

*Васильев Василий Васильевич*

*Старший преподаватель кафедры «Английский язык и перевод»,*

*Северо-Восточный федеральный университет им. М.К. Аммосова*

## **КОМПЬЮТЕРНЫЙ АНАЛИЗ ТОНАЛЬНОСТИ РЕГИОНАЛЬНЫХ СМИ**

*Аннотация: Анализ тональности — в одном из своих применений, класс методов контент-анализа в компьютерной лингвистике, основная задача которого заключается в классификации текста по его настроению. С помощью анализа тональности исследователи могут обобщать тональность текстов и делать выводы по разным темам.*

*Ключевые слова: анализ тональности; коммуникация эмоция; семантика.*

*Vasiliev Vasily Vasilyevich*

*Senior Lecturer, Department of English Language and Translation,*

*M.K. Ammosov North-Eastern Federal University*

## **SENTIMENT ANALYSIS OF REGIONAL MEDIA**

*Abstract: sentiment analysis is, in one of its applications, a class of content analysis methods in computational linguistics, the main task of which is to classify text by its mood. With the help of sentiment analysis, researchers can generalize the tonality of texts and draw conclusions on various topics.*

*Keywords: sentiment analysis; communication emotion; semantics.*

Анализ тональности стал мощным инструментом для масштабной обработки мнений, выражаемых в любых текстовых источниках. [2] Цель исследования - изучение и составление языкового образа новостных порталов на примере регионального СМИ «ЯСИА», «Якутия24», с целью выявления наиболее обсуждаемых категорий, единиц, анализа общественного мнения и реакции на данную тему. Для исследования будет использоваться текстовый материал в виде статей смешанной тематической направленности с сайтов региональных СМИ. Основными характеристиками данных будут являться выведение именованных сущностей составляющих ключевую тематику,

обсуждаемую в СМИ и выявление семантических характеристик данных текстов. Составляется корпус новостей, который в дальнейшем будет подвергнут как машинной обработке, так и лингвистическому анализу.

Для исследования будет использоваться текстовый материал в виде статей смешанной тематической направленности с сайтов региональных СМИ. Основными характеристиками данных будут являться выведение именованных сущностей составляющих ключевую тематику, обсуждаемую в СМИ и выявление семантических характеристик данных текстов. Составляется корпус новостей, который в дальнейшем будет подвергнут как машинной обработке, так и лингвистическому анализу.

Методы исследования: Для определения тональности текста используются определенная выборка слов и загружается в менеджер корпусов AntConc. Более точное определение тональности текста будет ясно указываться при изменении размера кластера на значение 2. При этом программа будет выдавать словосочетания, по которым можно будет в значительной степени легче определять тональность.

Sentistrength —использует словари, оценки тональности слов и правила для определения общей тональности текста. Sentistrength является одним из эффективных методов анализа тональности текста, так как он учитывает не только оценки слов, но и их контекст в предложении. Однако, он может быть неэффективен в случае, если отзывы содержат много ошибок или нетипичных конструкций. Чтобы программа показала достоверные результаты, надо предварительно обработать данные: удалить стоп-слова (часто используемые слова, не несущих смысловой нагрузки) и пунктуации, привести слова к единому регистру, лемматизировать. [3]

Алгоритм исследования:

- Сбор текстового корпуса – новостные статьи, периода 2023-2024 года, продолжается сбор более ранних статей. Как правило региональный

портал новостей представляет собой собрание материалов, охватывающих разные области жизни региона, таким образом представленный материал будет разноплановым, что в свою очередь соответствует критериям полноты и репрезентативности, представляемым к текстовым корпусам.

- Обработка языковых данных - введение данных в аналитическую систему.
- Индексация материала, на данном этапе возможна и частеречная разметка
- Фильтрация и очистка списка слов
- Ранжирование по частотности, сочетаемости.
- На полученных статистических данных – составление языкового портрета СМИ
- Разбиение по семантическим областям, полям
- Извлечение именованных сущностей, частотных объектов обсуждения
- Автоматизированная оценка тональности
- Лингвистический стилистический анализ
- Интерпретация и характеристика полученных данных
- Оценка восприятия, стимул – ключевые слова, ассоциативный сбор
- Возможен перевод специалистом и машинный перевод
- Оценка тональности переведенного текста
- Сравнение полученных результатов, оценка потенциала передачи тональности при выполнении перевода
- Формулировка рекомендаций к улучшению инструментария областей оценки тональности и машинного перевода.

