




<https://doi.org/10.22363/2687-0088-35624>

EDN: DFZGBB

Research article / Научная статья

The difference in positivity of the Russian and English lexicon: The big data approach

Valery D. SOLOVYEV   and Anna I. IVLEVA 

*Institute of Philology and Intercultural Communication,
Kazan Federal University, Kazan, Russia*
 maki.solovyev@mail.ru

Abstract

Psychological cross-cultural studies have long noted differences in the degree of cognition positivity, or optimism, in various cultures. Herewith, the question whether the difference shows up at the level of the language lexicon remains unexplored. Linguistic positivity bias has been confirmed for a number of languages. The point of it is that most words have a positive connotation in the language. This begs the question: is linguistic positivity bias the same for different languages or not? In a sense, the issue is similar to the hypothesis of linguistic relativity suggesting the language impact on the human cognitive system. The problem has been researched only in one work (Dodds et al. 2015), where data on the positivity bias values are given for different languages and the comparison for each pair of languages is based on merely one pair of dictionaries. In the present study, we radically increase the computational baseline by comparing four English and five Russian dictionaries. We carry out the comparative study both at the level of vocabularies and at the level of texts of different genres. A new, previously untapped idea is to compare positivity ratings of translated texts. Also, English and Russian sentiment dictionaries are compared based on the scores of translation-stable words. The results suggest that the Russian language is somewhat slightly more positive than English at the level of vocabulary.

Keywords: *sentiment, dictionaries, the Pollyanna principle, linguistic relativity, translated texts*

For citation:

Solovyev, Valery D. & Anna I. Ivleva. 2024. The difference in positivity of the Russian and English lexicon: The big data approach. *Russian Journal of Linguistics* 28 (2). 266–293. <https://doi.org/10.22363/2687-0088-35624>

© Valery D. Solovyev & Anna I. Ivleva, 2024



This work is licensed under a Creative Commons Attribution 4.0 International License
<https://creativecommons.org/licenses/by-nc/4.0/legalcode>

Различие позитивности в лексиконах русского и английского языков: подход, основанный на больших данных

В.Д. СОЛОВЬЕВ  , А.И. ИВЛЕЕВА 

Казанский федеральный университет, Казань, Россия

maki.solovyev@mail.ru

Аннотация

В психологических кросс-культурных исследованиях давно замечены различия в степени позитивности мышления, или оптимизма, в различных культурах. Напрашивается вопрос, является ли преимущественная лингвистическая позитивность (*linguistic positivity bias*) одинаковой для разных языков или нет. В определенном смысле этот вопрос схож с гипотезой лингвистической относительности, касающейся влияния языка на когнитивную систему человека. Эта проблема рассматривалась только в одной работе (Dodds et al. 2015), в которой представлены данные о разной величине преимущественной лингвистической позитивности для разных языков и где сравнение для двух языков проводилось с использованием всего одной пары словарей. В настоящем исследовании мы существенно увеличиваем вычислительную базу, сравнивая английский и русский языки на основе 4 английских и 5 русских словарей. Сравнение проводится как на уровне лексикона языков, так и на уровне текстов разных жанров. Новой, ранее не использовавшейся идеей, является сопоставление рейтингов позитивности переводных текстов. Также словари английского и русского языков сравниваются по значениям рейтингов слов, устойчивых к переводу (*translation-stable words*). Результаты позволяют предположить, что на уровне лексикона русский язык несколько более позитивен, чем английский.

Ключевые слова: *тональность, словари, принцип Полианны, лингвистическая относительность, переводные тексты*

Для цитирования:

Solovyev V.D., Ivleva A.I. The difference in positivity of the Russian and English lexicon: The big data approach. *Russian Journal of Linguistics*. 2024. V. 28. № 2. P. 266–293. <https://doi.org/10.22363/2687-0088-35624>

1. Introduction

One of the lines of cross-cultural psychological research is to compare the degree of cognition positivity in different cultures. In (Gallagher et al. 2013), the optimism universality is asserted as a property of the human race. In (Ji et al. 2021), East Asian are noted to engage in less positive thinking than Westerners. Most of studies compare the cultures of these regions. The study (Kirchner-Häusler et al. 2021) concerns Japan and Belgium, while (Ji et al. 2021) is devoted to China and Canada. Russia occupies an intermediate position between the West and the East, but there are very few works dealing with this country. A fairly long-standing work (Kassinove & Sukhodolsky 1995) compares the attitude of Russian and American youth to various current socio-political problems. Russians and Americans are shown to perceive various aspects more or less positively.

The above-mentioned and similar works compare respondents' evaluations of some specific situations. They take into account the impact of a small number of parameters, for instance the standard of living. The impact of the context is emphasized in (Ji et al. 2021). Meanwhile, a different way of assessing the degree of optimism inherent to the culture as a whole is available. The approach is to compare the positivity of the language lexicons of different cultural groups. Lexicon, being an element of culture, reflects the general attitude of native speakers to the world. The dictionaries with positivity/negativity scores for words obtained by respondents survey are the material for the above research. Such dictionaries have long been created within the framework of sentiment analysis.

Sentiment analysis (opinion mining) is an important applied problem that has been studied for several decades (Pang & Lee 2008, Liu 2012). The purpose of sentiment analysis is to automatically determine the positive or negative subjective evaluation of the text or its parts. A classical application of sentiment analysis is evaluation of customer reviews of goods and services (Solovyev & Ivanov 2014). Another significant application is to determine the sentiment of tweets and texts from other social networks to evaluate mood of the society as a whole. For instance, in (Mitchell et al. 2013) they study the way positive sentiment of Twitter users depends on demographics and geography. The latest research (Gower et al. 2023) reports on positively biased social media consumption under the condition of chronic stress during COVID-19. The term “sentiment analysis” is customary in computational linguistics. The term valence is used in psychology for the similar purpose to denote a bipolar “positive-negative” evaluation scale which is applicable both in theoretical and practical psychology research. The concept of valence was introduced by Osgood (1952) within the framework of the concept of semantic differential. In a number of works, it is applied to classify emotions. However, it is suitable for applying to any objects. For example, the database (Warriner et al. 2013) contains the values of this parameter for a large number of English words which obviously do not signify emotions: table, zoom, etc. In the present paper, we use the terms “sentiment” and “valence” as synonyms. The term “score” stands for the numeric value of this parameter.

The term “lemma” is used as is customary to denote the basic (root) form of the word. “Token” means a sequence of characters between two successive spaces. The term “token” is accepted in computational linguistics, it corresponds to the concept of a word form in linguistics. When we use the “word”, one can understand whether it is a token or a lemma depending on context.

In (Kloumann et al. 2012), they shift the research focus from separate texts to the entire language. Most words of the (English) language are shown to have positive connotation on the basis of a large-scale respondents survey. This result seems to be not application-specific but of global significance. For the first time, the idea of the Pollyanna principle was proposed in (Boucher & Osgood 1969) and is also known as linguistic positivity bias (LPB) (Iliev et al. 2016). It was rigorously confirmed by methods of computational linguistics based on universal text corpora,

including Google Books Ngram, etc. It was also checked for specialised corpora: children's and adolescent literature (Jacobs et al. 2020), tweets (Frank et al. 2013), negative customer reviews (Aithal & Tan 2021), etc.

Later, the Pollyanna principle was shown to be valid for 9 other languages (Dodds et al. 2015). Having confirmed the positive bias, one naturally comes to the question whether the principle is similar in different languages or not. The first study of this kind was given in (Dodds et al. 2015). However, there are several controversial points that could affect the result of the study. First, there are different methods of positivity calculation, as well as different dictionaries of sentiment ratings. Besides, the research methodology is not well-established, there are no generally recognised approaches in this field.

In general, the question — which language is more positive — is analogical with the well-known principle of linguistic relativity. In the present study, we are interested in the emotional-evaluative component of the human cognitive system. We put the question: is sentiment evaluation of certain concepts predetermined by cultural traditions only? Does it depend on the language itself as well? Evidently, there are many culture-specific concepts being evaluated by various peoples in essentially different ways. For instance, a pig was a sacred animal in ancient Egypt, while in the Russian language the word *pig* is a swear word. In the Hedonometer (Russian) dictionary discussed below, this word has a rating of 0.32 on a scale of $[0, 1]$, where 1 is the highest positive score.

Nevertheless, the sentiment evaluation of many words hardly depends on the cultural traditions of peoples. At least, it is true for close cultures, such as European ones. Such words as *pencil*, *socks*, *display* are unlikely to cause significantly different sentiment responses. Recently, the authors (Jackson et al. 2019) have drawn attention to some variability of emotions in different cultures. However, they found the sentiment parameter (valence) to be predominant and universal for differentiating emotional and neutral words. This is one of the factors motivating to study sentiment not only for purely applied problems. Thus, we do not ignore the cultural contribution to the sentiment evaluation of concepts in different languages/cultures. It is appropriate to put the question of language systematic impact on the sentiment of vocabulary as a whole, as well as on the emotional-evaluative perception of the world.

In the present study, we suggest some approaches to solve this problem. We compare the positive sentiment bias for the Russian and English languages. This aspect is considerably less studied for the Russian language than for English. So, one of the goals of the present paper is to introduce the data for the Russian language into scholarly discourse. To obtain reliable results, we conduct the comparative study from different aspects. The first point is to compare the languages as such, represented by the sentiment dictionaries. Then we compare the sentiment of texts of various genres in Russian and English.

The research issues covered in our study are:

- (1) How do different Russian and English sentiment dictionaries correlate?
- (2) What is the optimal way of evaluating text sentiment — by taking into account all words of the text or different words only (regardless of word frequency)?
- (3) How does the sentiment of English and Russian texts of different genres correlate? How does the choice of sentiment evaluation parameters affect the results of the comparative study?
- (4) How do the sentiment ratings of the original and translated text correlate?

The structure of the paper is as follows. The second section provides an overview of up-to-date publications on computer verification of the Pollyanna principle. The third section gives a brief description of methods and Russian sentiment dictionaries hardly covered in the English-language literature. The fourth section contains numerous statistical data for comparing positivity of the Russian and English languages, as well as discussions. Finally, the fifth section of the article summarises the results of the study and outlines the prospects for its development.

