

Using emotional evaluation of text in a foreign language learning app

Bulat Shakirov ¹, Liliya Khairullina ¹, Leysan Mingalieva ¹, Veronika Bronskaya ^{1,2}, Guzel Khabibullina ¹, Elena Fadeeva ¹ and Olga Kharitonova ²

¹Kazan Federal University, 35 Kremlyovskaya Street, Kazan 420008, Russia

²Kazan National Research Technological University, 68 Karl Marx Street, Kazan 420015, Russia

Abstract

The current level of development of computational linguistics is characterized by the involvement of more and more complex levels of linguistic analysis in the field of automatic analysis, the use of hybrid approaches in solving computer text processing problems that combine machine learning and algorithmic methods. At the same time, the levels of complexity of modern text processing tasks, such as extracting time reference in a text, analyzing the structure of discourse and many others, require the active involvement of expert linguistic knowledge. Nowadays, learning a foreign language is the goal of a very large number of people. There are many tools, methods and techniques for self-development of a foreign language that can significantly increase the effectiveness of training. Currently, sentiment analysis is used in monitoring, analytical and alarm systems, as well as in document management systems and advertising platforms. At the same time, this technology is practically not used in foreign language teaching systems, although it has a huge untapped potential. Almost all currently existing applications for learning foreign languages do not use the ability to analyze the tonality of the text. Mastering the emotional vocabulary of a language is a difficult task for a student, so this aspect should be given special attention. This paper is devoted to the description of a web application developed by us for learning German language, which is based on the use of sentiment analysis. This solution can be extended to other foreign languages as well.

Keywords

sentiment analysis, web application, learning app

1. Introduction

Mastering reading in a foreign language is a complex mechanism. Reading is a source of development and enrichment of oral speech, improvement of its quality side. The use of authentic texts, including fiction, in teaching a foreign language is becoming more and more widespread. Russian and foreign researchers emphasize the importance of fiction in teaching a foreign language: it teaches understanding of the peculiarities of different cultures, forms creative thinking skills, and helps to improve communication and language skills [1-7]. When translating a literary text, difficulties may arise associated with understanding emotive vocabulary. Understanding the tonality of the text is an important and necessary condition, firstly, to identify the motives and characteristics of the behavior of the characters, their inner world and the personality of the author of the work, which contributes to the

III International Workshop on Modeling, Information Processing and Computing (MIP: Computing-2021), May 28, 2021, Krasnoyarsk, Russia

EMAIL: bulat.crespo@gmail.com (Bulat Shakirov); Liliya-v1@yandex.ru (Liliya Khairullina); leysmy@yandex.ru (Leysan Mingalieva); dveronika@mail.ru (Veronika Bronskaya); hg21980@rambler.ru (Guzel Khabibullina); lenoktgpy@mail.ru (Elena Fadeeva); olga.220499@mail.ru (Olga Kharitonova)

ORCID: 0000-0003-3151-0815 (Bulat Shakirov); 0000-0003-2178-2948 (Liliya Khairullina); (Leysan Mingalieva); 0000-0003-2611-2679 (Veronika Bronskaya); 0000-0002-5881-0300 (Guzel Khabibullina); 0000-0003-2178-2948 (Elena Fadeeva); 0000-0001-6278-1744 (Olga Kharitonova)



© 2021 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

understanding of the ideological meaning of the work; secondly, for understanding human psychology, culture, mentality of people speaking a foreign language [7]. One of the ways to solve this problem is to use sentiment analysis when translating authentic texts.

Despite the fact that emotional analysis is already actively used in many types of human activity (when studying consumer preferences, in recommendation systems designed to help the consumer when choosing goods or services; when analyzing news resources, etc.), in the study of foreign languages this technology has not yet found widespread use. This work fills this gap to some extent. The work is aimed at studying the rule-based method of recognizing the sentiment of texts and the implementation of an interactive application for learning the German language with automatic sentiment determination of both the text as a whole and its individual parts.