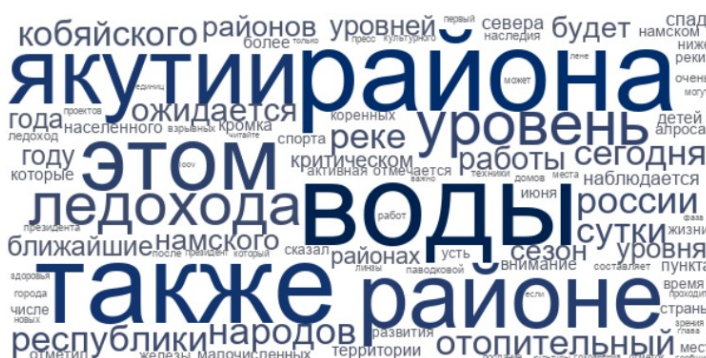
На материале корпуса ЯСИА выведена следующая частотность:

|   |          |   |    |   |
|---|----------|---|----|---|
| 1 | воды     | 1 | 60 | 1 |
| 2 | района   | 2 | 43 | 1 |
| 3 | также    | 3 | 35 | 1 |
| 4 | якутии   | 4 | 33 | 1 |
| 5 | районе   | 5 | 29 | 1 |
| 6 | этом     | 6 | 26 | 1 |
| 7 | уровень  | 7 | 24 | 1 |
| 8 | ледохода | 8 | 21 | 1 |

|    |              |    |    |   |
|----|--------------|----|----|---|
| 9  | народов      | 8  | 21 | 1 |
| 10 | россии       | 8  | 21 | 1 |
| 11 | республики   | 11 | 20 | 1 |
| 12 | сутки        | 11 | 20 | 1 |
| 13 | отопительный | 13 | 18 | 1 |
| 14 | реке         | 13 | 18 | 1 |
| 15 | сегодня      | 13 | 18 | 1 |

Фильтрация для удаления служебных слов произведена по регулярному выражению: `\w{4,14}`. Также возможно выполнение фильтрации посредством создания и включения списка стоп-слов.

Облако слов новостей за период - май 2024 г.



## Комплексный анализ тональности.

Ныне существующий компьютерный инструментарий средств исследования тональности не полностью удовлетворяет нужды лингвиста в данной области, поскольку в сфере компьютерной лингвистики направление анализа тональности ограничено исследованием полярности и извлечением эмоций в тексте. Для исследования коммуникативной тональности потребуется комбинированный гибридный подход с явной долей участия человека в человеко-машинном процессе анализа. [4]

Первичным пунктом в алгоритме мы поставим выявление дискурсивных зон, тем обсуждения и объекты обсуждения. Как путем статистического анализа корпуса текстов, так и с применением нейротехнологий в компьютерном анализе тональности, в рамках которой автоматически могут быть распознаны именованные сущности и их классы.

Следующим пунктом будет выделение единиц тональности, как путем автоматического анализа системами сентимент-анализа, извлекающими тонально окрашенные единицы и определяющие по ним настроение, полярность текста, так и проведение собственного экспертного анализа. Экспертный анализ будет состоять в отборе из текста единиц тональности, описания текста в целом, описание общей коммуникативной ситуации.

Результаты исследования подчеркивают значимость классификации текстовой тональности как важного аспекта лингвистических исследований. Это позволяет лучше понять, как язык формирует эмоциональное восприятие и коммуникативные намерения. Анализ существующих методов классификации и выявление ключевых индикаторов открывают новые возможности для их применения в различных областях, таких как маркетинг, психология и социология. Это делает тему актуальной и востребованной в условиях современного информационного общества.

### **Список литературы:**

1. Котельников Е.В. Функция оценки информативности гипотез для анализа тональности текстов на основе ДСМ-метода // Фундаментальные исследования. — 2014. — № 11. — С. 2150–2151.
2. Liu, Bing. Sentiment Analysis and Subjectivity // Handbook of Natural Language Processing (англ.) / под ред. N. Indurkha и F. J. Damerau. — 2010.

3. Pang, Bo, Lee, Lillian. Opinion Mining and Sentiment Analysis (англ.) // Foundations and Trends in Information Retrieval : журнал. — 2008. — No. 2. — P. 1—135.
4. Богданова Т.Ф. Лексическая тональность в деловой интернет-переписке на русском и английском языках // Вестник Костромского государственного университета. 2023. Т. 29, № 3. С. 194—199.