только негатив, тогда как из “формального” позитива чуть ли не половина оказывается скрытым негативом, иронией, или же попросту перечеркивается рядом стоящим негативом. Это становится очевидным, если учесть хотя бы тот факт, что обругать хорошими словами весьма естественно, а похвалить плохими – весьма затруднительно. По нашему мнению, выделение позитива с приемлемой точностью в общем случае невозможно, во всяком случае, при анализе текстов из “желтой” прессы. Возможно, правда, что данное утверждение относится только к локальным фрагментам текста, так как эксперименты по совокупной оценке тональности текста на базе оценки его составляющих нами пока не проводились.

С учетом сказанного, основным утилитарным результатом применения системы компьютерной оценки тональности является не подсчет соотношения “хвалят/ругают”, а формирование частотного портрета всех позитивных и негативных событий, связываемых в тексте (прямо или ассоциативно) с объектом. На наш взгляд, целевая система должна позиционироваться именно как инструмент для отсева и классификации событий, формирующих образ объекта в информационном поле. Не столь важно точно узнать, где и как хвалят или ругают, важнее узнать - за что?

Список литературы:

- 1) “Проект ВААЛ”: <http://www.vaal.ru/>
- 2) Ермаков А.Е. Референция обозначений персон и организаций в русскоязычных текстах СМИ: эмпирические закономерности для компьютерного анализа. // В настоящем сборнике
- 3) Киселев С.Л., Ермаков А.Е., Плешко В.В. Поиск фактов в тексте естественного языка на основе сетевых описаний // Компьютерная лингвистика и интеллектуальные технологии: труды Международной конференции Диалог’2004. – Москва, Наука, 2004. – С. 282-285.