2. Literature review

Let us note the basic differences between classical applied sentiment analysis and the line of fundamental research of the language and texts sentiment. The latter dates back to (Kloumann et al. 2012). The present paper is devoted to it as well. Both of these research lines rely on sentiment dictionaries. However, there is a considerable difference in their goals and methods. Applied tasks, such as analysis of customer reviews of products, require not only to identify an opinion positivity/negativity, but also to understand which aspect of the product it concerns. This approach implies a very detailed analysis of sentences, including syntactic analysis. It is either impossible or non-relevant for a global evaluation of positivity/negativity of a language as a whole or a large corpus of texts. So, within the line of fundamental research of the language and texts sentiment, the ratio of positive/negative vocabulary is determined only.

Such an overall sentiment analysis was applied not only to the language as a whole, but also to individual texts: lyrical songs, American President's messages to Congress (State of the Union) (Dodds & Danforth 2010), fiction (Dodds et al. 2015) and other diverse types of texts. This research line is based on the use of special dictionaries of positivity/negativity ratings. A large number of such dictionaries have been compiled to date. One can find a detailed review of English dictionaries in (Reagan et al. 2017), and Russian dictionaries in (Kotel'nikov et al. 2020). Usually, such dictionaries are made on the basis of respondents surveys. A number of works study the multidimensional evaluation of texts. Sometimes, in addition to the general positivity/negativity scale, each of the basic emotions is evaluated. For instance, texts evaluation based on the degree of arousal and dominance Osgood factors (Osgood 1952) is studied less thoroughly. Special dictionaries have also

been compiled for these purposes (Warriner et al. 2013). Methods of computational linguistics play an important role in all works (Solovyev et al. 2022), specifically, for preprocessing texts and word form normalization.

Almost all studies use one dictionary only. The exception is (Reagan et al. 2017), where 6 different dictionaries are applied to sentiment analysis of various texts. One should note that the evaluations significantly differ when using various dictionaries. For example, the sentiment evaluation of Society section of New York Times differs by 0.56 on a 9-point scale for ANEW and LabMT dictionaries. For LIWC (Linguistic Inquiry and Word Count (Tausczik & Pennebaker 2014)) and MPQA (Multi-Perspective Question Answering (Yoonjung & Wiebe 2014)) dictionaries it differs by 0.48 on a scale of (-1, 0, +1), i.e. almost a quarter of the scale. The study does not explain which features of dictionaries result in such score differences.

Comparing ratings for different languages presents extra difficulties. The study (Dodds et al. 2015) covers 10 languages. Spanish is shown to be the most positive language, while the most negative language according to this study is Chinese. Also, Russian turned out to be less positive than English. The difference of average score is 0.0263 on a scale of [0, 1]. The main drawback of this research is the application of one dictionary for each language only. The above-mentioned paper (Reagan et al. 2017) shows that using another dictionary can change the evaluation up to 0.56 on a 9-point scale, which corresponds to 0.07 on a scale of [0, 1]. Another dictionary can give a completely different result. So, it is necessary to compare languages on the basis of several dictionaries. There are other disputable points that do not allow us to consider this result as significant.

First of all, positivity evaluations were made for different languages based on different texts. Evidently, texts of various types can have different degrees of positivity, regardless of the language. Moreover, even social networks, being seemingly similar, have different degrees of positivity. For instance, as shown in (Jaidka 2022), Facebook users have a higher level of life satisfaction (positivity of messages) than Instagram users. We assume it is worth comparing the degree of positivity for pairs of translated texts. Due to their equivalence at different levels, the valence score difference is mainly determined by the language.

Secondly, in (Dodds et al. 2015) they use original tokens from texts without lemmatizing them in order to avoid dealing with morphology of various languages. The authors believe this would not strongly affect the result. Probably, it is true for languages with simple morphology, such as English, whereas it can alter the result for languages with complex morphology, such as Russian. For instance, in the Russian language there are 12 inflectional noun forms and several dozen, up to 100, verb forms. Many of inflectional forms of a word have approximately the same frequency. In English, a noun has only two cases. As for verbs, most combinations of tense, aspect, mood and voice are expressed periphrastically, using constructions with auxiliary verbs. It causes the following situation: among the most frequent 5000 words used in (Dodds et al. 2015) there are many forms of the same high-

frequency lemma. That is why many of the first most frequent 5000 lemmas from the frequency dictionary of the language are not taken into account. As a result, when comparing English and Russian, many more different lemmas will be included in calculations for English than for Russian. In the present study, we lemmatise both English and Russian texts.

Various computational approaches can be applied for the research. In the first works (Kloumann et al. 2012, Dodds et al. 2015), text sentiment was evaluated based on dictionary scores of words regardless of their frequency in texts. However, in (Dodds & Danforth 2010) texts sentiment calculations are made by means of the formula $v_{text} = \frac{\sum_{i=1}^n v_i f_i}{\sum_{i=1}^n f_i}$, where v_i is a word score, and f_i is a word frequency, i.e. the frequency of words is taken into account. The same approach is applied in (Hills et al. 2019). This formula summarises the contribution of all the words, both positive and negative. And in (Iliev et al. 2016), they calculate the ratio between the number of positive words and the number of negative ones. If we take the above formula, we get the difference of the corresponding values. In this way, the degree of predominance of positive words over negative ones is determined. We can see different approaches used in different works. There is no common approach yet, as well as systematic comparison of existing ones.

Changes in sentiment evaluation of languages over time has been studied as well. The most frequent data source is the Google Books Ngram corpus (ENA, April 11, 2024)¹.

It is of considerable interest to understand how the use of positive/negative vocabulary changes over time, and what factors can affect these changes. In (Iliev et al. 2016) they show the positive bias decreases over time, and the trend is well approximated by a linear law. The result was obtained for several English corpora only. (Hills et al. 2019) introduced the concept of National Valence Index, calculated according to the above formula. The study was carried out for English, German, and Italian. In both works, they note the correlation of positive bias with the subjectively assessed (in the course of population surveys) happiness level. But at the same time, Gross domestic product does not correlate with National Valence Index in the long run. In (Bochkarev et al. 2023), this problem is studied for the Russian language. High level of dependence on the dictionary is stated. The possibility of direct comparison of the above results for different languages is limited by the fact that each language has its own sentiment dictionaries and the correlation of these dictionaries is unclear.

3. Data and methods

In the present study, we have selected five Russian and four English sentiment dictionaries to eliminate the dependence on a dictionary used. We also focus on the score differences for individual words as well. For each Russian-English pair of

¹ <https://books.google.com/ngrams/>

dictionaries we make sentiment calculation for languages/texts to have twenty alternative options. Among numerous English sentiment dictionaries, we have decided on the most well-known and frequently cited ones: Hedonometer (Dodds et al. 2015), BRM (Warriner et al. 2013), ANEW (Affective Norms for English Words (Bradley & Lang 1999)), NRC-VAD (National Research Council Canada Valence, Arousal, and Dominance Lexicon (Mohammad 2018)). As for the Russian dictionaries, we use such well-known ones as KartaSlovSent (Kulagin 2021), LinisCrowd (Koltsova et al. 2016), Hedonometer (Russian), as well as the KFU² Sentiment human dictionary and the KFU Sentiment BERT machine dictionary that have recently been compiled in our laboratory. BERT stands for Bidirectional Encoder Representations from Transformers, which is an up-to-date method of machine learning developed by Google for the natural language processing. A detailed description of the latter two dictionaries is given in (Solovyev et al. 2022). Formal characteristics of all the dictionaries are given in Table 1 and Table 2.

Table 1. Basic information on Russian sentiment dictionaries

Dictionary	Source	Total number of words	Scale
KFU Sentiment	https://kpfu.ru/tehnologiya-sozdaniya-semanticheskikh-elektronnyh.html	1000	Continuous: [1, 9]
KFU Sentiment BERT	https://kpfu.ru/tehnologiya-sozdaniya-semanticheskikh-elektronnyh.html	25468	Continuous: [1, 9]
KartaSlovSent	https://kartaslov.ru	46127	Discrete: -1, 0, 1
Hedonometer (Russian)	https://hedonometer.org/words/labMT-ru-v2/	9941	Continuous: [1, 9]
LinisCrowd	http://linis-crowd.org/	6860	Discrete: -2, -1, 0, 1, 2

Table 2. Basic information on English sentiment dictionaries

Dictionary	Source	Total number of words	Scale
Hedonometer (English)	https://hedonometer.org/words/labMT-en-v2/	10187	Continuous: [1, 9]
BRM	https://github.com/meadej/twitter-sentiment-analysis?ysclid=1h0bctge6l466946169	13915	Continuous: [1, 9]
ANEW	https://github.com/eriq-augustine/sentiment-data/blob/master/anev.csv	1034	Continuous: [1, 9]
NRC-VAD	https://emilhvitefeldt.github.io/textdata/reference/lexicon_nrc_vad.html	19971	Continuous: [0, 1]

² KFU is abbreviation of the Kazan Federal University

The KFU Sentiment dictionary includes the most frequent words from the dictionary by Lyashevskaya, Sharova³ (2009), the proportions of nouns, adjectives, verbs being equal. The dictionary was obtained through respondents survey on the Yandex.Toloka service. At least 50 scores on a 9-point scale were received for each word. The KFU Sentiment BERT dictionary was derived from the KFU Sentiment dictionary by extrapolating human estimates by the BERT neural network⁴.

The Hedonometer (Russian) dictionary includes most commonly used words from a number of sources: Google Books, New York Times articles, Music Lyrics, Twitter messages translated into Russian. The respondents survey was conducted on Amazon's Mechanical Turk on a 9-point scale. The LinisCrowd dictionary was also created using the crowdsourcing method and is focused on emotional-evaluative words. The KartaSlovSent dictionary is based on the survey method with at least 25 scores for each word. The paper (Kulagin 2021) states that it contains commonly understood words. In fact, it has about a 4-fold predominance of negative words over positive ones, i.e. it is focused on emotional-evaluative words.

In general, the above dictionaries differ significantly both in the method of their compilation and in the set of words included. We compare dictionaries in terms of the scores distribution and consistency of scores between dictionaries. To improve readability and compare the texts evaluations, we normalise all the scores to values on a scale [0, 1] by linear transformation.

