

ДЕТЕКТИРОВАНИЕ ЭМОЦИЙ И СЕМАНТИКА ЭМОТИВОВ: ГРУСТЬ И ГНЕВ В АННОТИРОВАННОЙ ВЫБОРКЕ ТЕКСТОВ

Колмогорова А. В.

Сибирский федеральный университет (Красноярск, Россия)
ORCID ID: <https://orcid.org/0000-0002-6425-2050>

Аннотация. Статья посвящена поиску максимально приближенного к интерпретативному ощущению обычного носителя языка способа описания эмоциональной лексики. Новизна исследования связана с тем, что оно опирается на данные эмоционального аннотирования 3920 русскоязычных интернет-текстов, проведенного на значительной выборке информантов при помощи специального компьютерного интерфейса, позволяющего высчитывать «вес» 8 эмоций (грусть, радость, гнев, удивление, стыд, воодушевление, отвращение, страх) в тексте. Материал исследования, результаты которого представлены в данной публикации, составляют две выборки по 150 интернет-текстов, получивших, согласно агрегированной оценке 2000 информантов, наивысшие индексы двух эмоций – грусти и гнева. Предмет исследования – семантика эмотивов грусть и гнев через призму коллективной интроспекции информантов – носителей языка. Цель статьи – обсудить опыт экспертного семантического описания эмотивов грусть и гнев на материале интернет-текстов, аннотированных группой информантов-носителей языка как наиболее «грустные» и «гневные» тексты соответственно. Применение методов психолингвистического эксперимента, корпусного и семантического анализа привело к следующим результатам: семантические описания эмотивов грусть и гнев, полученные таким образом, 1) представляют собой социальные сценарии прототипического переживания эмоции, отражающие интерпретативные стратегии коллективного языкового сознания, а не только интроспективные ощущения эксперта-лингвиста; 2) дают возможность объяснить, почему технологии машинного обучения лучше детектируют в текстах гнев, чем грусть; 3) создают прецедент использования новых технологий для проведения семантического описания эмотивной лексики. Результаты исследования могут найти применение в эмотиологии, лексикографической практике, лингводидактике.

Ключевые слова: эмотивы; эмоции; грусть; гнев; эмоциональный анализ текстов; детектирование эмоций в тексте; семантическое описание; эмоциональное аннотирование; корпусный анализ.

EMOTION DETECTION AND SEMANTICS OF EMOTIVES: DISTRESS AND ANGER IN ANNOTATED TEXT DATASET

Anastasia V. Kolmogorova

Siberian Federal University (Krasnoyarsk, Russia)
ORCID ID: <https://orcid.org/0000-0002-6425-2050>

Abstract. The article explores the ways of making emotional lexemes semantic description consistent with interpretative intuition of the ordinary language speaker. The research novelty is determined by the fact that it is based on the data retrieved from the emotional assessment of 3920 internet-texts in Russian made by informants via using a specially designed computer interface. When applied this interface, we can aggregate the weight of 8 emotions (distress, enjoyment, anger, surprise, shame, excitement, disgust, fear) in text. Thus, the data we have used for this publication includes two sets of 150 internet-texts assessed by 2000 informants with the highest score of emotions of distress or anger. The scope of the study covers the semantics of two mentioned above lexemes (*grust'* and *gnev*) analyzed through the prism of collective introspection of informants. The article purpose is to discuss the case when a semantic description of emotives is given by an expert, which largely uses “the best texts” of corresponding emotions, according to the collective opinion of informants. Our methods include psycholinguistic experiment, corpus and semantic analysis. The research led us to three main conclusions. Firstly, the semantic descriptions of emotives *grust'* and *gnev* obtained in proposed way represent prototypical scenarios of living an emotion in social context and take into account not only the introspective sensations of an expert-ling-

guist, but the interpretative strategies of language users. Secondly, such semantic explanation provides us with keys for explaining, why machine learning technologies are better at detecting anger than sadness in text. Finally, it creates a precedent in using new technologies for making an ecological semantic description of emotive vocabulary. The research results can find application in emotiology, lexicographic practice and didactics.

Keywords: emotives; emotions; distress; anger; emotional text analysis; emotion detection; semantic description; emotional annotation; corpus analysis.

Для цитирования: Колмогорова, А. В. Детектирование эмоций и семантика эмотивов: грусть и гнев в аннотированной выборке текстов / А. В. Колмогорова. – Текст : непосредственный // Филологический класс. – 2021. – Т. 26, № 2. – С. 78–89. – DOI: 10.51762/1FK-2021-26-02-06.

Благодарности: исследование выполнено при финансовой поддержке РФФИ, проект № 19-012-00205 «Разработка классификатора русскоязычных интернет-текстов по критерию их тональности на основе модели эмоций „Куб Лёвхейма“».

For citation: Kolmogorova, A. V. (2021). Emotion Detection and Semantics of Emotives: *Distress and Anger* in Annotated Text Dataset. In *Philological Class*. Vol. 26. No. 2, pp. 78–89. DOI: 10.51762/1FK-2021-26-02-06.

Acknowledgments: the research leading to these results received funding from the Russian Foundation for Basic Research, Grant No. 19-012-00205 “Design of Sentiment Classifier for Internet Texts in Russian backed by Lövhheim’s Cube Emotion Model”.

Введение

Со времен классической работы Ч. Морриса [Morris 1938] семантика рассматривалась как область, в которой реализуется отношение языкового знака к тому элементу мира «Действительность», который этот знак замещает.

Когда таким объектом замещения становится эмоция, семантический аспект языкового знака существенно усложняется в силу того, что «Действительность» в сфере эмоциональных переживаний у каждой личности своя. При этом вступают в силу различные факторы: от индивидуального уровня эмпатии человека [Vasilyuk 2016] до степени сформированности его эмоционального интеллекта и успешности социализации. Эта двойственная природа эмоции в языке – биологическая и социальная – позволила исследователям в области биосемиотики предложить в качестве дополнения к известной типологии Ч. С. Пирса четвертый тип знаков – эмоны [Kull 2019].

