

5. Никитина В.В. Изучение профориентации абитуриентов при поступлении в вуз по программам целевого обучения // Современные проблемы науки и образования. – 2015. – № 3.

ПРЕДСКАЗАНИЕ ЭМОЦИОНАЛЬНОГО ВЕКТОРА ПРИЛАГАТЕЛЬНЫХ РУССКОГО ЯЗЫКА НА ОСНОВЕ МОДЕЛИ ДИСТРИБУТИВНОЙ СЕМАНТИКИ.

Мещанкин Андрей Вячеславович

*Аспирант кафедры педагогической психологии
Московского Городского Педагогического Университета,
г. Москва*

[DOI: 10.31618/ESU.2413-9335.2019.7.60.60-64](https://doi.org/10.31618/ESU.2413-9335.2019.7.60.60-64)

АННОТАЦИЯ.

Знание об эмоциональном содержании слов важно, как для эмоционального анализа текста, так и для психолингвистических основ диалоговых систем. Эмоциональная разметка слов представляет собой довольно сложную и дорогую процедуру. Чаще всего перед группой людей ставится задача обозначить соответствующее слово одной из предложенных эмоций, которая, на их взгляд, ассоциируется с ним. Эта крайне трудоемкая и время затратная процедура. Целью данной работы является автоматическое распространения информации об эмоциональном содержании слов на основе подходов дистрибутивной семантики.

ANNOTATION.

Knowledge of the emotional content of words is important, both for the emotional analysis of the text, and for the psycholinguistic foundations of dialogue systems. Emotional marking of words is a rather complicated and expensive procedure. Most often, a group of people asked for identifying the appropriate word with one of the proposed emotions, which, in their opinion, is associated with it. This is an extremely hard and time-consuming procedure. The purpose of this work is to make algorithm to automatically distribute information about the emotional content of words based on the approaches of distributional semantics.

Ключевые слова: психолингвистика, базовые эмоции, дистрибутивная семантика, распространение эмоций, эмоции в прилагательных

Keywords: psycholinguistics, basic emotions, distributive semantics, the spread of emotions, emotions in adjectives

1. Введение и новизна.

В последние десятилетия, одной из основных целей психолингвистики является объяснение того, как слова, которые используют люди в своей повседневной жизни, отражают то, кто они такие, чем они занимаются. Благодаря обширному распространению компьютеров и возможности общаться в социальных сетях, специалисты получили доступ к большому количеству данных, представленных естественным языком, что позволяет исследователям значительно расширить фронт работ по поиску взаимосвязей между языком и личностью.

Работа, проводимая в рамках подходов психолингвистики, имеют большое значение в психологических исследованиях.

Исходя из работ, посвященных выделению так называемых дескрипторов (слов, явно отражающих ту или иную эмоцию), а также последними исследованиями в области психолингвистики, известно, что наибольшая эмоциональная составляющая передается в письменной речи посредством прилагательных [11].

Многие исследования были сосредоточены на сборе рейтингов, связанных с такими свойствами слов, как частота, сложность, конкретность, образность, возраст, приобретение, знакомство и аффективные состояния [1,2]. Сбор рейтинга предполагает привлечение большого количества респондентов (носителей данного языка), которым предлагалось определить эмоциональную окраску слов.

Эти рейтинги в психолингвистике могут быть использованы для количественной оценки различных свойств большого объема текста на естественном языке, например, при анализе лексического выбора между демографическими группами [3] или предпочтений в музыке [4].

Основными единицами в построении эмоционального рейтинга слов являются валентность (разные эмоции) и возбуждение (аффект – интенсивность эмоции). Используя данные единицы, мы имеем несколько более расширенную модель циркумплекса Рассела [5], который утверждает, что все состояния представлены в виде линейной комбинации этих двух независимых величин. Тем не менее, сбор рейтинга слов занимает очень длительное время и является дорогой по стоимости процедурой. Кроме того, хотя сбор рейтинга происходит под контролем, часто опрашиваемые люди дают абсолютно противоположные оценки, что свидетельствует о различном предыдущем личностном опыте. Независимое автоматическое построение подобных рейтингов стало единственным выходом, обеспечивающим исследователей данными с достаточной полнотой и точностью. Автоматический подход к построению эмоциональных рейтингов слов был изучен на основании интуитивного предположения, что слова с одинаковой семантической дистрибуцией будут иметь схожие значения рейтинга.