Another distinct feature of different dictionaries is that they include either lexemes/lemmas or word forms. We believe that this factor does not impact our results as different word forms of the same lexeme in a dictionary have very proximate valence scores.

The present research background is made up of the following ideas:

- (1) Interlanguage comparison of English and Russian sentiment dictionaries can be based on pairs of words in two languages that correspond to each other as precisely as possible. These are the so-called translation-stable words (Dodds et al. 2015). Their semantics is assumed to be preserved in most cases when translated into another language. That is why we follow the idea from (Dodds et al. 2015) that comparing the scores of translation-stable words from dictionaries, it is possible to determine the language impact on the scores.
- (2) The comparison of either texts or language lexicon can be affected by a dictionary. To eliminate this impact, it is necessary to consider a number of dictionaries. As dictionaries were created by different methods, we expect not to deal with any systematic error that will distort the results. We compare scores for 20 pairs of dictionaries (four English and five Russian dictionaries) created by separate research teams at different times by various methods.

³ <http://dict.ruslang.ru/freq.php>

⁴ Note that the KFU Sentiment BERT dictionary is derived from the KFU Sentiment dictionary by extrapolating scores by using the BERT neural network. It includes the KFU Sentiment dictionary as a subset that is why their data correlate completely (see the next section).

- (3) The central takeaway of all modern approaches is that translation is a kind of cross-linguistic, cross-cultural, and cross-social communication having the following purpose — to establish as much equivalence between the source text and the target text as possible. There are different classifications of translation equivalence levels (see, for instance, (Panou 2013)). But the predominant idea and strategy of any translator is to ensure equivalence at more macroscopic levels (which are called pragmatic, functional, situational, etc. by different scholars and include communication purpose, text style, emotiveness, expressiveness, evaluation aspect, etc.). So, a translator strives for the original and translated texts to be, among other things, emotively equivalent using language units from such sublevels as lexis, syntax, punctuation, etc. Due to it, original and translated texts can serve as a sort of standards, allowing us to evaluate the degree of language lexicon impact on positivity of the whole text by calculating the average positivity shift.

4. Results

4.1. Review and comparison of open Russian and English sentiment dictionaries

As can be seen from Figure 1, the KFU Sentiment dictionary is the most positive both according to the average score and the percentage of positive words. The least positive dictionary by these two criteria is LinisCrowd. The ratio of the number of positive words (with scores > 0.5) to the number of negative words (with scores < 0.5) is 5.85 for KFU Sentiment, 2.94 for KFU Sentiment BERT, 1.08 for KartaSlovSent, 3.34 for Hedonometer (Russian), 0.23 for LinisCrowd. Consequently, all the Russian dictionaries, except for LinisCrowd, confirm the Pollyanna principle. KartaSlovSent dictionary is almost at the boundary of subsets.

Hedonometer dictionary is the most positive in terms of average scores among the English-language dictionaries (Figure 2). The ratio of the number of positive words (with scores > 0.5) to the number of negative words (with scores < 0.5) is 2.34 for Hedonometer (English), 1.26 for BRM, 1.30 for ANEW, 1.06 for NRC-VAD. Consequently, all the English-language dictionaries considered confirm the Pollyanna principle, but the proportion of positive words in Russian-language dictionaries is generally greater than in English ones. NRC-VAD dictionary is almost at the boundary of subsets.

Table 3 and Table 4 give us correlation characteristics for Russian and English sentiment dictionaries respectively. The characteristics are as follows: N is the number of words in the intersection of dictionaries, R_s is a Spearman correlation coefficient for words in the intersection of dictionaries (p -value does not exceed $5 \cdot 10^{-49}$ for all R_s), M_1 and M_2 are average scores for dictionary 1 and dictionary 2 respectively through all words from the intersection of dictionaries. The diagrams embedded in the tables are illustrative, the values on X and Y axes range from 0 to 1.

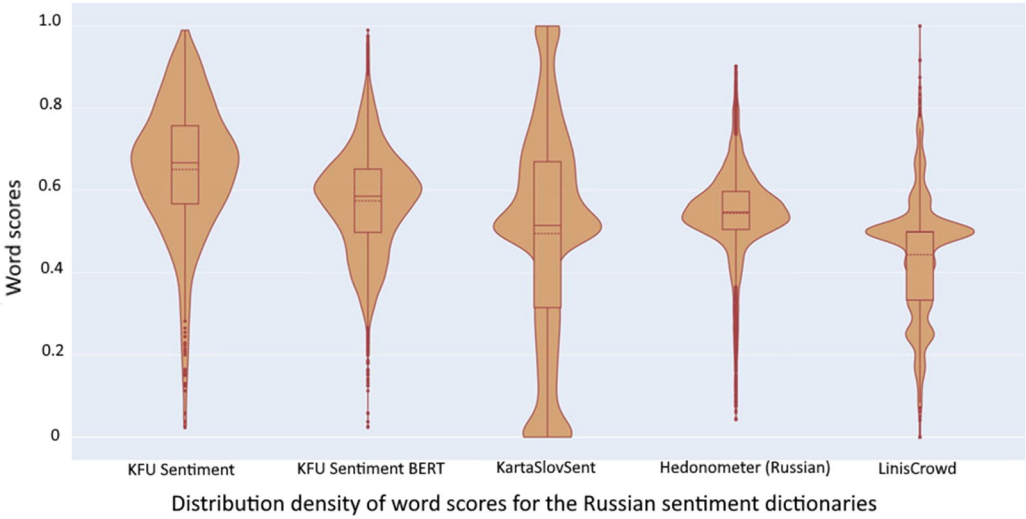


Figure 1. Word scores distributions for the Russian sentiment dictionaries

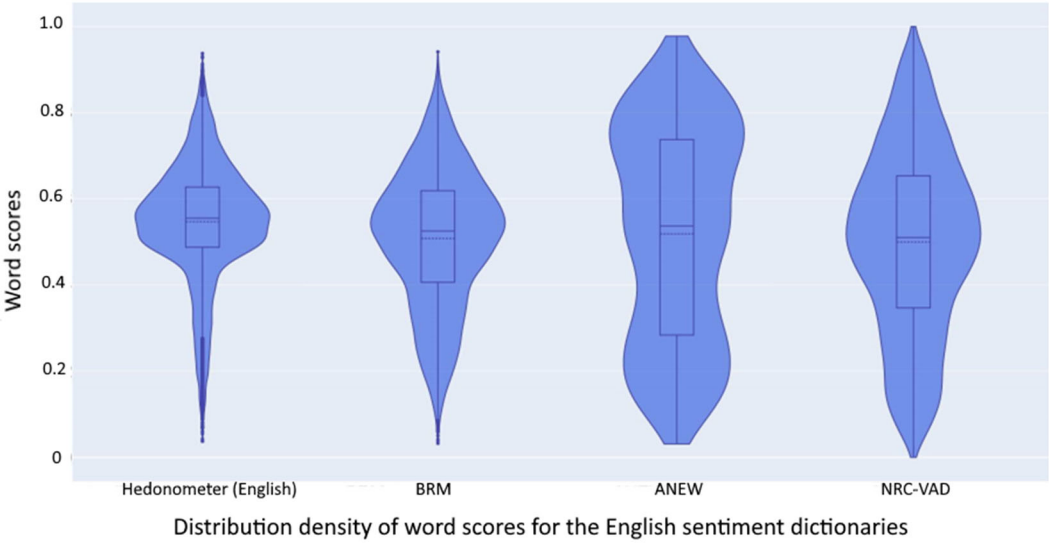


Figure 2. Word scores distributions for the English sentiment dictionaries

In general, the English-language dictionaries correlate much better than the Russian-language ones. This is probably since the methods of compiling Russian dictionaries vary greater. In many cases, the correlation coefficient between the English-language dictionaries exceeds 0.9. For the Russian language, the lowest correlation coefficient is expectedly between the most positive KFU Sentiment and the most negative LinisCrowd dictionary. At the same time, Russian sentiment dictionaries have a better correspondence to the Pollyanna principle.

Table 3. Intersection and correlation of Russian sentiment dictionaries

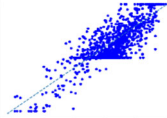
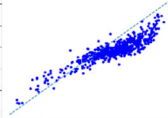

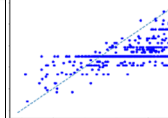
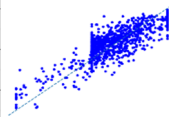
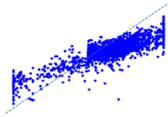
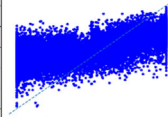
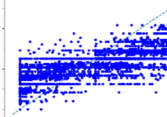
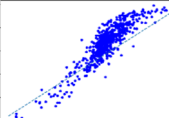
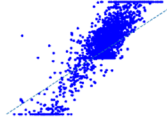
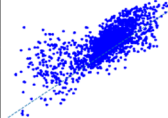