Таким образом, если отвлечься от идеи системности семантических отношений в языке, а сфокусироваться на биосемиотической и экологической реальности знаковых структур, то следует, по-видимому, признать, что семантическое описание эмоциональной лексики не может обойтись без интроспекции, но последняя должна быть преимущественно коллективной или распределенной [Cowley 2004], т. е. должна не исходить из ощущений

самого исследователя-эксперта, а суммировать подобные семантические впечатления некоторого коллектива носителей языка. Текст и знак не имманентны в своем эмоциональном содержании – последнее есть только некоторый потенциал, который может быть актуализирован в контексте каждой отдельной личности, являющейся одновременно и частью социума, языкового коллектива.

Современные технологии позволяют достигать подобной коллективной распределенной интроспекции и использовать ее результаты в экспертном семантическом анализе. В данной публикации представлен подобный опыт, ставший возможным благодаря тому, что в рамках проекта по разработке компьютерного классификатора интернет-текстов на русском языке исследовательской группой лаборатории прикладной лингвистики и когнитивных исследований Сибирского федерального университета был разработан интерфейс для аннотирования текстов по критерию присутствующей в них эмоции [Kolmogorova, Kalinin, Malikova 2020]. Это позволило использовать краудсорсинг для оценки эмоций, присутствующих в текстах, и формализовать эмоции в каждом из текстов в виде векторов.

Таким образом, цель данной публикации состоит в обсуждении опыта экспертного семантического описания эмотивов *грусть* и *гнев* на материале интернет-текстов, аннотированных группой информантов-носителей

языка как наиболее «грустные» и «гневные» тексты соответственно.

1. Семантическое описание эмоциональной лексики: к истории вопроса

Разговор о способах репрезентации эмоций в языке был начат еще на заре становления структурной лингвистики Шарлем Балли, который утверждал, что, с одной стороны, эмоция может быть манифестирована в слове «рационально» – в понятийном ядре значения, а с другой – ассоциативно, непрямо, например, через коннотативный ореол (*effets par évocation*) [Bally 1921]. В рамках системно-структурной парадигмы велась работа по описанию структурной организации эмотивного значения слова [Бабенко 1989]. Перемещение же категории эмотивности на авансцену лингвистических штудий связано с идеей антропоцентризма в гуманитарных науках XX века.

Так, в Московской семантической школе была разработана семантическая «рамка» для описания эмоции в наивной языковой картине мира [Апресян 1995], представляющая собой подобие психологического сценария переживания эмоции: 1) первопричина эмоции; 2) непосредственная причина эмоции; 3) собственно эмоция, или состояние души, обусловленное тем, что человек воспринял, что он созерцал и как он это оценил; 4) желание продлить или пресечь существование причины эмоции; 5) внешнее проявление эмоции.

В работах А. А. Зализняка эмоции вписаны в иную систему координат – аксиокультурную: исследователь дает толкования эмотивов, например, *радость, удовольствие, счастье*, отталкиваясь от ценностных оппозиций, организующих русскую языковую картину мира [Зализняк 2006: 256–267].

В контексте идей ключевых концептов культуры А. Вежбицкая описывает эмотивы *гнев* и *грусть* в русском языке интроспективно, используя разработанный ею язык семантических примитивов [Вежбицкая 1999].

Когнитивный вектор в толковании лексики эмоций прослеживается в работах, выполненных в рамках лингвopsихологии, где значение лексической единицы рассматривается через призму восьми вариантов когнитивной интерпретации эмоции: эмоциональное со-

стояние; эмоциональное отношение, становление эмоционального состояния и отношения, эмоциональное воздействие, внешнее выражение эмоций, эмоциональная характеристика, эмоциональное качество, человек как средоточие и носитель эмоций [Бабенко 2020]. Парадигма концептуального анализа и функциональный подход к описанию манифестаций эмоций в языке нашли свое воплощение в работах Волгоградской школы лингвоэмотиологии В. И. Шаховского [Жура 2000; Димитрова 2001].

Появление корпусных технологий способствовало становлению так называемых *corpus-driven* исследований [Заячкова 2015], в которых автоматизированными или полуавтоматизированными методами анализируются системные контекстные связи лексем, их статистически значимые коллокации и на этой основе устанавливается их значение / функционал.

Наконец, следует констатировать, что сегодня с лингвистикой эмоций серьезно конкурирует направление, находящееся в «ведении» технологий искусственного интеллекта, – детектирование эмоций в тексте (*Emotion Detection* – например, [Mashal, Asnani 2017]). Как правило, здесь используются модели машинного обучения, которые «обучаются» выполнять поставленную задачу узнавания эмоции на основе коллекции размеченных данных (обучающего множества), где каждому тексту уже приписана определенная эмоциональная метка. Фокус исследовательского внимания распределен между тремя основными точками: 1) преселекция текстов для обучающего датасета; 2) организация процедуры аннотирования текстов выборки, в том числе – процедуры достижения согласованности аннотаций; 3) подбор «экологичных» по отношению к процедуре аннотирования алгоритмов и моделей обучения.

В данном случае интроспекция эксперта-лингвиста, характерная для исследований в области лингвистики эмоций, уступает место интроспекции ассессоров – информантов, перед которыми ставится задача оценить ведущую эмоцию в каждом тексте из обучающего датасета.

В статье мы используем в качестве материала именно такую оцененную информантами

коллекцию интернет-постов, полученную в ходе выполнения проекта в области эмоциональной оценки текстов.

2. Описание эксперимента

Поскольку целью выполняемого проекта является создание компьютерного классификатора текстов по критерию доминирующей в них эмоции, ее достижение невозможно без применения машинного обучения. Для решения задачи классификатора на базе существующих алгоритмов (как классических – например, XGBoost, модель Линейного дискриминантного анализа, некоторые модели, обученные при помощи Логистической регрессии, k-ближайших соседей и Наивного Байеса, – так и современных нейронных сетей) необходима так называемая обучающая выборка. В основе такой выборки – аннотированная коллекция текстов, где каждому тексту информанты приписывают определенную эмоцию. Для того чтобы создать подобную выборку, потребовалось решить две промежуточные задачи: 1) выбрать классификацию эмоций и 2) отобрать тексты для предъявления информантам. Несмотря на разнообразие существующих классификаций эмоций [Ekman 1992; Tomkins 1962; Plutchik 1984], мы остановились на относительно малоизвестной модели, которую называют Кубом Лёвхейма [Lövhheim 2012]. Ее преимущества связаны с трехмерностью самой модели, позволившей разра-

ботать интерфейс для недискретного аннотирования, и с разумным количеством выделяемых эмоций: грусть, радость, отвращение, удивление, воодушевление, гнев, стыд, страх.