Часто рейтинг неизвестного слова вычисляют как среднее соответствующих величин его k -ближайших соседей из низкоразмерного семантического пространства [6]. Однако, недостатком такого подхода, является то, что антонимы также семантически похожи, что, как ожидается, уменьшает точность этих методов. Орфографическое сходство показало немного лучшие результаты [7].

Научная новизна нашего подхода автоматического предсказания эмоционального вектора прилагательных заключается в использовании методов дистрибутивной семантики и анализа графов, что позволяет избежать вышеуказанных трудностей. Такой подход впервые использован для русского языка. Нами разработан программный комплекс для эмоциональной разметки прилагательных русского языка, позволяющий в автоматическом режиме рассчитывать эмоциональный вектор слов.

2. Методика исследований

2.1 Сбор Данных.

Наш рейтинг представляет собой сопоставление слов определенным эмоциям вместе с количественной мерой аффекта. Данные собраны при помощи онлайн-опросника, который предлагался для прохождения добровольцам. Опрос давал возможность выбора для каждого слова одной из 7 эмоций (гнев, предвосхищение, отвращение, страх, радость, удивление, доверие). Словами в опросе являлись различные прилагательные русского языка. Алгоритм предлагал для разметки наиболее частотные в употреблении естественной речи прилагательные таким образом, чтобы каждое слово имело возможность получить статистически значимое значение эмоции, либо признавалось эмоционально неопределенным. Таким образом, каждое слово получало 7-ми мерный вектор эмоций.

2.2 Методология анализа.

Метод состоит из двух шагов. Первый шаг заключается в применении модели смысловой близости слов, обученной на большом корпусе текстов естественного русского языка. Дистрибутивная гипотеза, использованная для построения модели, заключается в том, что смысловая близость между двумя словами может быть построена исходя из частоты появления данных слов с одинаковым лексическим окружением.

Во втором шаге мы предполагаем, что эмоциональный контекст слова так же, как и смысл слова,

может быть рассчитан на основании дистрибутивной гипотезы.

Следовательно, имея контрольную группу и заранее размеченных добровольцами слов, мы можем автоматически распространить эмоциональный вектор на близкие им неразмеченные слова.

2.3 Дистрибутивная семантика.

Основной идеей дистрибутивной семантики является так называемая дистрибутивная гипотеза, состоящая в том, что о грамматических и семантических свойствах языковой единицы (морфемы или лексемы) можно многое узнать из её окружения другими языковыми единицами в тексте, не привлекая никаких дополнительных сведений [8].

Общепринятым в лингвистике является следующий способ: используя корпуса текстов как можно большего размера собираются сведения о совместной дистрибуции слов в одинаковом окружении, далее каждому слову присваивается вектор w , являющийся вектором частот разных слов в его контекстных окружениях заданной ширины. Частоты могут быть помноженными на веса, приписанные словам-признакам (элементам базиса) в соответствии с тем, насколько тесна их связь с w . Величина семантической близости между w и произвольным словом w' вычисляется как расстояние в какой-либо метрике между векторами, описывающими их [8].

Психолингвистические исследования показали, что данный метод показывает хороший результат при оценке синонимичных слов. В работе [12] была использована модель близости слов, составленная на основании корпуса из текстов. Далее добровольцев просили определить синонимы, для некоторой подгруппы слов. После того, как синонимы были названы, результат сравнивался с синонимами, найденными моделью. Данные показали хорошую сходимость. Среди прочих таким путём идут Л. Ли [9], Дж. Уидс [10] и Дж. Каррен [11], в чьих работах исследовано поведение целого ряда весовых функций и метрик относительно эталонов.

3. Результаты эксперимента и их анализ

Нами был собран эмоциональный рейтинг слов на основании ответов 116 респондентов. В таблице 1 представлены результаты работы одного респондента из общей выборки.

Таблица 1. Экспериментальные данные

Слово	гнев	предвосхищение	отвращение	страх	радость	удивление	доверие
модный		1					
расширенный					1		
лимфатический						1	
плачущий		1					
модный		1					
бестолковый		1					
альтернативный				1			

Как правило, ответы респондентов, отражали их личностную позицию, с учетом их предыдущего опыта. Разные респонденты по-разному размечали одни и те же предложенные слова. После этого для

каждого уникального слова, суммировалась разметка от всех респондентов. В большинстве случаев была найдена доминирующая для респондентов эмоция, характеризующая данное слово. Пример такого выбора представлен в табл. 2.