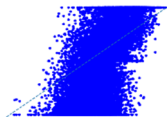
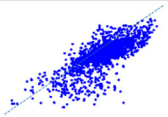
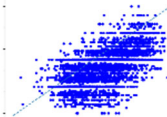
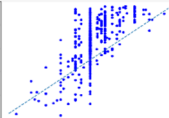
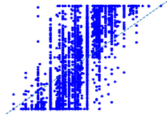
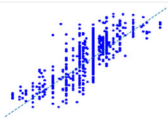
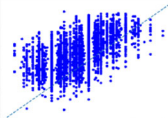
Dictionary 2 \ Dictionary 1	KFU Sentiment	KartaSlovSent	Hedonometer (Russian)	KFU Sentiment BERT	LinisCrowd
KFU Sentiment		N = 967 Rs = 0.755 M ₁ = 0.652 M ₂ = 0.634 	N = 663 Rs = 0.856 M ₁ = 0.663 M ₂ = 0.573 	N = 1000 Rs = 1.000 M ₁ = 0.650 M ₂ = 0.650 	N = 681 Rs = 0.523 M ₁ = 0.651 M ₂ = 0.509 
KartaSlovSent	N = 967 Rs = 0.755 M ₁ = 0.634 M ₂ = 0.652 		N = 2052 Rs = 0.799 M ₁ = 0.609 M ₂ = 0.555 	N = 21498 Rs = 0.636 M ₁ = 0.480 M ₂ = 0.577 	N = 5785 Rs = 0.780 M ₁ = 0.492 M ₂ = 0.449 
Hedonometer (Russian)	N = 663 Rs = 0.856 M ₁ = 0.573 M ₂ = 0.663 	N = 2052 Rs = 0.799 M ₁ = 0.555 M ₂ = 0.609 		N = 2015 Rs = 0.726 M ₁ = 0.552 M ₂ = 0.632 	N = 1309 Rs = 0.620 M ₁ = 0.533 M ₂ = 0.485 
KFU Sentiment BERT	N = 1000 Rs = 1.000 M ₁ = 0.650 M ₂ = 0.650 	N = 21498 Rs = 0.636 M ₁ = 0.577 M ₂ = 0.480 	N = 2015 Rs = 0.726 M ₁ = 0.632 M ₂ = 0.552 		N = 5834 Rs = 0.602 M ₁ = 0.576 M ₂ = 0.444 
LinisCrowd	N = 681 Rs = 0.523 M ₁ = 0.509 M ₂ = 0.651 	N = 5785 Rs = 0.780 M ₁ = 0.449 M ₂ = 0.492 	N = 1309 Rs = 0.620 M ₁ = 0.485 M ₂ = 0.533 	N = 5834 Rs = 0.602 M ₁ = 0.444 M ₂ = 0.576 	

Table 4. Intersection and correlation of English sentiment dictionaries

Dictionary 2 \ Dictionary 1	Hedonometer (English)	BRM	ANEW	NRC-VAD
Hedonometer (English)		N = 4517 Rs = 0.900 M ₁ = 0.557 M ₂ = 0.548	N = 707 Rs = 0.944 M ₁ = 0.551 M ₂ = 0.555	N = 5431 Rs = 0.842 M ₁ = 0.555 M ₂ = 0.561
BRM	N = 4517 Rs = 0.900 M ₁ = 0.548 M ₂ = 0.557		N = 1033 Rs = 0.941 M ₁ = 0.519 M ₂ = 0.509	N = 13867 Rs = 0.792 M ₁ = 0.508 M ₂ = 0.507
ANEW	N = 707 Rs = 0.944 M ₁ = 0.555 M ₂ = 0.551	N = 1033 Rs = 0.941 M ₁ = 0.509 M ₂ = 0.519		N = 1025 Rs = 0.915 M ₁ = 0.518 M ₂ = 0.507
NRC-VAD	N = 5431 Rs = 0.842 M ₁ = 0.561 M ₂ = 0.555	N = 13867 Rs = 0.792 M ₁ = 0.507 M ₂ = 0.508	N = 1025 Rs = 0.915 M ₁ = 0.507 M ₂ = 0.518	

Ten common words having the highest score differences (from the most positive KFU Sentiment and the most negative LinisCrowd) are: *смеяться* (to laugh), *природа* (nature), *герой* (hero), *активный* (active), *смешной* (funny), *результат* (result), *море* (sea), *друг* (friend), *жить* (to live), *стремиться* (to strive). Their rating differences are more than 0.424. The scores of these words in the LinisCrowd dictionary are close to 0.5, while in the KFU Sentiment dictionary they have a fairly positive rating. Totally, 38% of words in LinisCrowd dictionary (including those with emotional connotation) have a score of 0.5. It significantly reduces the correlation between this dictionary and the other ones. For some words,

neutral scores can be explained by ambiguous interpretations of the meaning, so that the respondents' opinions are not unanimous. For example, *смеяться* (*to laugh*): to laugh heartily and laugh at someone; *смешной* (*funny*): funny meaning amusing or ridiculous; *природа* (*nature*): nature surrounding us or the nature of things; *море* (*sea*): the sea as a physical object or as a large number of things. However, for the rest of the 10 listed words, as well as for many other words from the LinisCrowd dictionary, the neutral score is difficult to explain by any other reasons but a small number of respondents and/or the lack of their careful selection and control.

4.2. Comparison of Russian and English sentiment dictionaries

Applying the approach described in (Dodds et al. 2015), we have performed an interlanguage comparison of English and Russian sentiment dictionaries. The comparative study was carried out for translation-stable words from the intersection of dictionaries. The word is considered translation-stable in case the result of machine translation from the original to the target language, and then back to the original language does not change. We used Yandex Translate machine translation. The results are presented in Table 5, where N is the number of common translation-stable words for a pair of dictionaries, R_s is a Spearman correlation coefficient for common translation-stable words (p -values do not exceed $3 \cdot 10^{-24}$ for all R_s), M_1 and M_2 are the average score values in dictionary 1 and dictionary 2 respectively for all the common translation-stable words; $\Delta M = M_1 - M_2$. From this point onward, scores for Russian words are subtracted from scores of English words, i.e. the negative value of ΔM means a greater positivity of the Russian language.

Table 5 shows the score differences between English and Russian (ΔM), which is based on the average scores of translation-stable words (proposed in (Dodds et al. 2015)). ΔM values vary significantly both in absolute value and sign, depending on a pair of dictionaries.

If we calculate the average score difference through all pairs of dictionaries, we get the value of $\Delta M = -0.0183$. However, dictionaries (particularly the Russian-language ones) differ greatly and make different contributions to this value. We try to account for this different contribution in the following way.

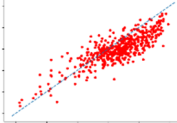
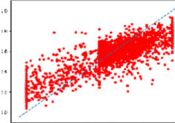
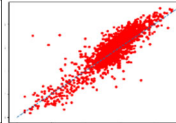
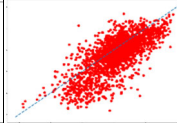
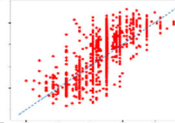
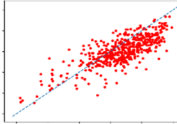
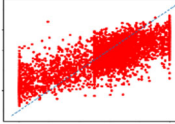
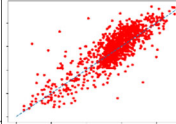
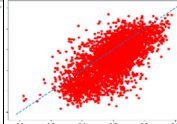
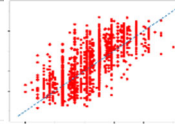
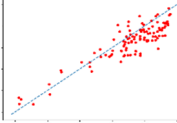
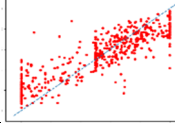
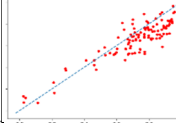
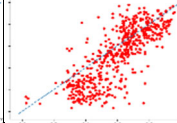
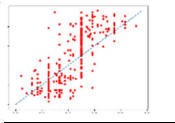
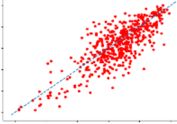
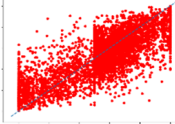
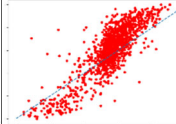
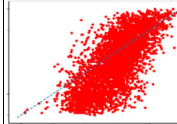
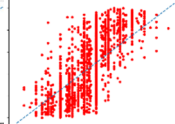
To decide on the ΔM value, we compare sentiment dictionaries based on the two criteria:

- Deviation of the average dictionary score (according to Figures 1 and 2) from the mean of the average scores of all the dictionaries of one language.
- The average correlation of the dictionary scores with the scores of other dictionaries of one language (according to Tables 3 and 4). The correlation of KFU Sentiment BERT and KFU Sentiment dictionaries is not taken into account.

These criteria were normalised by dividing them by the sum of the corresponding criteria values. The first criterion is negative, whereas the second criterion is positive, so they are taken into account with the "-" and "+" sign respectively. As a result, we obtain the following (descending) order of dictionaries

as evaluation alternatives. English dictionaries and their significance weights are: ANEW (1.254), BRM (1.066), NRC-VAD (0.920), Hedonometer (0.760); Russian dictionaries and their significance weights are: Hedonometer (1.205), KFU Sentiment BERT (1.074), KartaSlovSent (1.051), LinisCrowd (0.843), KFU Sentiment (0.828). The weight of each Russian-English pair of dictionaries was determined as the normalised sum of two weights.

Table 5. Intersection and correlation of Russian and English sentiment dictionaries

Dictionary 2 \ Dictionary 1	KFU Sentiment	KartaSlovSent	Hedonometer (Russian)	KFU Sentiment BERT	LinisCrowd
Hedonometer (English)	N = 520 Rs = 0.769 M ₁ = 0.602 M ₂ = 0.667 ΔM = -0.065 	N = 2544 Rs = 0.764 M ₁ = 0.566 M ₂ = 0.585 ΔM = -0.019 	N = 2073 Rs = 0.829 M ₁ = 0.573 M ₂ = 0.549 ΔM = 0.024 	N = 2561 Rs = 0.703 M ₁ = 0.562 M ₂ = 0.616 ΔM = -0.054 	N = 1458 Rs = 0.660 M ₁ = 0.544 M ₂ = 0.483 ΔM = 0.061 
BRM	N = 498 Rs = 0.733 M ₁ = 0.606 M ₂ = 0.668 ΔM = -0.062 	N = 4987 Rs = 0.743 M ₁ = 0.522 M ₂ = 0.539 ΔM = -0.017 	N = 1096 Rs = 0.783 M ₁ = 0.578 M ₂ = 0.556 ΔM = 0.022 	N = 4422 Rs = 0.661 M ₁ = 0.519 M ₂ = 0.597 ΔM = -0.078 	N = 2088 Rs = 0.698 M ₁ = 0.501 M ₂ = 0.460 ΔM = 0.041 
ANEW	N = 112 Rs = 0.782 M ₁ = 0.657 M ₂ = 0.722 ΔM = -0.115 	N = 547 Rs = 0.884 M ₁ = 0.532 M ₂ = 0.516 ΔM = 0.016 	N = 246 Rs = 0.889 M ₁ = 0.594 M ₂ = 0.561 ΔM = 0.033 	N = 558 Rs = 0.786 M ₁ = 0.600 M ₂ = 0.522 ΔM = -0.078 	N = 322 Rs = 0.769 M ₁ = 0.503 M ₂ = 0.443 ΔM = 0.060 
NRC-VAD	N = 505 Rs = 0.736 M ₁ = 0.636 M ₂ = 0.665 ΔM = -0.029 	N = 5895 Rs = 0.776 M ₁ = 0.516 M ₂ = 0.534 ΔM = -0.018 	N = 1274 Rs = 0.793 M ₁ = 0.603 M ₂ = 0.553 ΔM = -0.050 	N = 5044 Rs = 0.692 M ₁ = 0.514 M ₂ = 0.594 ΔM = -0.080 	N = 2253 Rs = 0.733 M ₁ = 0.499 M ₂ = 0.457 ΔM = 0.042 

The ultimate score difference ΔM for all pairs of dictionaries with corresponding significance weights turned out to be equal to -0.0185 . The resulting estimate is quite close to the average difference through all pairs of dictionaries, which is $\Delta M = -0.0183$. So, our estimate points out that Russian words are slightly more positive than English ones. At the same time, the paper (Dodds et al. 2015) gives the value equal to $+0.0263$.