Для решения второй задачи при помощи предварительно валидированных в психолингвистическом эксперименте хештегов (например, #_фуу или #Грустненько) была сделана преселекция из 15000 постов (например, #_фуу или #Грустненько) из пабликов «Подслушано», «Палата № 6» и «Карамель» в социальной сети ВКонтакте. Из каждого класса текстов, отобранных по хештегам и прошедших преселекцию, было случайным образом взято по 490 образцов. В итоге объем этой второй выборки, предложенной для эмоциональной оценки, составил 3920 текстов из трех указанных выше пабликов. В качестве платформы для проведения эмоциональной оценки был использован краудсорсинговый ресурс «Яндекс Толока», где можно выбрать информантов согласно их рейтингу, который составляется по итогам оценки качества их работы в других проектах. В данном случае мы постарались отобрать информантов, входящих в среднюю когорту. В общей сложности все тексты были оценены 2000 информантами.

Подход к недискретной разметке текстов, разработанный нами на предыдущем этапе, был адаптирован к техническим условиям платформы¹. Реализованный интерфейс представлен на рисунке 2.

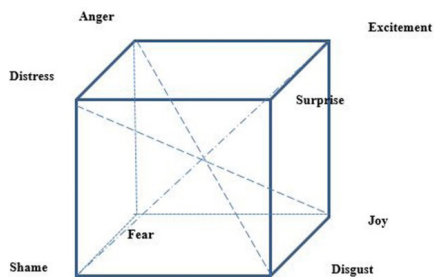


Рис. 1. Внутренние диагонали куба

ae6a61a35eed1dcbefc039483f402dc3 Действия ▾

Хочу сказать спасибо всем работодателям, не принявшим меня на работу по специальности. Благодаря вам, я получила хороший толчок к созданию своего дела. Теперь у меня свой бизнес, который успешно развивается и приносит хороший доход. Самое главное, я не стала рабом системы, а стала воистину хозяйкой своей жизни.)

Стыд	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Воодушевление
Отвращение	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Гнев
Страх	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Удивление
Удовольствие	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Тоска/Грусть

Рис. 2. Интерфейс для недискретной разметки на платформе «Яндекс Толока»

¹ Данные могут быть получены по ссылке <https://storage.yandexcloud.net/nlp-dataset-bucket-1/toloka-vk-proceedings-2020/toloka-vk-raw-unprocessed.tsv>.

Каждая из четырех шкал для эмоциональной оценки представляла собой внутреннюю диагональ куба Лёвхейма (рис. 1). В качестве нейтральной (исходной) позиции понимается центр куба, а на шкалах – середина шкалы. Каждое из четырех измерений трактовалось как вектор со своим углом в трехмерном пространстве и магнитудой. Эти четыре вектора затем агрегировались в один результирующий вектор путем усреднения. С помощью конечной точки результирующего вектора мы определяли «эмоциональные» координаты текста в пространстве Куба¹.

3. Материал и методы

Таким образом, мы получили коллекцию текстов, в которой каждый текст, согласно «усредненному» мнению 2000 информантов (ассессоров) имеет определенный эмоциональный вес. Кроме того, нам удалось извлечь из данной коллекции и некоторые специфические данные, а именно: 1) если составить топ-150 текстов, имеющих, по мнению информантов, самый высокий индекс той или иной эмоции, т. е. визуализировать место этих текстов внутри Куба Лёвхейма, то соответствующие им точки окажутся ближе всего к вершинам Куба; 2) получить статистику о том, как каждый из информантов расставлял ползунок на четырех предложенных шкалах.

Первый тип данных будет использован нами как материал 1) для того, чтобы показать через корпусные методы, чем очень эмоциональные тексты отличаются от нейтральных, поскольку корпус нейтральных текстов также имеется в распоряжении исследовательской группы (подробнее см. [Kolmogorova, Kalinin, Malikova 2021]), и сделать набросок семантической рамки для толкования эмоции «вообще»; 2) для семантического анализа имен соответствующих эмоций (грусть и гнев), дополняющего и специфицирующего заданную семантическую рамку.

Отметим, что мы остановимся только на данных двух эмоциях, поскольку грусть, согласно результатам работы всех использованных алгоритмов машинного обучения, хуже

всего поддается детектированию в тексте на русском языке независимо от используемой модели, а гнев, наоборот, – хорошо распознается независимо от применяемого алгоритма. Возможно, объяснение данного феномена кроется в семантической природе этих эмотивов.

Второй тип данных мы коротко рассмотрим в следующем разделе для того, чтобы проиллюстрировать наше мнение о том, что интерпретация эмоциональной семантики языковых единиц – процесс, подверженный влиянию самых различных индивидуально-обусловленных факторов, учесть которые помогает «коллективная интроспекция» посредством краудсорсинговой эмоциональной оценки.

Для анализа первого из упомянутых типов данных с целью семантического описания эмотивов *грусть* и *гнев* используется инструментарий корпусного менеджера «Sketch Engine».

4. Описание распределения стратегий оценки, продемонстрированных информантами в экспериментальной работе

Анализ деятельности информантов во время выполнения процедуры эмоциональной оценки текстов продемонстрировал, что один и тот же текст провоцирует различные стратегии оценки присутствующей в нем эмоции. В то же время у каждого информанта есть свои стереотипные модели оценки.