Таблица 2. Результаты сопоставления слов с доминирующими выборами эмоций на основании опроса 116 респондентов

Слово	гнев	предвосхищение	отвращение	страх	радость	удивление	доверие
оживленный					7		
невзрачный				3	5		
вооруженный				1	4		
живой					3		
молодой					3		1
новый					3		1
прекрасный					3		

В табл. 3 представлена ситуация, когда при суммарной эмоциональной разметки наблюдались несколько равноправных эмоций.

Таблица 3. Результаты сопоставления слов с равновероятным выбором эмоций на основании опроса 116 респондентов

Слово	гнев	предвосхищение	отвращение	страх	радость	удивление	доверие
водный					3	2	3

Это можно связать как с недостаточным количеством респондентов, так и с отсутствием единой эмоции для характеристики слова. Поэтому мы используем в нашем рейтинге понятие «эмоциональный вектор», а не конкретную эмоцию, что позволяет не упустить малых в абсолютных величинах, но значимых для алгоритма распространения, значений аффекта эмоции.

Распространение эмоционального вектора между словами происходило по следующему алгоритму:

1. С помощью заранее обученной модели для размеченного респондентами слова находились его ближайшие смысловые соседи
2. Найденные слова проверялось на соответствие всех лексических правил исходному

3. При соответствии лексических правил, эмоциональный вектор передавался как его произведение на скалярную величину, которой являлась мера близости слов согласно п.1

4. Для слова, которому был спрогнозирован эмоциональный вектор, записывалось количество ссылающихся на него изначально размеченных слов.

5. Если слово, на которое ссылались первоначально размеченное, уже имело эмоциональный вектор, то он поэлементно складывался с предыдущим.

6. После завершения всего цикла распространения, предсказанный вектор эмоций, поэлементно был разделен на количество ссылок из п.4

На первом этапе мы получили таблицу из 200 слов, с эмоциональными векторами по нашему алгоритму (табл. 4).

Таблица 4. Пример результата автоматического прогнозирования сопоставления слов с эмоциями.

Слово\ эмоция	Гнев	Предвосхищение	Отвращение	Страх	Радость	Удивление	Доверие
британский	0,37	0	0	0	0,37	0,37	0
французский	0	0	0	0	0	0,79	0
кремовый	0	0,37	0,77	0	0,41	0	0
чернявый	0	0	0,84	0	0	0,84	0,84
черный	0	0	0	0	0	0,76	0,76
безбрежный	0	0	0,81	0	0	0	0
меньший	0	0,86	0	0	0	0,86	0
небольшой	0	0	0	0,71	0	0,71	0,71
жизнерадостный	0,25	0	0	0	0,78	0	0
вероятный	0	0	0,77	0	0,77	0	0
субботный	0	0	0,76	0	0,76	0	0
невысокий	0	0	0,78	0	0,78	0,78	0,78
квалифицированный	0	0,80	0	0	0	0	0,79

Представленные в таблице коэффициенты являются мерой аффекта (интенсивности) эмоции. И вычисляются как отношение суммарного предсказанного коэффициента к количеству ссылающихся на слово изначально размеченных слов.

Для наглядности мы построили график плотности однозначности выбора эмоций (рис. 1), где по

оси Y отложено количество эмоций, определенных для каждой десятки слов, а по оси X – номер каждой десятки слов. Причем, декады отсортированы в порядке убывания суммарного количества входящих ссылок.