Manual analysis of 520 and 2073 translation-stable words from KFU Sentiment – Hedonometer (English) and Hedonometer (Russian) – Hedonometer (English) with the highest deviations allowed us to single out the following groups of mismatch:

(1) The inadequacy of machine translation. Examples: *очень* is formally translated as *sooo* (“very” would be adequate); *нз* is transliterated as *ng* (meaning “new year” in Russian with unclear meaning in English), etc.

(2) Homonymy in the Russian and English languages: *мисс* — *miss* (the address to women and the verb); *представляет* (the verb meaning “imagines”, “represents”) — *presents* (the noun meaning “gifts”); *камера* (meaning “prison” as well) — *camera*; *кривые* (meaning “incorrect” as well) — *curves*; *напряжение* (referring to current and to human) — *voltage*; *рак* (meaning disease and constellation) — *cancer*; *дорогой* (meaning “important”, “close”) — *expensive*; *вентилятор* — *fan* (about a person as well); *корона* (virus) — *crown* (a thing).

(3) Polysemantic words within one or both languages. The words with semantic fields that only partially intersect in different languages, for example, *выход* — exit (in Russian *выход* also stands for a way out of the situation), *дело* — case (in Russian *дело* is a polysemantic word with more meanings, such as business, activity and so on), *кредит* — credit (the principle meaning of *кредит* is loan), *вспышка* — flash (in Russian *вспышка* is a polysemantic word with more meanings, including outbreak of epidemic).

To eliminate the impact of the above phenomena on the comparative analysis of sentiment evaluation, we manually excluded mismatches of translation-stable words for the pair Hedonometer (Russian) – Hedonometer (English). As a result, the sample decreased from 2022 to 1886 words, the correlation coefficient slightly increased, whereas the average score difference remained almost the same for this pair of dictionaries. So, we can see the sentiment score difference is robust despite inadequate interlanguage correspondences.

In Table 6 and Table 7 we give 10 words with the highest and 10 words with the lowest score difference for NRC-VAD and KartaSlovSent (the pair of dictionaries having the greatest number of translation-stable words). Pairs of words with high score differences belong to one or more of the three groups of mismatch listed above. Words with zero score difference are unambiguous in both languages and have adequate machine translation.

Table 6. A sample of translation-stable word valence in Russian (KartaSlovSent) and English (NRC-VAD) – the highest score differences.

Russian word	Score in KartaSlovSent dictionary	English word	Score in NRC-VAD dictionary	Score difference
раскаяние	0.860	remorse	0.103	0.757
любовник	0.190	lover	0.881	-0.691
зажигательный	0.945	incendiary	0.281	0.664
сдержанность	0.815	restraint	0.167	0.648
дерзость	0.120	audacity	0.760	-0.640
ябеда	0.000	tattletale	0.633	-0.633
напористый	0.695	pushy	0.080	0.615
кропотливый	1.000	painstaking	0.396	0.604
утешение	1.000	consolation	0.408	0.592
гордый	0.340	proud	0.906	-0.566

Table 7. A sample of translation-stable word valence in Russian (KartaSlovSent) and English (NRC-VAD) – the lowest (zero) score differences.

Russian word	Score in KartaSlovSent dictionary	English word	Score in NRC-VAD dictionary
щедрый	1.000	generous	1.000
гибрид	0.500	hybrid	0.500
реалист	0.720	realist	0.720
подсластитель	0.670	sweetener	0.670
трафарет	0.500	stencil	0.500
экваториальный	0.550	equatorial	0.550
обертка	0.625	wrapper	0.625
иностранец	0.500	foreigner	0.500
половина	0.500	half	0.500
бухгалтер	0.625	accountant	0.625

4.3. Comparison of sentiment evaluation of translated texts

The second stage of our research involves comparing texts. Naturally, texts of different genres can contain various words with different frequency. Let us use translated texts to evaluate the score difference between English and Russian, as well as to assess the effect of dictionaries on the evaluation. We have compared sentiment evaluation of 16 English and Russian literary works with their translations. Text preprocessing included removal of non-letter characters, lowering the case, tokenization, lemmatization by pymorphy2.MorphAnalyzer for Russian⁵ and by WordNetLemmatizer for English⁶. Sentiment analysis is carried out regarding words frequency in the text. In addition, following the calculation method from (Dodds et al. 2015), words with relatively neutral scores (> 0.3 and < 0.7) are not taken into account.

In Figure 3 and Figure 4, we can see the medians of distributions of the score difference for translated texts to be close to our previous evaluation from

⁵ <https://pymorphy2.readthedocs.io/en/stable/user/guide.html>

⁶ https://www.nltk.org/_modules/nltk/stem/wordnet.html

section 4.2. of this article (-0.0185). Taking into account the significance weights previously obtained for pairs of dictionaries, we get the difference in average scores, which is -0.0135 . This difference for Russian-language texts and their translations into English is -0.0076 , whereas for English-language texts and their translations into Russian it is -0.0105 . The conclusion is: for literary works, the positivity of the Russian language is slightly greater, although to somewhat lesser extent than it was obtained while comparing dictionaries.

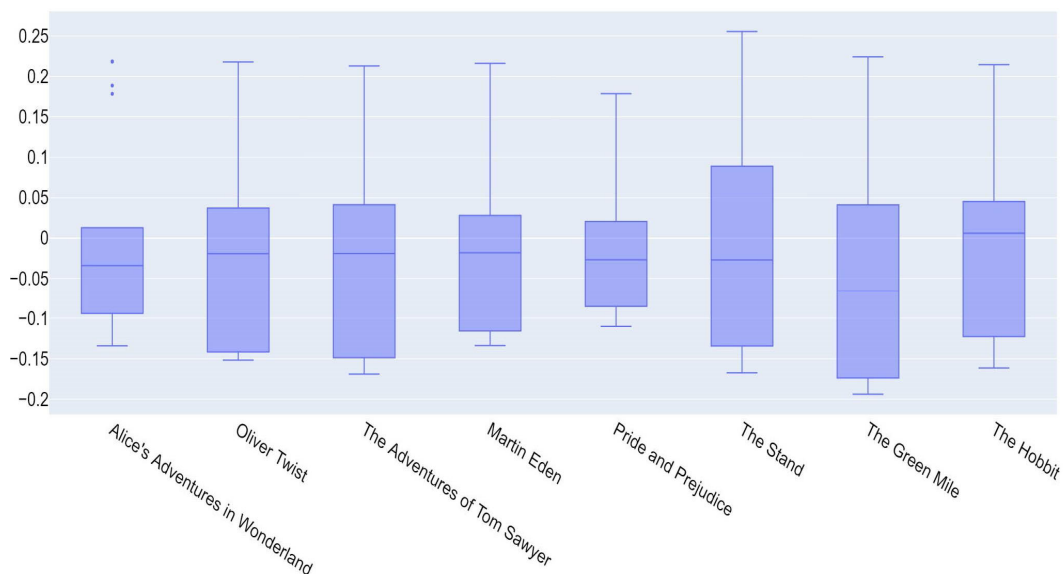


Figure 3. Diagrams of the score difference distributions for English literary works and their translations into Russian (based on 5 Russian and 4 English sentiment dictionaries)

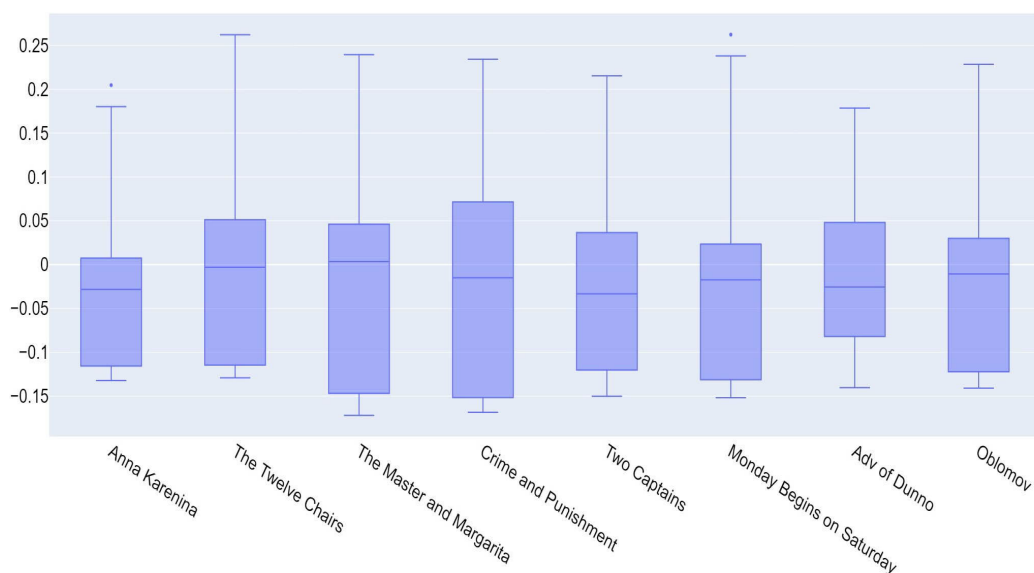


Figure 4. Diagrams of the score difference distributions for Russian literary works and their translations into English (based on 5 Russian and 4 English sentiment dictionaries)

In Figure 5, we can see that all pairs of dictionaries which include LinisCrowd overrate the score difference, i.e. the scores of Russian texts are undervalued. Also, all pairs of dictionaries which include KFU Sentiment and KFU Sentiment BERT undervalue the score difference, i.e. the scores of Russian texts are overestimated. The most exact (close to -0.0185) evaluation of the score difference is obtained by KartaSlovSent in pairs with all English-language dictionaries, except for ANEW. Though KartaSlovSent is one of the most negative Russian dictionaries (by its average score and the number of negative words), it gives the sentiment scores of texts that are neither significantly overvalued nor underestimated.