Таких стратегий мы выделили пять (рис. 3). Две стратегии лидируют: сбалансированная стратегия (balanced) – после прочтения текста информант приписывает приблизительно равные, но не максимальные, «веса» трем и более эмоциям; стратегия двух вершин (two edges) – на двух из четырех шкал информант ставит ползунок на крайние отметки, выделяя таким образом 2 ведущие эмоции в тексте. Следующие две стратегии являются несколько менее частотными: используя стратегию одной вершины (one edge), информант приписывает максимальные значения интенсивности только одной эмоции, а на остальных шкалах перемещает ползунок на нулевую отметку; опираясь на экстремальную стратегию

¹ Алгоритм пересчета значения векторов измерений в результирующую точку можно найти по этой ссылке: <https://colab.research.google.com/drive/1ASaVNleaHrYxwsr8XK37hPqygoNkawLi?usp=sharing>; Данные с пересчитанными значениями доступны здесь: https://storage.yandexcloud.net/nlp-dataset-bucket-1/toloka-vk-proceedings-2020/toloka_recalc_lean.csv.

(extreme), участник эксперимента отмечает максимальные значения на каждой из четырех шкал. Наконец, наименее востребованная стратегия – пассивная (passive): информант на всех шкалах ставит ползунок на нулевую отметку, что эквивалентно полной эмоциональной нейтральности текста. Однако одновременно такой тип поведения аннотатора можно интерпретировать как свидетельство его нежелания решать поставленную задачу. Тот факт, что в наших данных эта стратегия оказалась наименее частотной, косвенно указывает на ответственное отношение информантов к разметке. Шестая категория объединяет в себе типы поведения, которые мы затруднились классифицировать (other).

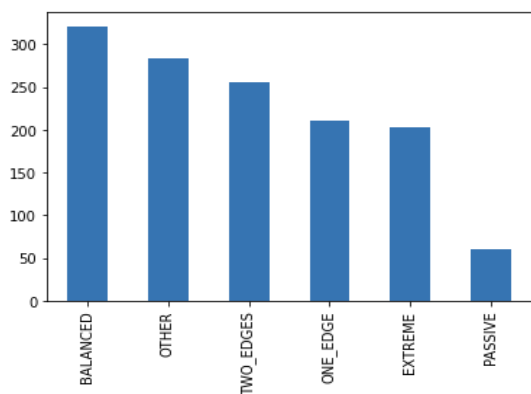


Рис. 3. Распределение стратегий, использованных всеми информантами

Таким образом, те 150 текстов, которые получили наиболее высокий индекс эмоций гнева и грусти по результатам проведенной разметки, являются, по сути, квинтэссенцией тех представлений о данных эмоциях, которые сложились в современном русскоязычном социуме. Причем эмоциональный «рейтинг» этих текстов отражает и индивидуальные стратегии носителей языка, обобщенные до уровня коллективной интроспекции.

5. Семантическое описание эмотивов гнев и грусть

Для того чтобы задать некоторую семантическую рамку для будущих толкований, следует ответить на вопрос о том, чем тексты с высокой эмоциональной насыщенностью отличаются от нейтральных текстов. Для этого

сравним два подкорпуса текстов – подкорпус из 150 самых эмоциональных текстов для 8 эмоций (1200 текстов) и подкорпус нейтральных текстов, полученный за счет автоматического извлечения их из социальных сетей предварительно предобученным классификатором.

Сравнение двух подкорпусов при помощи инструмента «Key n-gram», встроенного в платформу «Sketch Engine», дает следующие результаты в отношении самых специфичных для эмоционального подкорпуса триграмм: *что у меня, когда мне было, с тех пор, и как же, так и не, а у меня, потому что я, потому что ее, ни разу не.*

Они далеко не тривиальны, а достаточно показательны, поскольку очерчивают основные границы ситуации переживания эмоции.

1. Признак «обладание» ((что) у меня), который может принимать разные формы: обладание жизненным опытом (*у меня был секс*), материальным ресурсом (*у меня есть деньги*), социальными связями (*у меня есть поклонник, дочь, подруга, сестра, парень, коллега*), даже некоторым девиантным телесным или психическим качеством (*у меня есть бзик, сломан позвоночник, аносмия, отит, гайморит, фронтит*).

2. Присутствует некая хронологическая отметка в персональной истории (*когда мне было*). Это может быть обозначение какого-то точного, знакового момента, как правило, в детстве или в подростковом возрасте (*когда мне было 11 месяцев, 2, 3, 4 года, 5, 6, 11, 12, 15 лет*) или события во взрослой жизни, после которого все резко изменилось (*с тех пор*): *с тех пор как умерла моя мамочка я стала пить; с тех пор никаких автоматов, езжу исключительно на механике; с тех пор все наперекосяк, облом за обломом.*

3. Идея сфокусированности на своем внутреннем мироощущении, сосредоточенности на нем. Подобная интроспективность маркируется коллокациями и *как же меня бесит, раздражает / и как же мне обидно, стыдно, офигенно, непередаваемо круто, хорошо осознавать, меня* и т. д.

4. Признак «несоответствие норме / противопоставление себя чему-либо» (*так и не и а у меня*), например, несоответствие желаемого / нормы и реальности (*Я так и не решилась взять щенка; А водилу, который меня сбил, так и не нашли*) или противопоставление себя

другим людям (*семья у него была зажиточная, а у меня – как и у большинства в 90-х; мужики всегда любили в отношениях со мной то, что я не ною как другие бабы*), социальным нормам и правилам (*а у меня никаких ургрызений совести*).

5. Данное противопоставление дополняется признаком исключительности ситуации, в которой оказался говорящий (*ни разу не*), относительно социальных стереотипов (*за все это время ни разу не готовила еду, даже яичницу, все готовит мой любимый*), примет и ритуалов (*она мне ни разу не снилась после смерти*).

6. Наконец, ситуация переживания эмоции дополняется поисками ее первопричины, которую говорящий видит либо в себе, либо в других людях: *мы не ругаемся, потому что каждый раз я боюсь, что это наша последняя встреча; она молчала до последнего, потому что ее «лекари» сказали, что опухоли можно лечить, но в полной тайне, иначе слезают*.

Таким образом, прототипическая ситуация переживания сильной эмоции может быть представлена следующим образом: в определенный момент своей жизни Т человек Х оказывается в ситуации Y, которую Х считает а) исключительной, б) заставляющей сосредотачиваться на своих переживаниях, в) дающей ему ощущение обладания некоторыми объектами и отношениями, которые г) заставляют его чувствовать свою противопоставленность по отношению к другим людям и / или д) заставляют искать причину эмоции.