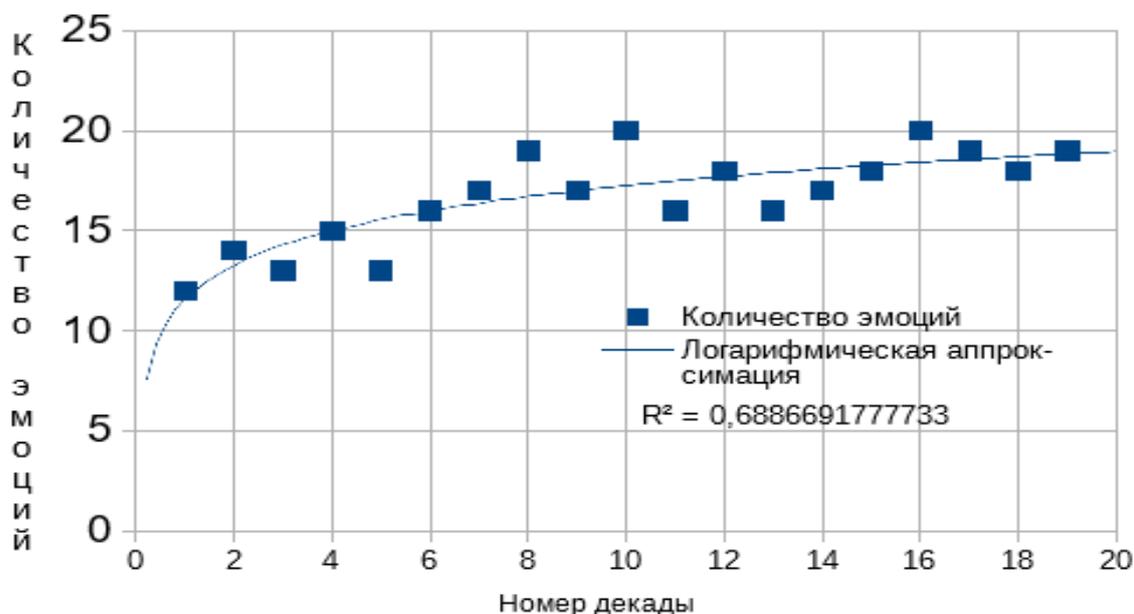


Рис. 1 График плотности однозначности выбора эмоций.

Как видно из графика, плотность неоднозначности эмоции слова возрастает с уменьшением ссылок на него. Таким образом мы определили, что неоднозначность выбора эмоции алгоритма зависит от количества изначально размеченных слов.

Заключение.

Разработанный нами программный комплекс для эмоциональной разметки прилагательных русского языка позволяет в автоматическом режиме рассчитывать эмоциональный вектор слов. Доказано, что точность выбора эмоции зависит от количества уже размеченных слов. Программный комплекс, включающий более 4000 прилагательных (с разной частотой употребления), ведет сбор данных для в режиме автономной работы в сети Интернет.

Conclusion.

We have developed a software package for the emotional marking of adjectives of the Russian language. Also, based on the approaches of distributive semantics, an algorithm was developed for automatically propagating the emotional vector to unmarked words. It is proved that the accuracy of the choice of emotions depends on the number of already marked words.

Список литературы.

1. Victor K., Hans S., Marc B. Age-of-acquisition ratings for 30,000 English words // Behavior Research Methods. 2012, №4, Vol.44, p 978–990.
2. Marc B., Amy B., Victor K. Concreteness ratings for 40 thousand generally known English word lemmas // Behavior Research Methods. 2014, №3, Vol.46, p 904–911.

3. Liu L., Preotiu-Pietro D., Samani Z., Moghadam E., Ungar L. Analyzing Personality through Social Media Profile Picture Choice // Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM 2016), p 211 – 220.

4. Maulidyani A., Manurung R. Automatic Identification of Age-Appropriate Ratings of Song Lyrics // 2015. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, p 583 – 587.

5. Russell J. A. A circumplex model of affect. // Journal of Personality and Social Psychology. 1980. № 6. Vol 39, p 1161—1178.

6. Bestgen Y., Vincze N. Checking and bootstrapping lexical norms by means of word similarity indexes // Behavior Research Methods. 2012, №3, Vol.44, p 998–1006.

7. Recchia G., Louwerse. Reproducing affective norms with lexical co-occurrence statistics: Predicting valence, arousal, and dominance // The Quarterly Journal of Experimental Psychology. 2015. №8, Vol. 68, p 1584 – 1598.

8. Ortony A., Turner T. What's Basic About Basic Emotions? // Psychological Review, 1990. V. 97, №3, P 315-331

9. Pelevina M., Arefyev N., Biemann C., Panchenko A. Making Sense of Word Embeddings // In Proceedings of the 1st Workshop on Representation Learning for NLP co-located with the ACL conference. Berlin, Germany. Association for Computational Linguistics (2016) P 174 – 183

10. Pennington J., Socher R., Manning C. Glove: Global vectors for word representation // Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). – 2014. – С. 1532-1543.