2. Sentiment analysis of the text

The term "tonality analysis" has been used since 2003. One of the first papers to use the term was a scientific paper by T. Nasukawa and J. Yi, which investigated the extraction of opinions in relation to a given subject [8, 9]. Note that scientific work on tonality analysis began several years before this term was introduced. One of the first works on the analysis of tonality can be considered the work of J. Vibi on identifying the point of view of characters in works of fiction. The aim of the work was to automatically classify text fragments into objective fragments and fragments that Express the opinion or emotions of one of the characters [10-14]. Later, the object of research interest was the semantic orientation of adjectives [4, 5]. J. Vibe, R. Bruce, and T. O'hara began testing the application of statistical methods to the analysis of subjectivity [15]. The number of studies in the field of text tonality analysis is growing every year [1, 2, 10-13, 16-17]. Tonality analysis is a useful tool in almost any language processing task, since evaluative and emotive are important characteristics of speech.

Tonality or sentiment (opinion orientation, sentiment) is the emotional color expressed in the text. There are three types of tonality: positive, negative, and neutral. However, in the scientific literature, neutral tonality is defined differently, so very often only the first two types of tonality are taken into account. Some researchers define neutral tonality as some intermediate position between positive and negative, and some define it as the absence of any emotional coloring.

The object of text tonality analysis is any entity (product, service, problem, personality, etc.) that the text expresses an opinion about. Very often, objects can consist of separate parts (components) and properties (features). Components and properties make up many aspects.

In the dictionary of Efremova, an opinion is defined as "1) a judgment expressing an assessment of someone, something, an attitude to someone, something, a view of someone, something, 2) an official conclusion, decision" [2]. In the analysis of the emotional coloring of a text, an opinion is defined through a tuple [8] consisting of five elements:

$$(e, a_i, s, h, t),$$

where

e (entity) - an object about aspect a_i of which the author *g* expressed an opinion at time *t*. The object can be a product, person, event, organization or topic of discussion;

a_i (aspect) - property of the object in relation to which the opinion is expressed;

s (sentiment) - sentiment of opinion in relation to the *i*-th aspect of the essence *e*;

h (holder) - opinion exponent (subject);

t (time) - time of expression of opinion.

All the five elements are related to each other, if there is a mismatch – a big mistake. That is, if the source of the opinion *h* is incorrectly set, the key *s* is incorrectly determined.

Sentiment analysis, or text sentiment analysis, is a developing direction of computational linguistics, the main task of which is to identify emotionally colored vocabulary and emotional assessment of objects by the author in a document. Sentiment analysis can be viewed as a method for quantitatively describing qualitative data, implemented by assigning some sentiment estimates [3].

The tonal sentence analysis is shown in Figure 1:.

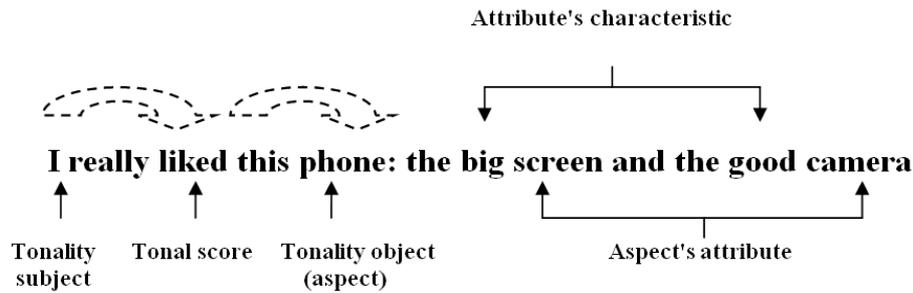


Figure 1: Parsing a tonal sentence

3. Rule-based method for evaluating emotionally coloured texts

Emotional text evaluation methods fall into one of the following four categories [11]:

- rule-based methods;
- methods based on the use of dictionaries of emotional and evaluative vocabulary (sentiment lexicon);
- supervised learning methods;
- unsupervised (deep) teaching methods.

The web application we have developed uses a rule-based approach to text sentiment analysis. This approach is based on linguistics; in this approach, the semantics of words and the rules for constructing sentences play an important role. This approach implies a tonal vocabulary containing words or collocations. Each word is assigned two attributes indicating tonality and / or strength of tonality (for example, on a scale of one to ten, where ten is a strong positive tonality). The tonal dictionary can be taken from the outside, or it can be generated statistically. A syntax tree is formed from each sentence or part of it, containing chains of words or collocations that depend on each other. The object of analysis and the direction of the sentiment are determined. The linguistic rules that form syntactic trees can be formed according to the researcher's preferences. A linguistic rule, or pattern, is a template that specifies a certain linguistic regularity, that is, a specification of a property of a set of examples, defined in terms of some formal language. Further, certain words or their combinations are compared with words from tonal dictionaries and, thus, the direction of tonality and its strength are assigned to individual sentences. The tonality of the entire text can be formed based on the tonality of its parts. This method allows you to get syntactically correct emotional expressions and, with good filling of tonal vocabulary lists, allows you to achieve good completeness (coverage of emotive vocabulary).