Иначе говоря, описание эмотива может включать в себя пять ключевых признаков: *исключительность, обладание, противопоставление, причинность, интроспективность*. При описании грусти и гнева мы посмотрим роль каждого из этих признаков в экспликации эмоции и при необходимости дополним их.

Для семантической интерпретации грусти признак «противопоставление» играет достаточно важную роль. Он, в частности, просматривается на уровне системных лексических отношений и морфологии.

Так, например, в списке 50 наиболее частотных прилагательных подкорпуса из 150 самых грустных текстов обнаруживается достаточно яркое присутствие антонимичных пар (табл. 1).

Антонимы как бы задают общую идею внутреннего конфликта между собой и миром, существенного для субъекта эмоции.

Таблица 1. Словоформы антонимичных лексем в списке частотных прилагательных подкорпуса грустных текстов

№	СЛОВОФОРМЫ АНТОНИМИЧНЫХ ЛЕКСЕМ (НОРМАЛИЗОВАННЫЕ ЧАСТОТЫ В 1РМ)
1	огромный – маленький (231.178 – 154.119)
2	молодой – пожилой (308.231 – 231.178)
3	новый – старенький (370.123 – 77.9)
4	лишний – нужный (154.19 – 154.19)
5	нормальный – странный (154.19 – 154.19)

На фоне подкорпуса нейтральных текстов в «грустных» шестью наиболее специфическими триграммами являются сочетания, включающие отрицательную частицу *не*: *и я не, что я не, (до) сих пор не, я не могу, я не знаю, все равно не*. «Настойчивость», с которой появляются в корпусе формы отрицания, также становится маркером признака «противопоставленность», ощущаемой говорящим по отношению к социуму и его нормам. Однако эта противопоставленность, как правило, со знаком «–»: говорящий ощущает, что «не дотягивает» до социально приемлемой нормы по каким-либо параметрам. Например:

1) *Я «скрываю» своего двоюродного брата от друзей. Он очень добрый, щедрый и отзывчивый парень, но он абсолютно не образован и очень часто несет несусветную чушь, даже не понимая, как глупо он выглядит. И я не могу решиться пригласить его на свой день рождения уже который год, потому что среди моих начитанных и остроумных друзей он будет выглядеть глупо и унижительно. Я не стесняюсь его и очень люблю, но не хочу выставлять его на посмешище.*

Хотя первоначально в социальной сети пост (пример 1) был размещен под хештегом #стыдно, информанты оценили его как один из самых грустных. Очевидно, поводом для такой оценки стало противоречие между отношением девушки к брату и ее ощущением, что он не соответствует нормам того круга, в котором она общается.

Для гнева признак «противопоставление» также релевантен, но проявляется по-другому: субъект эмоции находится в доминирующей позиции по отношению к социуму и его нормам. На это указывает, например, тот факт, что подавляющее большинство ключевых коллокаций «гневного» подкорпуса при анализе на фоне нейтрального включают в

себя лексемы с ярко выраженной оценочной семантикой, например: *на хрена так, драгоценное время, гениальнейший ответ, назойливый смс-спам, банальное отношение, великолепная идея, разжиревший муж, дурной вкус, беспробудное пьянство, блондинистая мразь*. Аксиологическая поляризация косвенно указывает на то, что говорящий считает себя вправе выносить оценочные суждения относительно «других».

Достаточно интересно проявляет себя признак «обладание».

В «грустных» текстах, по сути, речь идет об «антиобладании», т. е. об обладании тем / кем, кто составляет трудность, проблему, требует заботы. Среди номинаций родственников – лексемы *сестра, дети*, номинирующие самых незащищенных и зависимых членов семьи (*у меня сестра / дети*). Неодушевленные объекты обладания также окружены негативным коннотативным ореолом: *у меня ипотека / галлюцинации / ужасный характер*. Самую обширную группу составляют номинации одушевленных и неодушевленных объектов, которых у субъекта эмоции нет: *у меня нет друзей / друга / любимой девушки / вечеринок / посиделок / совместных проектов / просмотров фильмов вдвоем*. Следующий текстовый фрагмент хорошо отражает те вербальные манифестации признаков «обособленность» и «антиобладание», которые были описаны выше:

2) Больше не могу... Не могу. Смотрю на бывших одноклассников, завидую и плачу по ночам. Мне так стыдно, так больно. У меня уже давно нет девушки. Так давно, что я уже и не верю, что когда-нибудь ее найду.

В «гневных» же текстах наоборот: речь идет об обладании со знаком «+» (*у меня парень, очередь из желающих, крутая тачка, возможности, все хорошо, кожа белая, глаза красивые* и т. д.).

Семантический признак «интроспективность» хорошо просматривается в «грустных» текстах посредством применения инструмента «поиск ключевых слов» при анализе обсуждаемого подкорпуса на фоне нейтральных текстов. Так, в шорт-листе из десяти ключевых глаголов для «грустных» текстов присутствуют четыре глагола, номинирующие внутренние предикаты, которые характеризуют ситуацию, недоступную внешнему наблюдению – о ней знает только переживающий

ее субъект: узнала, знала, чувствую, боялся. Среди десяти ключевых словосочетаний три апеллируют соответственно к температурному, слуховому восприятию и обонянию: *ледяная вода, шумный город, вонючий носок*.

Для эмоции гнева интроспективность, по видимому, не характерна. В списках частотных слов, ключевых лексем или коллокаций (на фоне нейтрального подкорпуса) не обнаружено ни одного из перечисленных выше маркеров. Кроме того, из 50 ключевых для гневного подкорпуса слов 17 лексем – это существительные конкретной семантики (*посуда, наушник, тарелка, мусор, изюм, песок, мыло, штора, вагон, камера, руль, печенка, прокладка, рюкзак, сиг, домик, станок*), в то время как для «грустных» текстов в подобном списке доминировали абстрактные существительные и глаголы ментальной деятельности / восприятия.