11. Lee L. Similarity-based approaches to natural language processing. // Ph.D. thesis, Harvard University, 1997

К ВОПРОСУ О ПСИХОЛОГИЧЕСКИХ ПРИЧИНАХ РЕГУЛЯЦИИ ВОЛОНТЕРСКОЙ АКТИВНОСТИ И СМЫСЛОЖИЗНЕННЫХ ОРИЕНТАЦИЙ, ДОБРОВОЛЬЦЕВ, ОКАЗЫВАЮЩИХ ПОМОЩЬ ДЕТЯМ С НАРУШЕНИЯМИ РАЗВИТИЯ И ИХ СЕМЬЯМ

Ткачева Виктория Валентиновна

профессор кафедры олигофренопедагогики и специальной психологии, профессор, доктор психологических наук,

Московский педагогический государственный университет (ФГБОУ ВО МПГУ), Москва, Россия.

Неклюкова Мария Борисовна

магистрант кафедры специальной психологии и реабилитологии ФГБОУ ВО МППУ, Москва, Россия.

[DOI: 10.31618/ESU.2413-9335.2019.7.60.64-70](https://doi.org/10.31618/ESU.2413-9335.2019.7.60.64-70)

АННОТАЦИЯ.

Статья посвящена исследованию психологических причин регуляции волонтерской помощи детям с тяжелыми и множественными нарушениями развития, с онкозаболеваниями и слепоглухим. В качестве гипотезы высказано предположение, что существуют различия уровне альтруизма, эгоизма, стремления помогать людям, в эмпатии и смысложизненных ориентациях между участвующими и не участвующими в волонтерской поддержке детям с ограниченными возможностями здоровья (ОВЗ). Также рассматривается гипотеза о различиях в структуре мотивации волонтеров, принимающих участие в разных направлениях помощи детям с нарушениями развития. Анализ данных не подтвердил гипотезу о различиях в показателях уровня эмпатии и СЖО, другие гипотезы исследования были подтверждены.

ABSTRACT.

The article is devoted to the study of the psychological reasons for the regulation of volunteer assistance to children with severe and multiple developmental disorders, with cancer and deaf and blind. As a hypothesis, it was suggested that there are differences in the level of altruism, egoism, the desire to help people, in empathy and life-meaning orientations between participating and not participating in volunteer support for children with disabilities. The hypothesis of differences in the structure of the motivation of volunteers taking part in different areas of helping children with developmental disorders is also considered. Data analysis did not confirm the hypothesis of differences in the indicators of the level of empathy and meaning of life evaluation, other research hypotheses were confirmed.

Ключевые слова: волонтеры, добровольческое движение, дети с ограниченными возможностями здоровья (ОВЗ), ребенок с тяжелыми и множественными нарушениями развития (ТМНР), мотивация волонтерской деятельности, альтруизм, эгоизм, просоциальное поведение, эмпатия, СЖО

Keywords: volunteer, volunteer movement, children with disabilities (HVD), child with severe and multiple developmental disorders (MSDD), motivation for volunteering, altruism, egoism, empathy, prosocial behavior, meaning of life orientations.

Введение

Волонтерская деятельность как форма общественной активности населения России является проявлением просоциального поведения, альтруизма и определенной структуры личности. В условиях развития рыночных отношений и глобальной конкуренции, акцентуации индивидуализации реальности человек неосознанно стремится к поиску проявления гуманистических ценностей. На сегодняшний день активно развивается направление волонтерской поддержки детям с ОВЗ и онкобольным детям. Организация комплексной поддержки для развития личности необходима каждому ребенку, а в особенности детям с тяжелыми множественными нарушениями развития, которые долгое время (до 2012 г.) в России оставались в категории необучаемых и являлись подопечными социальных служб. Успешная интеграция таких детей в общество возможна только при формировании качественной

развивающей среды и непрерывной поддержке дома или в сиротских учреждениях [9,10,11].

Проблема изучения психологических причин регуляции волонтерской активности затрагивается во многих зарубежных и отечественных исследованиях. По мнению Ф.Фуэнтос, Р.М. Газали, Е.С. Азарова, Х.К. Анхайер, Л.М. Саламон мотивация к волонтерской деятельности является мультифакторным феноменом, включающим в себя как внешние, так и внутренние причины появления потребности стать добровольцем [1, 8]. Информированность и признание позитивного статуса волонтеров способствует формированию внешних причин для развития добровольческого движения. Отечественные и зарубежные авторы отмечают, что мотивы к волонтерской деятельности не всегда являются только альтруистическими, основанными на идеалистических, религиозных взглядах (Л.А. Кудринская 2006, Е.С. Азарова 2008, У.П. Косова 2012, Кисиленко