Note that compiling a set of rules is a very laborious process. In our work, about 20 linguistic patterns were compiled to assess emotionally colored texts. The negation words influencing the polarity of the sentence were taken into account - "nicht", "kein" ("not", "not"); opposite conjunctions "aber" ("but"), "nicht ..., sondern" ("not ..., but"). Complex sentences were broken down into simple ones. The clauses were associated with the word being defined. In the case of a verb with zero tonality, the emotional color of the sentence was determined by the tonal environment of the object, its qualitative characteristics.

Here is an example of the simplest rule, formalized as a pattern, for the sentence "Arzt heilte grippepatienten" ("The doctor cured a patient with the flu"):

if the predicate is included in a positive set of verbs, and there are no negatives and opposite conjunctions in the sentence, then classify the tonality as "positive"

The algorithm for emotional evaluation of the text consists of several stages:

1. Pre-processing of text, selection and classification of found words:

Arzt <noun, pos> heilte <verb, pos> grippepatienten <noun, neg>

Abbreviations are introduced here: noun - a noun, verb - a verb, pos - positive, neg - negative.

2. Combining the found words into chains linked to each other

Arzt <nPos> heilte <vPos> grippepatienten <nNeg>

3. Assessment of the tonality of the entire proposal

Arzt heilte grippepatienten <Pos>

4. Description of the application functionality

The service «Wir lerne Deutsch» is implemented using the Vue.js framework and Firebase cloud data storage using the open source library of AFINN sentiment created by Finn Orup Nielsen. AFINN is one of the most popular libraries used for emotional analysis of text. The AFINN library contains over 3300 words with an assessment of the polarity (rating) of each word [12-13]. The rating ranges from -5 to 5. For example,

- hopefulness - 2 points
- smart - 1 point
- slow - (-2) points
- amazing - 4 points
- breathtaking -5 points

As you can see from the list, words can represent different parts of speech. They all have their own rating. There are also neutral words with a rating of 0. This means that they do not carry either a positive or negative connotation [14-16].

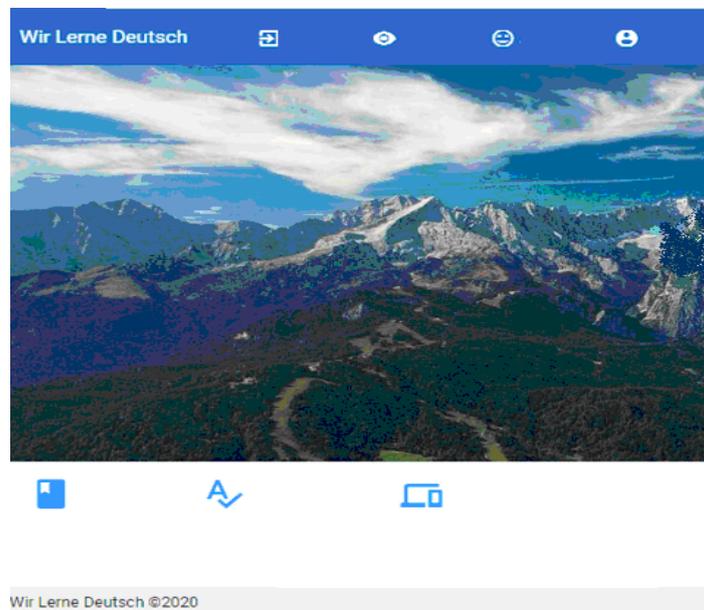


Figure 2: Administrative panel of the application

The developed web application allows you to read and translate fiction and popular science literature in German at various levels of difficulty, from basic (A1) to the level of a native speaker (C2). Literature is downloaded from sources that are freely available. To search for books of interest, the user can use multiple searches by the title of the book, as well as by the level of difficulty:

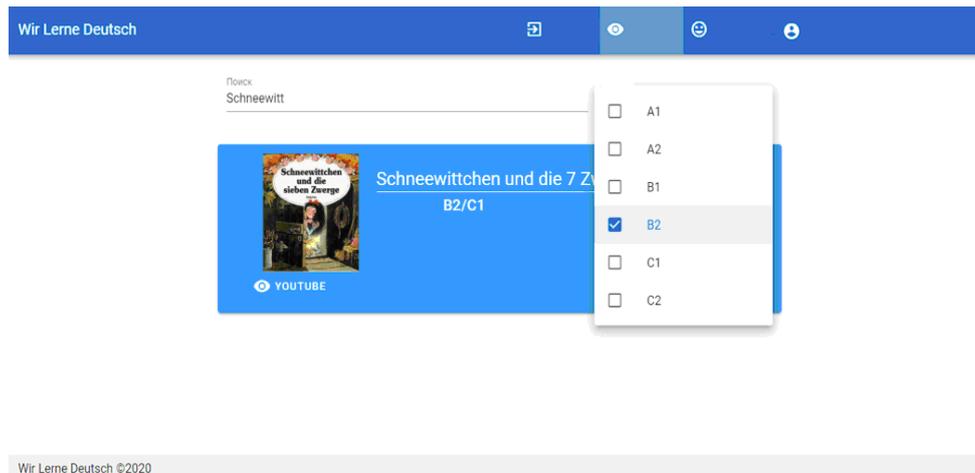


Figure 3: Application interface

After selecting literature, the user is directed to a page with a list of its chapters. Each chapter is presented with an audio recording (read by a native speaker), text with tips and words for study on this chapter.

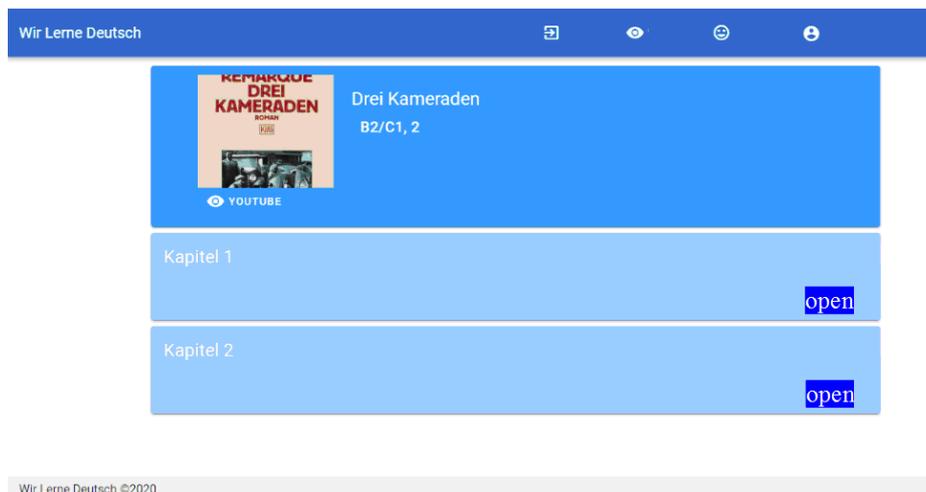


Figure 4: List of book chapters

At his discretion, the user can choose two reading modes:

1. text with prompts, that is, generating a translation and emotional colouring only by pressing the corresponding button;
2. parallel translation.

When choosing the first mode, the user can take two types of hints during reading:

Text translation is carried out by clicking on the symbol - 

Hint of the emotional context of the sentence by clicking on the symbol - 

When choosing a hint of an emotional context, three assessments are given in the form of pictograms, which make it possible to understand the message of the author of a foreign text:

- Positive assessment. Symbol - 
- Neutral assessment. Symbol - 
- Negative assessment. Symbol - 