Признак «причинность» обнаруживается и в «грустных», и в «гневных» текстах, но имеет разнонаправленные векторы. Так, в «грустных» текстах причину эмоции говорящий видит в социальной ситуации, спровоцированной им самим (пример 3) – я страдаю, потому что у меня самой никого нет:

3) Я одинока. Живу одна и много работаю. Как-то со мной в автобусе ехали старушка и женщина с маленьким ребенком в коляске. Старушка жаловалась на новые дома – что их построили некрасиво и ненадежно. Спустя пару остановок бабушка стала заигрывать с ребенком: она хлопала в ладоши, пытаясь привлечь его внимание, и смеялась. А я стояла рядом и плакала. Я вдруг ясно увидела свое будущее: через много лет, будучи уже немолодой и некрасивой женщиной, радоваться вниманию чужих детей, потому что у меня самой НИКОГО нет.

В «гневных» же текстах причина эмоции кроется в социальных ситуациях, спровоцированных другими, поэтому маркером причинности в таких текстах является не союз *потому что*, а коллокация *бесит, когда* (пример 4):

4) Меня дико бесит, когда обо мне или о ком-либо говорят «молодой человек». Для меня это звучит так, как будто ко мне обращается какая-то бабушка или как-то по-старинному что ли. Это все равно, что парни называли бы своих девушек «мадемуазель». Так ладно бы еще молодым человеком называли действительно молодых.

Наконец, высокая частотность (второй и третий ранг в частотном списке прилагатель-

ных) лексем прошлый и бывший (например, *бывшая военная часть, бывшая одноклассница, прошлый год*) в «грустных» текстах указывает на то, что ситуация, ставшая триггером эмоции грусти, находится в прошлом. Что касается самой этой ситуации, то она относится к повседневной жизни и жизни семьи. Частотные сочетания здесь – *в будни, на похоронах, в семье, в гостях*.

Признак времени не представлен в «гневных» текстах, а вот сама триггерная ситуация локализуется в разнообразных публичных пространствах, а не в семье, как в «грустных» текстах. Здесь речь идет об эпизодах *в маршрутке, в магазине, в офисе, в коридоре, в квартире, в госучреждении, в цеху, в подъезде, в метро, в садику, в школе, в больнице, в аптеке, в торговом центре*.

Суммируя, представим набросок семантического описания эмотивов *грусть* и *гнев* с опорой на пять признаков, включенных нами в семантическую рамку для описания эмоции (табл. 2). Отметим, что данные описания отнюдь не носят модельного характера – они выстроены исключительно на материале нашей выборки «лучших текстовых представителей» той или иной эмоции.

Таблица 2. Семантические профили лексем *грусть* и *гнев*

Семантический признак	ГРУСТЬ	ГНЕВ
противопоставление	+ ↓	+ ↑
обладание	+ (негатив)	+ (позитив)
исключительность	–	–
причинность	+ (в себе)	+ (в других)
интроспективность	+	–

Можно заметить, что четыре из пяти выделенных признаков валидны для описания грусти. При этом признак «обладание», присутствуя, имеет «негативное» наполнение: речь идет об обладании объектами, причиняющими беспокойство и проблемы. Признак «противопоставление» в координатах «я – другие» реализуется за счет принижения себя (стрелка, направленная вниз).

Для описания гнева оказались релевантны три признака из пяти. Признак «обладание» имеет помету «позитив», поскольку речь идет о владении престижными объектами. Идея противопоставленности себя другим реализуется с учетом собственного субъективно бо-

лее высокого статуса (стрелка, направленная вверх). Таким образом, исходя из нашего материала, согласно показаниям коллективной интроспекции информантов-носителей русского языка, семантика лексемы *грусть* может быть описана в виде следующего прототипического сценария:

– в определенный момент своей жизни в прошлом человек оказывается в будничной ситуации, связанной преимущественно с семьей; ситуация воспринимается им как та, которая вынуждает его сосредотачиваться на своем внутреннем мире; она дает ему ощущение обладания объектами, отношениями, *затрудняющими его жизнь*, которые заставляют его чувствовать свою противопоставленность по отношению к другим людям *в силу субъективного ощущения неполноценности*; заставляют видеть в себе причину ситуации, провоцирующей эмоцию.

Для гнева, по-видимому, следует предложить такую трактовку:

– в некоторый момент своей жизни человек оказывается в ситуации *У*, разворачивающейся в публичном социальном пространстве; эта ситуация воспринимается им как такая, в которой он ощущает себя обладателем некоторых *внешних атрибутов успеха, материальных ценностей*, заставляющих его чувствовать свою противопоставленность по отношению к другим людям *в силу некоторого превосходства*; заставляют видеть причину ситуаций, провоцирующих эмоцию *не в себе, а в других*.

Заключение

Проведенный анализ позволяет сделать несколько как практических, так и теоретических выводов.

Практический вывод касается ответа на вопрос, почему «грустные» тексты классифицируются гораздо хуже (значение точности и полноты классификации f_1 , независимо от применяемого алгоритма машинного обучения, не превышает 78%), чем «гневные» (значение точности и полноты классификации f_1 , независимо от применяемого алгоритма, не опускается ниже 94%). По-видимому, это связано с более внешним, эксплицитным переживанием эмоции гнева, каждый семантический признак которой так или иначе связан с субъективным ощущением говорящим своего социально-

го превосходства. Последнее, в свою очередь, имеет в языке яркие лексические средства манифестации: экспрессивно-оценочные лексемы (преимущественно с негативной коннотацией), специфические чрезвычайно частотные специфические коллокации, имеющие экспрессивный оттенок (*бесит, когда*). Их маркированный с точки зрения стилистической нормы характер легко идентифицируется алгоритмами машинного обучения.

Эмоция грусти манифестируется в языке, во-первых, большим количеством признаков, а во-вторых, более имплицитными и сложными способами. Так, на первый план выходит не стилистическая окраска лексем, а их ассоциативный ореол (*ипотека*, например, не имеет фиксированных сем негативной оценки даже в периферийной зоне значения, но ассоциативное значение чрезвычайно негативно нагружено). Актуализируются системные отношения в лексике (антонимы) и морфологические особенности (категория отрицания). Не последнюю роль играет семантика предикатов, обработать и формализовать которую чрезвычайно трудно, и это также затрудняет автоматическую идентификацию данной эмоции.

