

# project

Предмет исследования: Тонально окрашенные глаголы русского языка из словаря оценочных фреймов RuSentiFrames.

Методика сбора данных: Глаголы были размечены на наличие/отсутствие тональности пятью респондентами.

Те глаголы, которые один и более респондентов отметили, как тонально окрашенные, были извлечены в отдельную таблицу, и размечены на позитив/негатив.

Так же был проведен эксперимент с предобученными моделями Fasttext, взятыми с ресурса RusVectores. Были выбраны сидовые множества из сильно тонально окрашенных глаголов(позитивное и негативное), и с помощью модели были измерены косинусные расстояния до этих множеств, для каждого из глаголов.

Глаголы, более близкие к позитивному множеству были оценены, как позитивные, и более близкие к негативному - как негативные.

Целью исследования является оценка результата работы модели, и наличия связи между оценкой разметчиков и оценкой модели на позитив негатив, и наличия зависимости количества совпадений между этими оценками от коэффициента тональности, если принять его равным количеству человек, которые сочли глагол тональным(данные из разметки глаголов на наличие отсутствие тональности в целом), а косинусную близость до сидового множества считать коэффициентом тональности, который присвоила глаголу модель.

Гипотеза состоит в том, что связь между оценкой тональности, присвоенной разметчиками, и оценкой тональности и полярности при помощи модели есть.

Еще одна гипотеза состоит в том, что негативные глаголы более тонально окрашены, чем позитивные.

Данные - разметка на наличие отсутствие тональности.

```
tonality_grisha <- read.csv("https://raw.githubusercontent.com/DmitryKutsev/NIS_SentiFrame/master/annotations/annotation_arshinov.csv")
tonality_masha <- read.csv("https://raw.githubusercontent.com/DmitryKutsev/NIS_SentiFrame/master/annotations/annotation_masha.csv")
tonality_ira <- read.csv("https://raw.githubusercontent.com/DmitryKutsev/NIS_SentiFrame/master/annotations/annotation_ira.csv")
tonality_olya <- read.csv("https://raw.githubusercontent.com/DmitryKutsev/NIS_SentiFrame/master/annotations/annotation_olya.csv")
tonality_lesha <- read.csv("https://raw.githubusercontent.com/DmitryKutsev/NIS_SentiFrame/master/annotations/annotation_lesha.csv")
head(tonality_olya)
```

```
##          verb X0.1
## 1  сообщать      0
## 2  сообщить      0
## 3  принять       1
## 4  задержать     1
## 5  провести      0
## 6  пройти        0
```

Отобранные глаголы, размеченные на позитив/негатив.

```
polar_dima <- read.csv("https://raw.githubusercontent.com/DmitryKutsev/NIS_SentiFrame/master/annotations/polarity_annotation/polar_annotation_dima.csv")
polar_ira <- read.csv("https://raw.githubusercontent.com/DmitryKutsev/NIS_SentiFrame/master/annotations/polarity_annotation/polar_annotation_ira.csv")
polar_masha <- read.csv("https://raw.githubusercontent.com/DmitryKutsev/NIS_SentiFrame/master/annotations/polarity_annotation/polar_annotation_masha.csv")
head(polar_dima)
```

```
##          verb tonality polarity
## 1  агитировать        1         1
## 2  адаптироваться     1         0
## 3  арестовать         6         0
## 4  арестовывать       6         0
## 5  атаковать          6         0
## 6  бастовать          6         0
```

Данные - ручная разметка на полярность тональности, плюс определение полярности тональности с помощью модели fasttext.

manual\_tonality - сумма голосов, присвоенных разметчиками глаголу(от 1 до 5), разделенная на кол-во разметчиков.

fasttext\_tonality - косинусное расстояние до наиболее близкого сидового множества

manual\_polarity и fasttext\_polarity - полярности(позитив/негатив), присвоенные разметчиками и моделью.

```
model_manual_df <- read.csv("https://raw.githubusercontent.com/DmitryKutsev/NIS_SentiFrame/master/annotations/polarity_annotation/my_full_df.csv")
head(model_manual_df)
```

```
##          verb manual_tonality fasttext_tonality manual_polarity
## 1   агитировать      0.1666667      0.5600420             1
## 2   адаптироваться  0.1666667      0.3728098             1
## 3   арестовать     1.0000000      0.3292224             0
## 4   арестовывать   1.0000000      0.2764541             0
## 5   атаковать      1.0000000      0.4387920             0
## 6   бастовать      1.0000000      0.5433168             0
## fasttext_polarity
## 1             1
## 2             1
## 3             1
## 4             0
## 5             0
## 6             1
```