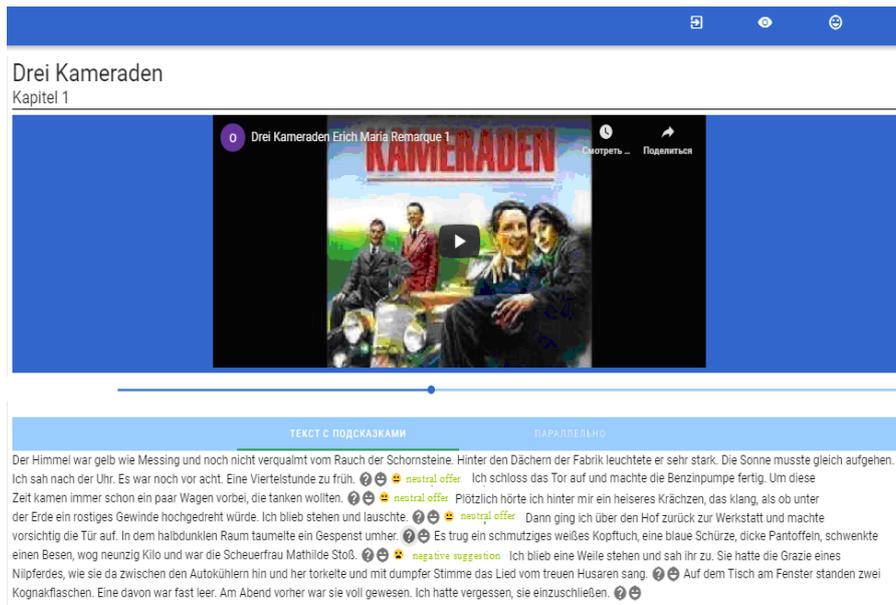


Figure 5: Page with the analysis of the sentiment of the text

Table 1
Text fragments and their emotional evaluation

Source code	Tonality
Der Himmel war wie Messing und noch nicht verqualmt vom Rauch der Schornsteine	Neutral
Es trug weites Koptfuch, eine blaue Schürze, dicke Pantoffein, schwenkte einen Besen, neunzig und war Stol	Negativ
Es war ein gut Tag	Positive

Note that you can also analyze the emotional coloring of a word, sentence or piece of text that interested the user in the Sentiment Analysis section. To do this, just fill in the special field “Enter text for sentiment analysis”. As a result, the user will receive an assessment of the text (positive, neutral, negative tone) and the number of points (from -5 to 5) corresponding to the given text. The larger the number, the more pronounced the positive tone is:

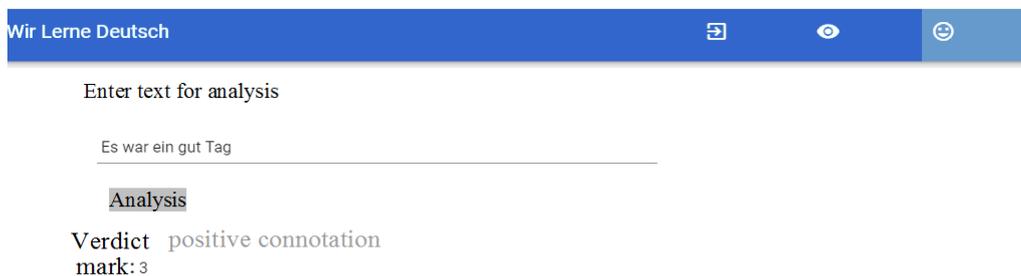


Figure 6: Sentiment analysis of the text in German

Recall that sentiment analysis of the text in the application is implemented using the method of using rules, as well as using the AFINN vocabulary.

To determine the emotional color of the text entered by the user, the getEmoji () function was developed:

```

getEmoji: function () {
  const ml = require ('ml-sentiment') ({lang: 'de'});
  this.senti = ml.classify (this.sentiAnalysis);
  if (this.senti <0) {
    this.emoji = "Negative Hue"
  }
  if (this.senti > 0) {
    this.emoji = "Positive Tint"
  }
  if (this.senti == 0) {
    this.emoji = "Neutral Tint"
  }
}

```

5. Outcome evaluation

At present, methods of objective testing of text tone marking systems have not yet been developed. Therefore, when analyzing the correctness of the automatic emotional assessment of the text, we based on the subjective assessment of small test sets by volunteers. To do this, we manually divided into fragments the novel "Three Comrades" by E.M. Remark and posted it on a lexical resource for crowdsourcing tonal markup. The volunteers were asked to put one of three emotional marks (positive, neutral, negative) on each fragment. In total, about 700 fragments were tested.