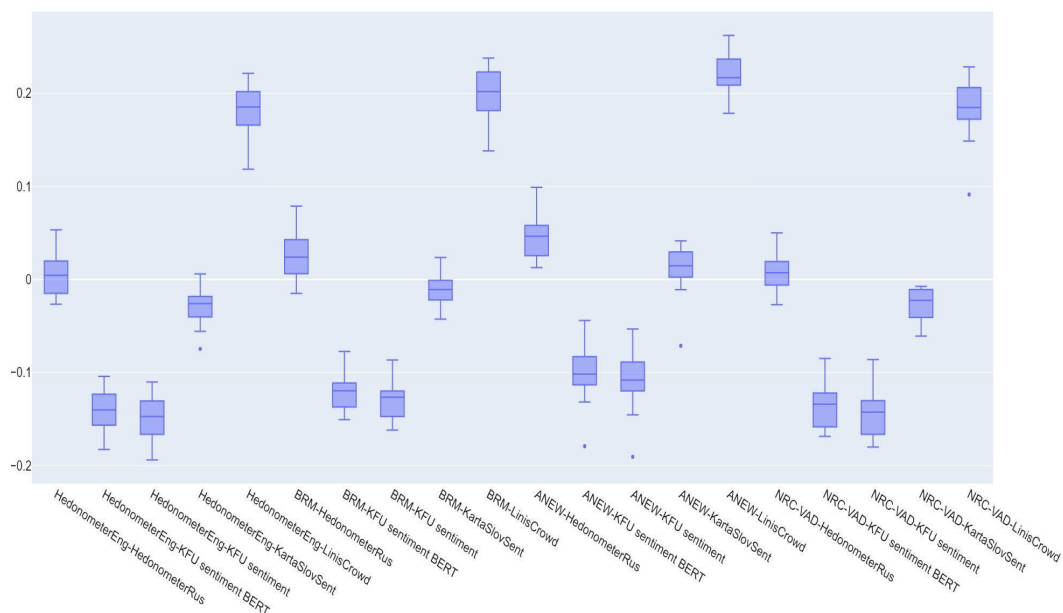


Figure 5. Diagrams of the score difference distributions for the literary works and their translations (based on 20 pairs of dictionaries)

4.4. Two LPB formulations

Let us recall LPB formulation. LPB generally states there are more positive words in the language vocabulary (dictionary) than negative ones. A number of studies are devoted to evaluate emotive vocabulary. For example, in (Tetior 2015) they mention about 150 emotions, and the ratio of negative to positive ones is 2 to 1. In our study, we estimate which words — positive or negative — are used more often in a well-balanced subcorpus of Russian National Corpus. This interpretation is applicable to the evaluation of texts. Let us assume a word is encountered n times in the text. Should we take the word into account once or n times when evaluating the sentiment? If we account for the words as many times as they are encountered, we call it token-approach, otherwise we call it type-approach. When we use the type-approach, a word contributes to the text sentiment evaluation only once, regardless of its frequency. With type-approach, the contribution of a word to the

assessment of the tonality of the text is taken into account only once, regardless of how many times the word has been encountered. The type-approach implies that the contribution of a word sentiment is taken into account only once, regardless of how many times the word has encountered in the text. This terminology corresponds to the well-known TTR (Type Token Ratio) parameter (McKee et al. 2000), which reflects the lexical diversity of the text and is widely used to evaluate the complexity of texts (Solnyshkina et al. 2022). In the first case, we take into account the contribution of the entire text, in the second case — of its vocabulary. The importance of distinguishing these two interpretations of LPB and their independence is stated in (Warriner et al. 2005).

Let us compare the results of these two approaches. The part of the Russian National Corpus, which is in free access⁷ (hereinafter RNC), was taken for this purpose. All texts are divided into sections: fiction, science, public, speech, and blogs. Table 8 shows the data for these RNC subcorpora according to the number of tokens and their types for positive and negative words. Let us consider, for instance, the Public subcorpus. It contains 126477 tokens with scores > 0.5 corresponding 11191 unique types. On average, tokens of one positive type occur 11.3 times. Tokens with scores in the range from 0 to 0.5 occur 18791 times and correspond 3002 types. On average, tokens of one negative type occur 6.3 times. Thus, positive words are used by the authors of texts almost 2 times more often than negative ones. Similar results are obtained for other types of texts, as well as individual texts. This gives the right to mention LPB in the two aspects — there are more positive words in the language, and positive words are encountered more often.

Table 8. Token/type ratio for positive and negative words

Subcorpus	The number of tokens with score > 0.5	The number of types with score > 0.5	Token/type	The number of tokens with score < 0.5	The number of types with score ≤ 0.5	Token/type
Public	126642	11191	11.3	18884	3000	6.3
Fiction	93766	9894	9.5	16351	3012	5.4
Science	108768	8080	13.5	11140	1895	5.9
Speech	91192	7306	12.5	11826	1864	6.3
Blogs	23422	4763	4.9	2896	1134	2.6

We can calculate similar token/type ratio for other frequency ranges. The data for the Public subcorpus are presented in Figure 6. The ratio ranges from 4.6 to 52.2 and is noticeably higher for positive words (> 0.6) than for neutral and negative ones (< 0.4). The most negative words stand apart (< 0.2). Increased frequency of extremely negative words is also noted in (Warriner & Kuperman 2015). So, the LPB hypothesis is confirmed in its two aspects under a more detailed check as well.

⁷ <https://ruscorpora.ru/new/>

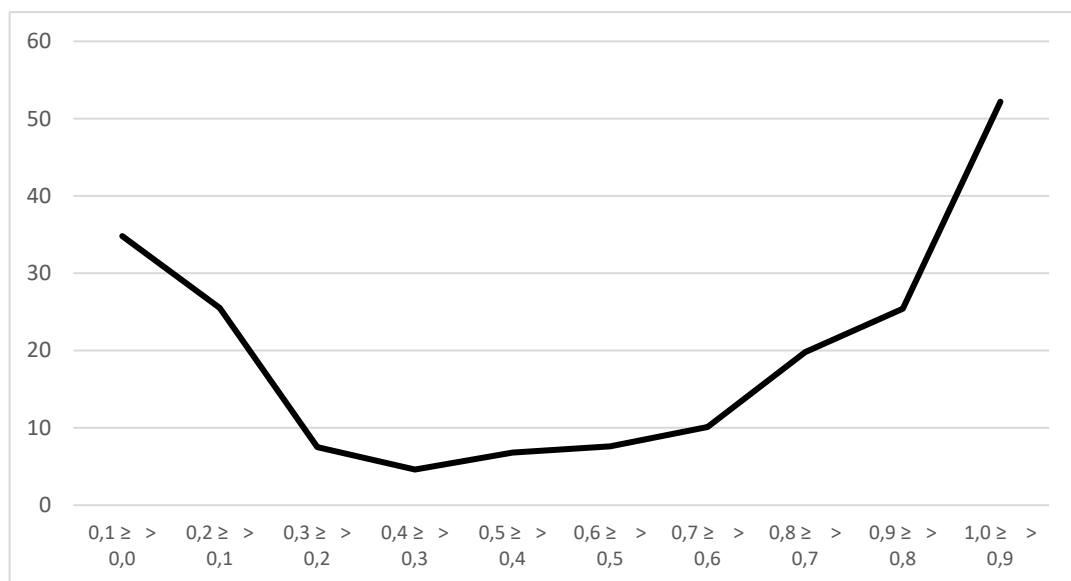


Figure 6. Token/type ratio of positive and negative words having different frequency levels in the Public subcorpus

Both approaches have been considered. Hereinafter, we will explicitly indicate which one is applied.

4.5. Sentiment comparison of English and Russian texts corpora

In this section we compare data on English and Russian corpora of approximately the same subjects. In (Kloumann et al. 2012), the following data are given for the English language: Twitter scores 28.00% of negative words, Google Books collection — 21.20%, New York Times — 21.62%, and Music lyrics collection — 35.86%. Negative words have scores < 0.5 . Type-approach is applied here. For the Russian language we provide data for type-approach as well to make comparative analysis. For the same corpora, the data for token-approach are given in (Solovyev et al. 2022).

We can compare the Public subcorpus with the New York Times subcorpus. English Google Books collection includes various texts — artistic, scientific and journalistic. It can be roughly compared with the Russian fiction, science and public subcorpora. Blogs subcorpus can be compared with English Twitter. Statistical data on Russian subcorpora are given in Table 9. Evaluation of positivity degree was calculated based on the KFU Sentiment BERT dictionary.

Public subcorpus includes almost the same number (0.47% less) of negative words as the English New York Times corpus. The average share of words scored ≤ 0.5 in the Russian fiction, science, public subcorpora is 20.95, which is also almost the same, but still is 0.35% less than in the English Google Books collection.

Table 9. Statistical parameters of sub-corpuses in the Russian National Corpus

Subcorpus	Volume (total number of words in the text)	The number of types (lemmas)	Average value	Median	Share of words scored ≤ 0.5
Public	145526	14193	0.587	0.597	21.15%
Fiction	110117	12806	0.580	0.592	22.74%
Science	119908	9975	0.594	0.604	19.00%
Speech	103018	9170	0.588	0.599	20.33%
Blogs	26318	5897	0.599	0.608	19.23%

There are a lot fewer negative words in the Russian Blogs subcorpus than in Twitter. However, this fact is in good agreement with the data from (Jaidka 2022) on the relatively negative life perception among Twitter users. So, it can concern the users of a particular social network but not the language as a whole.