Что касается теоретического наблюдения, то хочется отметить, что в процессе выполнения проекта, находящегося в области междисциплинарного взаимодействия лингвистики и машинного обучения, а именно – в парадигме аффективных вычислений (*Affective computing*), выкристаллизовалось понимание того, что новые технологии дают лингвистам перспективные возможности для того, чтобы

сделать семантические исследования максимально приближенным к интуитивным ощущениям рядового носителя языка. В эпоху научного позитивизма лингвистика довольно часто слышала от представителей естественных наук упреки в излишней интроспективности и субъективности исследовательских практик. Когда речь идет об эмоциях, кажется, что от данных принципов не уйти. Но апробированный нами кейс, основанный на эмоциональном аннотировании текстов с использованием краудсорсинга и специально разработанного интерфейса, позволяющего приписывать каждому тексту некоторый эмоциональный «вес», предлагает возможный выход за пределы индивидуальной субъективности в сферу интерсубъективности. В свое время Ян Хэкинг [Hacking 1999; цит. по: Cowley 2016] сформулировал мысль о том, что предназначение науки состоит в том, чтобы постоянно перекидывать мостик между нашим представлением об объекте X и самим этим объектом. Предложенная методика позволяет сделать такой мостик коллективным: речь идет о представлении реальных носителей языка об объекте-эмоции. В случае описаний эмотивов *грусть* и *гнев*, представленных в данной статье, мы можем наблюдать, что опора на «коллективную интроспекцию» позволяет получить толкования, отражающие прототипический *социальный* сценарий проживания данных эмоций. Такой подход имеет элемент новизны, поскольку отходит от психологически ориентированных, лингвоаксиологических и, в определенной степени, системно-структурных способов описания эмоциональной лексики.

Литература

- Апресян, Ю. Д. Избранные труды. Том II. Интегральное описание языка и системная лексикография / Ю. Д. Апресян. – М. : Языки русской культуры, 1995. – 767 с.
- Бабенко, Л. Г. Лексические средства обозначения эмоций в русском языке / Л. Г. Бабенко. – Свердловск : Изд-во Урал. ун-та, 1989. – 184 с.
- Вежицкая, А. «Грусть» и «гнев» в русском языке: неуниверсальность так называемых «базовых человеческих эмоций» / А. Вежицкая // Семантические универсалии и описание языков. – М. : Языки русской культуры, 1999. – С. 503–526.
- Димитрова, Е. В. Языковые средства трансляции эмотивных смыслов русского концепта «тоска» во французскую лингвокультуру : дис. ... канд. филол. наук / Димитрова Е. В. – Волгоград, 2001. – 173 с.
- Жура, В. В. Эмоциональный дейксис в вербальном поведении английской языковой личности (на материале англоязычной художественной литературы) : дис. ... канд. филол. наук / Жура В. В. – Волгоград, 2000. – 200 с.
- Зализняк, А. А. Многозначность и способы ее представления / А. А. Зализняк. – М. : Языки славянских культур, 2006. – 676 с.
- Заячковская, О. О. Концептуальный анализ семантики эмоционального лексикона / О. О. Заячковская // Методы когнитивного анализа семантики слова: компьютерно-корпусный подход. – М. : Языки славянской культуры, 2015. – С. 243–268.

Колмогорова, А. В. Лингвистика когнитивная и экологичная: к вопросу о перспективах применения концепции когнитивной экологии в лингвистических исследованиях / А. В. Колмогорова // Экология языка и коммуникативная практика. – 2019. – № 3. – С. 19–28.

Bally, Ch. *Traité de stylistique française* / Ch. Bally. – Heidelberg, 1921. – 349 p.

Cowley, S. J. Changing the idea of language: Nigel Love's perspective / S. J. Cowley // *Language Sciences*. – 2016. – Vol. 61. – P. 43–55. – DOI: <https://doi.org/10.1016/j.langsci.2016.09.008>.

Cowley, S. J. Contextualizing bodies: how human responsiveness constrains distributed cognition / S. J. Cowley // *Special issue on Integrational Linguistics and Distributed Cognition. Language Sciences*. – 2004. – № 26/6. – P. 565–591.

Ekman, P. An argument for basic emotions / P. Ekman // *Cognition & Emotion* – 1992. – 6 (3–4). – P. 169–200. – DOI: <https://doi.org/10.1080/02699939208411068>.

Hacking, I. *The Social Construction of What?* / I. Hacking. – Oxford University Press, 1999. – 278 p.

Kolmogorova, A. V. Non-discrete Sentiment Dataset Annotation: Case Study for Lövheim Cube Emotional Model / A. V. Kolmogorova, A. A. Kalinin, A. V. Malikova // *Communications in Computer and Information Science*. – 2020. – Vol. 1242. – P. 154–164. – DOI: [10.1007/978-3-030-65218-0_12](https://doi.org/10.1007/978-3-030-65218-0_12).

Kolmogorova, A. V. The problem of retrieving neutral classes of texts in Russian in multiclass emotional text analysis / A. V. Kolmogorova, A. A. Kalinin, A. V. Malikova // *EUR Workshop Proceedings*. – 2020. – Vol. 28522020.

Kull, K. Steps towards the natural meronymy and taxonomy of semiosis: Emotin between index and symbol? / K. Kull // *Sign Systems Studies*. – 2019. – № 47 (1/2). – P. 88–104. – DOI: <https://doi.org/10.12697/SSS.2019.47.1-2.03>.

Lövheim, H. A New Three-Dimensional Model for Emotions and Monoamine Neurotransmitters / H. Lövheim // *Medical Hypotheses*. – 2012. – Vol. 78. – P. 341–348. – DOI: <https://doi.org/10.1016/j.mehy.2011.11.016>.

Mashal, S. X. Emotion intensity detection for social media data / S. X. Mashal, K. Asnani // *Proceedings of the 2017 International Conference on Computing Methodologies and Communication (ICCMC)*. – 2017. – P. 155–158.

Morris, Ch. W. Foundations of the theory of signs / Ch. W. Morris // *International encyclopedia of unified science*. 1938. – Vol. 1, № 2.