The analysis showed that 93% of the text fragments were correctly assessed by the program. Most often, difficulties arose in the automatic assessment of the tonality of sarcastic sentences, for example, "Wer allein ist, kann nicht verlassen werden" ("... who is alone, he will not be abandoned"), "Wissen macht frei - aber unglücklich" ("Knowledge makes free - but unhappy"), "Die Liebe ist etwas Herrliches. Aber sie verdirbt den Charakter" ("Love is a wonderful thing. But it spoils the character"). Such expressions have a general tone, the opposite of the tone of individual words. In general, the application showed high speed of word processing and flexibility to new, unknown suggestions.

6. Conclusion

Over the past decade, interest in the field of analyzing the emotional sentiment of texts has increased greatly. In teaching a foreign language, an emotional assessment of a literary text can help to understand the author's message and translate without meaningful loss. In the developed application, sentiment analysis is based on the method of using rules. The AFINN library was used as an evaluation vocabulary. The user can assess the tonality of both a single sentence and a fragment of the text. This will allow him to better understand the characteristics of the behavior of the characters, their inner world, the personality of the author of the work and his emotional message. We believe that such an approach to teaching a foreign language can increase the effectiveness of teaching.

7. References

- [1] D. S. Denisova, Analiz emotsional'noy okraski teksta, Sinergiya nauk 19 (2018) 1415–1424.
- [2] T. F. Efremova, New dictionary of the Russian language. Explanatory-word – formation, M.: Russian language, 2000.
- [3] D. A. Gorbushin, D. V. Grinchenkov, V. A. Mokhov, Nguyen Fuk Khau. Sistemnyy analiz podkhodov k resheniyu zadachi identifikatsii tonal'nosti teksta, Izvestiya vuzov, Severo-Kavkazskiy region, Issue: Tekhnicheskkiye nauki 2 (2016).
- [4] V. Hatzivassiloglou, J. M. Wiebe, Effects of adjective orientation and gradability on sentence subjectivity, Proceedings of the 18th conference on Computational linguistics 1 (2000) 299–305.
- [5] V. Hatzivassiloglou, K. R. McKeown, Predicting the semantic orientation of adjectives, Proceedings of the 35th annual meeting of the association for computational linguistics and eighth

- conference of the European chapter of the association for computational linguistics, Association for Computational Linguistics (1997) 174–181.
- [6] How to Build a Speech to Emotion Converter with the Web Speech API and Node.js. [Electronic resource] <https://www.freecodecamp.org/news/speech-to-sentiment-with-chrome-and-nodejs/> (date accessed: 28.01.20).
- [7] R. Kalafato, *Khudozhestvennaya literatura v obuchenii angliyskomu yazyku v Rossii*, *Voprosy obrazovaniya* 2 (2018).
- [8] B. Liu, *Sentiment Analysis and Opinion Mining*, Toronto: Morgan & Claypool Publishers, 2012.
- [9] T. Nasukawa, J. Yi, *Sentiment analysis: Capturing favorability using natural language processing*, *Proceedings of the 2nd international conference on Knowledge capture, ACM* (2003) 70–77.
- [10] A. G. Pazel'skaya, A. H. Solov'yev, *A method of Sentiment Analysis in Russian Texts*, *Computational linguistics and intellectual technologies* (2011) 510–512.
- [11] T. A. Semina, *Tonality of the text: syntactic patterns of expression of relations between entities*, dissertation, 175.
- [12] *Sentiment Analysis Basics* [Electronic resource] <https://monkeylearn.com/sentiment-analysis/#the-basics-of-sentiment-analysis> (date of access: 03/01/2020).
- [13] *Sentiment Analysis with AFINN Lexicon* [Electronic resource] https://medium.com/@himanshu_23732/sentiment-analysis-with-afinn-lexicon-930533dfe75b (date of access: 05.02.20).
- [14] J. M. Wiebe, *Tracking point of view in narrative*, *Computational Linguistics* 20 (1994) 233–287.
- [15] J. M. Wiebe, R. F. Bruce, T. P. O'Hara, *Development and use of a gold - standard data set for subjectivity classifications*, *Proceedings of the 37th annual meeting of the Association for Computational Linguistics* (1999) 246–253.
- [16] S. A. Yermakov, L. M. Yermakova, *Metody otsenki emotsional'noy okraski teksta*, *Vestnik Permskogo universiteta*, Issue: matematika, mekhanika, informatika 1 (2012) 85–90.