In general, based on the material of the Russian language, we see the share of words scored ≤ 0.5 does not differ a lot for various genres and is close to 20%. Note these data are obtained by using the KFU Sentiment BERT dictionary.

5. Discussion

In this section, we will discuss the following two points: our results on the greater positivity of the Russian language and validity of the methods applied.

Due to numerous English-language publications like “How Can People Become Happier?” (Folk & Dunn 2023) and the well-known habit of Americans to smile, one can have an impression of a globally positive mood of Anglo-American society. In (Larina & Ponton 2022) they point out to the abundance of English lexemes with positive sentiment. The authors call the feature of communication style formed by such lexemes a demonstrative attractiveness. Specifically, it is formed as a result of the regular use of positive politeness strategies. At the same time, a number of linguistic and cultural publications drew attention to the fact that the Russian language is characterized by negative emotive words, such as *toska* (a sort of longing), which are absent in English (Wierzbicka 1992). However, these publications are based on extremely limited sets of lexemes. In our study, we analyze tens of thousands of words and large text corpora — hundreds of thousands of words. Based on these data, the results we have obtained do not confirm the opinion of more positive thinking and mood of native English speakers compared to Russian speakers. The confidence of our conclusions is also verified by the consistency of the results obtained by different methods based on various data — dictionaries, translated literary works, text corpora.

The above-mentioned direct sociological studies (Kassinove & Sukhodolsky 1995) of the psychological mood of Russian and American youth gave ambiguous results. All these grounds point to a great complexity of the problem. Anyhow, our results obtained by processing large linguistic material by rigorous statistical methods cast doubt on stating greater positivity of the English language as a whole (Dodds et al. 2015).

When comparing original and translated texts, the issue on the influence of a translator on a literary work sentiment arises inevitably. We acknowledge the impact, whereas one should note the process of translation cannot radically change a work sentiment — utopias remain utopias whatever language they are translated into, so do anti-utopias. Later on, we suppose to make supplementary research of the translation impact by comparing translations of the same work by different translators.

The results of this work preliminarily indicate a slight influence of translation on sentiment. The average difference in sentiment of the original and translated literary works is -0.0106 , while the average difference in the sentiment of translation-stable words for different pairs of dictionaries is -0.0185 . In other words, the impact of dictionaries is greater than the influence of translators. One should also note we have analyzed 16 literary works translated by various translators. This fact excludes evaluation bias due to linguistic and translation preferences of a translator.

To compare sentiment dictionaries we use the method of translation-stable words proposed in (Dodds et al. 2015). Currently, this is the only existing method for interlanguage comparison of sentiment of tens and hundreds of thousands of words. Alternatively, we have increased the objectivity of this method by increasing data volume: 20 pairs of dictionaries (four English and five Russian dictionaries) instead of one pair (Dodds et al. 2015). In addition, in contrast to this work, we applied words lemmatization to reduce the structural differences between the Russian and English languages. Also, de facto, it ensures more lemmas to be taken into account.

6. Conclusions

Comparing LPB for different languages is obviously a very difficult problem. On the one hand, translated words do not exactly match. On the other hand, various sentiment dictionaries have been made by different methods and their scores do not coincide. In the present paper, we undertake the first systematic attempt to make interlanguage comparative sentiment research based on the English and Russian languages. We approach the problem from three different aspects.

First, sentiment dictionaries of both languages are considered by comparing scores of translation-stable words. The correlation of scores is quite high, being in the range from 0.7 to 0.9 in most cases. The average score difference between four English and five Russian dictionaries is -0.0185 , i.e. Russian equivalents are more positive. The fact that dictionaries were created by separate research teams at different times by various methods allows us to expect no systematic error. Although there are relatively few translation-stable words, still they give us reference points to be used for interlanguage comparison. In addition, interlanguage sentiment analysis can help assess and improve the quality of text translations and machine translation post-editing. Interlanguage comparison of positivity scores can speed up the search and selection of semantically and emotionally adequate

translation equivalent, can help avoid semantic literalisms and other translation inaccuracies.

The second approach suggests comparing scores for pairs of translated texts. As far as we know, this idea is being implemented for the first time. The data obtained say for the hypothesis of our study: the difference in the positivity of languages (calculated for translation-stable words) matches the score difference between original texts and their correct translation.

A curious pattern has been discovered: the result depends on the direction of translation. If we deal with Russian original text translated into English, the positivity of the Russian text is slightly higher, otherwise the positivity of the Russian texts is more vivid. However, so far this pattern has been obtained for a small collection of 16 literary works, so further larger-scale studies are required.

The third approach is to compare the scores of all words in large text corpora of similar subject (using a pair of dictionaries). Our research shows the degree of texts positivity for similar subject-related corpora is approximately the same (except for social networks), with a slight bias towards greater positivity of Russian texts.

Compared to the work (Dodds et al. 2015), where different languages are considered as well, we have significantly increased the data baseline. In (Dodds et al. 2015) only one dictionary was used for a language. We make calculations using four English and five Russian dictionaries. It allows us to avoid the effect of peculiarities of individual dictionaries. The dictionaries include average scores of dozens of respondents. This allows us to assume they give fair objective words valence.

Thus, all the three approaches applied in our study show a bit higher positivity of the Russian lexicon. Probably, this result reflects some deep psychological patterns inherent to native Russian speakers, their more positive attitude to life and the world around them. Further research is to keep track of increasing the data — the number of languages, dictionaries, translated works. It also seems appropriate to go beyond the valence factor and account for the other Osgood factors. Our article presents a methodology to be applied in our prospective studies.

Disclosure statement

No potential conflict of interest was reported by the authors.

Data availability statement

The data that support the findings of this study are openly available at <https://kpfu.ru/tehnologiya-sozdaniya-semanticheskikh-elektronnyh.html>

Acknowledgment

This paper was supported by the Kazan Federal University Strategic Academic Leadership Program (“PRIORITY-2030”), Strategic Project №5.

References

- Madhusudhan, Aithal & Tan Chenhao. 2021. On positivity bias in negative reviews. <https://arxiv.org/pdf/2106.12056.pdf>
- Bochkarev, Vladimir, Valery Solovyev, Timofei Nestik & Anna Shevlyakova. 2023 Variations in average word valence of Russian books in response to social change over a century. *Proceedings of the Artificial Intelligence and Natural Language Conference. Zap. Nauchn. Sem. POMI* 529. 24–42.
- Bradley, Margaret M. & Peter. J. Lang. 1999. *Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings*. Technical Report C-1, The Center for Research in Psychophysiology, University of Florida.
- Boucher, Jerry & Charles E. Osgood. 1969. The Pollyanna hypothesis. *Journal of Verbal Learning and Verbal Behavior* 8. 1–8. [https://doi.org/10.1016/S0022-5371\(69\)80002-2](https://doi.org/10.1016/S0022-5371(69)80002-2)
- Dodds, Peter Sheridan, Eric M. Clark, Suma Desu, Morgan R. Frank, Andrew J. Reagan, Jake Ryland Williams, Lewis Mitchell, Kameron Decker Harris, Isabel M. Kloumann, James P. Bagrow, Karine Megerdooian, Matthew T. McMahon, Brian F. Tivnan & Christopher M. Danforth. 2015. Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences* 112 (8). 2389–2394. <https://doi.org/10.1073/pnas.1411678112>
- Dodds, Peter Sheridan & Christopher M. Danforth. 2010. Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. *Journal of Happiness Studies* 11. 444–456. <https://doi.org/10.48550/arXiv.1703.09774>
- Dodds, Peter Sheridan, Kameron Decker Harris, Isabel M. Kloumann, Catherine A. Bliss & Christopher M. Danforth. 2011. Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter. *PLoS ONE* 6 (12). e26752. <https://doi.org/10.1371/journal.pone.0026752>
- Folk, Dunigan & Elizabeth Dunn. 2023. How can people become happier? A systematic review of preregistered experiments. *Annual Review of Psychology* 75.
- Frank, Morgan R., Lewis Mitchell, Peter Sheridan Dodds & Christopher M. Danforth. 2013. Happiness and the patterns of life: A study of geolocated tweets. *Scientific Reports* 3. 2625. <https://doi.org/10.1038/srep02625>
- Gallagher, Matthew W., Shane J. Lopez & Sarah D. Pressman. 2013. Optimism is universal: Exploring the presence and benefits of optimism in a representative sample of the world. *Journal of Personality* 81 (5). 429–440. <https://doi.org/10.1111/jopy.12026>
- Gower, Tricia, Kimberly S. Chiew, David Rosenfield & Holly J. Bowen. 2023. Positive biases and psychological functioning during the coronavirus disease 2019 pandemic. *Cognition and Emotion* 37. 1–9. <https://doi.org/10.1080/02699931.2023.2221022>
- Hills, Thomas T., Eugenio Proto, Daniel Sgroi & Chanuki Illushka Seresinhe. 2019. Historical analysis of national subjective wellbeing using millions of digitized books. *Nature Human Behaviour* 3 (12). 1271–1275. <https://doi.org/10.1038/s41562-019-0750-z>
- Iliev, Rumen, Joe Hoover, Morteza Dehghani & Robert Axelrod. 2016. Linguistic positivity in historical texts reflects dynamic environmental and psychological factors. *Proceedings of the National Academy of Sciences* 113 (49). E7871–E7879. <https://doi.org/10.1073/pnas.1612058113>
- Jackson, Joshua Conrad, Joseph Watts, Teague R. Henry, Johann. M. List, Robert Forkel, Peter Mucha, Simon J. Greenhill, Russell D. Gray & Kristen A. Lindquist. 2019. Emotion semantics show both cultural variation and universal structure. *Science* 366 (6472). 1517–1522. <https://doi.org/10.1126/science.aaw8160>
- Jacobs, Arthur M., Berenike Herrmann, Gerhard Lauer, Jana Lüdtkke & S. Schroeder. 2020. Sentiment analysis of children and youth literature: Is there a Pollyanna effect? *Frontiers in Psychology* 11. <https://doi.org/10.3389/fpsyg.2020.574746>

- Jaidka, Kokil. 2022. Cross-platform-and subgroup-differences in the well-being effects of Twitter, Instagram, and Facebook in the United States. *Scientific Reports* 12 (1). 3271. <https://doi.org/10.1038/s41598-022-07219-y>
- Ji, Li-Jun, Thomas. I. Vaughan-Johnston, Zhiyong Zhang, Jill A. Jacobson, Ning Zhang & Xiaoye Huang. 2021. Contextual and cultural differences in positive thinking. *Journal of Cross-Cultural Psychology* 52 (5). 449–467. <https://doi.org/10.1177/00220221211020442>
- Kassinove, Howard & Denis G. Sukhodolsky. 1995. Optimism, pessimism and worry in Russian and American children and adolescents. *Journal of Social Behavior & Personality* 10 (1). 157–168.
- Kay, Paul & Chad K. McDaniel. 1978. The linguistic significance of meanings of basic color terms. *Language* 54 (3). 610–646. <https://doi.org/10.2307/412789>
- Kirchner-Häusler, Alexander, Michael Boiger, Yukiko Uchida, Yoko Higuchi, A. Uchida & Batja Mesquita. 2022. Relatively happy: The role of the positive-to-negative affect ratio in Japanese and Belgian couples. *Journal of Cross-Cultural Psychology* 53 (1). 66–86. <https://doi.org/10.3389/fpsyg.2020.01048>
- Kloumann, Isabel M., Christopher M. Danforth, Kameron Decker Harris, Catherine A. Bliss & Peter Sheridan Dodds. 2012. Positivity of the English Language. *PLoS ONE* 7 (1). e29484. <https://doi.org/10.48550/arXiv.1108.5192>
- Koltsova, Olesya Yu., Svetlana V. Alexeeva & Sergey N. Kolcov. 2016. An opinion word lexicon and a training dataset for Russian sentiment analysis of social media. *Computational Linguistics and Intellectual Technologies — Proceedings of the International Conference “Dialog”* 277–287.
- Kotel’nikov, Evgeniy V., Elena V. Razova, Anastasiya V. Kotelnikova & Sergey V. Vychezhnanin. 2020. Modern sentiment lexicons for opinion mining in English and Russian (analytical survey). *Informacionnye Processy i Sistemy* 12. 16–33.
- Kulagin, Denis I. 2021. Publicly available sentiment dictionary for the Russian language KartaSlovSent. *Computational Linguistics and Intellectual Technologies — Proceedings of the International Conference “Dialog”* 20. 1106–1119.
- Kušen, Ema, Mark Strembeck & Mauro Conti. 2019. Emotional valence shifts and user behavior on Twitter, Facebook, and YouTube. *Influence and Behavior Analysis in Social Networks and Social Media*. 63–83. https://doi.org/10.1007/978-3-030-02592-2_4
- Larina, Tatiana & Douglas Mark Ponton. 2022. I wanted to honour your journal, and you spat in my face: emotive (im) politeness and face in the English and Russian blind peer review. *Journal of Politeness Research* 18 (1). 201–226.
- Liu, Bing. 2012. *Sentiment Analysis and Opinion Mining*. Springer.
- McKee, Gerard T., David D. Malvern & Brian James Richards. 2000. Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing* 15 (3). 323–337. <https://doi.org/10.1093/lc/15.3.323>
- Mitchell, Lewis, Kameron Decker Harris, Morgan R. Frank, Peter Sheridan Dodds & Christopher M. Danforth. 2013. The geography of happiness: Connecting Twitter sentiment and expression, demographics, and objective characteristics of place. *PLoS ONE* 8 (5). e64417. <https://doi.org/10.48550/arXiv.1302.3299>
- Mohammad, Saif M. 2018. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 174–184. <https://doi.org/10.18653/v1/P18-1017>
- Osgood, Charles E. 1952. The nature and measurement of meaning. *Psychological Bulletin* 49. 197–237.
- Pang, Bo & Lillian Lee. 2008. Opinion mining and sentiment analysis (English). *Foundations and Trends in Information Retrieval* 2. 1–135. <https://doi.org/10.1561/1500000011>

- Panou, Despoina. 2013. Equivalence in translation theories: A critical evaluation. *Theory and Practice in Language Studies* 3 (1). 1–6. <https://doi.org/10.4304/tpls.3.1.1-6>
- Reagan, Andrew J., Christopher M. Danforth, Brian F. Tivnan, Jake Ryland Williams & Peter Sheridan Dodds. 2017. Sentiment analysis methods for understanding large-scale texts: A case for using continuum-scored words and word shift graphs. *EPJ Data Science* 6. 1–21. <https://doi.org/10.1140/epjds/s13688-017-0121-9>
- Solnyshkina, Marina I., Valery D. Solovyev, Elzara V. Gafiyatova & Ekaterina V. Martynova. 2023. Text complexity as interdisciplinary problem. *Voprosy Kognitivnoy Lingvistiki* 1. 18–39. <https://doi.org/10.20916/1812-3228-2022-1-18-39>
- Solovyev, Valery, Musa Islamov & Venera Bayrasheva. 2022. Dictionary with the evaluation of positivity/negativity degree of the Russian words. In S. R. Mahadeva Prasanna, Alexey Karpov, K. Samudra Vijaya & Shyam S. Agrawal (eds.), *Speech and computer. SPECOM 2022. Lecture notes in computer science, 13721*, 651–664. Springer.
- Solovyev, Valery, Vladimir Ivanov. 2014. Dictionary-based problem phrase extraction from user reviews. In Petr Sojka, Alex Horák, Ivan Kopeček & Karel Pala (eds.), *Text, speech and dialogue. TSD 2014. Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in Bioinformatics), LNAI, 8655*, 225–232. Springer.
- Solovyev, Valery D., Marina I. Solnyshkina & Danielle S. McNamara. 2022. Computational linguistics and discourse complexology: Paradigms and research methods. *Russian Journal of Linguistics* 26 (2). 275–316. <https://doi.org/10.22363/2687-0088-31326>
- Tausczik, Yla R. & James W. Pennebaker. 2014. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology* 29 (1). 24–54. <https://doi.org/10.1177/0261927X09351676>
- Tetior, Alexander N. 2015. The emotional sphere of a person: The predominance of negative emotions. *Eurasian Union of Scientists* 2 (11). 78–81.
- Warriner, Amy Beth, Victor Kuperman & Marc Brysbaert. 2013. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods* 45. 1191–1207. <https://doi.org/10.3758/s13428-012-0314-x>
- Warriner, Amy Beth & Victor Kuperman. 2015. Affective biases in English are bi-dimensional. *Cognition and Emotion* 29 (7). 1147–1167. <https://doi.org/10.1080/02699931.2014.968098>
- Whorf, Benjamin Lee. 2012. *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*. In John B. Carroll, Stephen C. Levinson & Penny Lee (eds.). The MIT Press.
- Wierzbicka, Anna. 1992. The Russian language. *Semantics, Culture and Cognition: Universal Human Concepts in Culture-specific Configurations*. 395–441. New York: Oxford University Press.
- Yoonjung, Choi & Wiebe Janyce. 2014. +/-EffectWordNet: Sense-level lexicon acquisition for opinion inference. *Proc. of EMNLP*. 1181–1191. <https://doi.org/10.3115/v1/D14-1125>

Dictionaries

- Hedonometer (English). Retrieved from <https://hedonometer.org/words/labMT-en-v2/> (accessed 18 March 2024).
- BRM. Retrieved from <https://github.com/meadej/twitter-sentiment-analysis?ysclid=lh0bctge61466946169> (accessed 18 March 2024).
- ANEW. Retrieved from <https://github.com/eriq-augustine/sentiment-data/blob/master/new.csv> (accessed 18 March 2024).
- NRC-VAD. Retrieved from https://emilhvitefeldt.github.io/textdata/reference/lexicon_nrc_vad.html (accessed 18 March 2024).
- KFU Sentiment. Retrieved from <https://kpfu.ru/tehnologiya-sozdaniya-semanticheskikh-elektronnyh.html> (accessed 18 March 2024).

KFU Sentiment BERT. Retrieved from <https://kpfu.ru/tehnologiya-sozdaniya-semanticheskikh-elektronnyh.html> (accessed 18 March 2024).

KartaSlovSent. Retrieved from <https://kartaslov.ru> (accessed 18 March 2024).

Hedonometer (Russian). Retrieved from <https://hedonometer.org/words/labMT-ru-v2/> (accessed 18 March 2024).

LinisCrowd. Retrieved from <http://linis-crowd.org/> (accessed 18 March 2024).

Article history:

Received: 07 August 2023

Accepted: 20 March 2024

Bionotes:

Valery D. SOLOVYEV is Doctor Habil. of Physical and Mathematical Sciences, Professor, Chief Researcher of “Text Analytics” Research Lab, Institute of Philology and Intercultural Communication of Kazan Federal University, Kazan, Russia. He is a member of the Presidium of the Interregional Association for Cognitive Research, author of four monographs and more than 60 publications on the computational linguistics.

e-mail: maki.solovyev@mail.ru

<https://orcid.org/0000-0003-4692-2564>

Anna I. IVLEVA is Ph.D. in Engineering Sciences and Senior Researcher of “Linguistics and AI” Research Lab, Institute of Philology and Intercultural Communication of Kazan Federal University, Kazan, Russia. The main areas of her research interests are quantitative linguistics, natural language processing and translation studies.

e-mail: ivleva.anna.igorevna@yandex.ru

<https://orcid.org/0000-0002-2670-6795>

Сведения об авторах:

Валерий Дмитриевич СОЛОВЬЕВ – доктор физико-математических наук, профессор, главный научный сотрудник НИЛ «Текстовая аналитика» Института филологии и межкультурной коммуникации Казанского федерального университета, Казань, Россия. Член президиума Межрегиональной ассоциации когнитивных исследований. Автор четырех монографий и более 60 публикаций по компьютерной лингвистике.

e-mail: maki.solovyev@mail.ru

<https://orcid.org/0000-0003-4692-2564>

Анна Игоревна ИВЛЕВА – кандидат технических наук, старший научный сотрудник НИЛ «Лингвистика и искусственный интеллект» Института филологии и межкультурной коммуникации Казанского федерального университета, Казань, Россия. Основные сферы научных интересов: квантитативная лингвистика, обработка естественного языка, теория перевода.

e-mail: ivleva.anna.igorevna@yandex.ru

<https://orcid.org/0000-0002-2670-6795>