Plutchik, R. Emotions: A general psychoevolutionary theory / R. Plutchik // *Approaches to emotion* / ed. by K. Scherer, P. Ekman. – Hillsdale: Lawrence Erlbaum Associates, 1984. – P. 197–219. – DOI: <https://doi.org/10.4324/9781315798806>.

Tomkins, S. S. *Affect Imagery Consciousness*. Vol. 1: The Positive Affects / S. S. Tomkins. – New York: Springer, 1962. – 588 p.

Vasilyuk, F. E. Semiotics and the technique of empathy / F. E. Vasilyuk // *Journal of Russian & East European Psychology*. – 2016. – № 53 (2). – P. 56–79. – DOI: <https://doi.org/10.1080/10610405.2016.1230994>.

References

Apresyan, Yu. D. (1995). *Izbrannyye yrudyy. Tom II. Integral'noe opisanie yazyka i sistemnaya leksikografiya* [Selected Works. Volume II. Integral Description of Language and Systemic Lexicography]. Moscow, Yazyki russkoi kul'tury. 767 p.

Babenko, L. G. (1989). *Leksicheskie sredstva oboznacheniya emotsiy v russkom yazyke* [Lexical Means of Emotion Denotation in Russian]. Sverdlovsk, Izdatel'stvo Ural'skogko universiteta. 184 p.

Vezhbitskaya, A. (1999). «Grust' i «gnev» v russkom yazyke: neuniversal'nost' tak nazyvaemykh «bazovykh chelovecheskikh emotsiy» [Distress and Anger in Russian: about the Non-Universality of Basic Human Emotions]. In *Semanticheskie universalii i opisanie yazykov*. Moscow, Yazyki russkoi kul'tury, pp. 503–526 p.

Dimitrova, E. V. (2001). *Yazykovyye sredstva translyatsii emotivnykh smyslov russkogo kontsepta «toska» vo frantsuzskuyu lingvokul'turu* [Linguistic Means of Translating the Emotive Meanings of the Russian Concept *Toska* into the French Linguoculture]. Dis. ... kand. filol. nauk. Volgograd. 173 p.

Zhura, V. V. (2000). *Emotsional'nyi deiksis v verbal'nom povedenii angliiskoi yazykovoi lichnosti (na materiale angloyazychnoi khudozhestvennoi literatury)* [Emotional Deixis in the Verbal Behavior of an English Language Personality (Based on the Material of English-Language Fiction)]. Dis. ... kand. filol. nauk. Volgograd. 200 p.

Zaliznyak, A. A. (2006). *Mnogoznachnost' i sposoby ee predstavleniya* [Polysemanticity and Ways of Its Manifestation]. Moscow, Yazyki slavyanskikh kul'tur. 676 p.

Zayachkovskaya, O. O. (2015). Kontseptual'nyi analiz semantiki emotsional'nogo leksikona [Conceptual Analysis of Emotional Lexicon Semantics]. In *Metody kognitivnogo analiza semantiki slova: komp'yuterno-korpusnyi podkhod*. Moscow, Yazyki slavyanskoi kul'tury, pp. 243–268.

Kolmogorova, A. V. (2019). Lingvistika kognitivnaya i ekologichnaya: k voprosu o perspektivakh primeneniya kontseptsii kognitivnoi ekologii v lingvisticheskikh issledovaniyakh [When Cognitive Linguistics Is Ecological Too: on the Question of the Prospects for the Application of the Concept of Cognitive Ecology in Linguistic Research]. In *Ekologiya yazyka i kommunikativnaya praktika*. No. 3, pp. 19–28.

Bally, Ch. (1921). *Traité de stylistique française*. Heidelberg, 1921. 349 p.

Cowley, S. J. (2016). Changing the Idea of Language: Nigel Love's Perspective. In *Language Sciences*. Vol. 61, pp. 43–55. DOI: <https://doi.org/10.1016/j.langsci.2016.09.008>.

Cowley, S. J. (2004). Contextualizing Bodies: How Human Responsiveness Constrains Distributed Cognition. In *Special issue on Integrational Linguistics and Distributed Cognition. Language Sciences*. No. 26/6, pp. 565–591.

Ekman, P. (1992). An Argument for Basic Emotions. In *Cognition & Emotion*. No. 6 (3–4), pp. 169–200. DOI: <https://doi.org/10.1080/02699939208411068>.

Hacking, I. (1999). *The Social Construction of What?* Oxford University Press. 278 p.

Kolmogorova, A. V., Kalinin, A. A., Malikova, A. V. (2020). Non-discrete Sentiment Dataset Annotation: Case Study for Lövheim Cube Emotional Model. In *Communications in Computer and Information Science*. Vol. 1242, pp. 154–164. DOI: 10.1007/978-3-030-65218-0_12.

Kolmogorova, A. V., Kalinin, A. A., Malikova, A. V. (2020). The Problem of Retrieving Neutral Classes of Texts in Russian in Multiclass Emotional Text Analysis. In *EUR Workshop Proceedings*. Vol. 28522020.

Kull, K. (2019). Steps towards the Natural Meronymy and Taxonomy of Semiosis: Emotin between Index and Symbol? In *Sign Systems Studies*. No. 47 (1/2), pp. 88–104. DOI: <https://doi.org/10.12697/SSS.2019.47.1-2.03>.

Lövheim, H. (2012). A New Three-Dimensional Model for Emotions and Monoamine Neurotransmitters. In *Medical Hypotheses*. Vol. 78, pp. 341–348. DOI: <https://doi.org/10.1016/j.mehy.2011.11.016>.

Mashal, S. X., Asnani, K. (2017). Emotion Intensity Detection for Social Media Data. In *Proceedings of the 2017 International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 155–158.

Morris, Ch. W. (1938). Foundations of the Theory of Signs. In *International encyclopedia of unified science*. Vol. 1. No. 2. Chicago, The University of Chicago Press.

Plutchik, R. (1984). Emotions: A General Psychoevolutionary Theory. In Scherer, K., Ekman, P. (Eds.). *Approaches to emotion*. Hillsdale, Lawrence Erlbaum Associates, pp. 197–219. DOI: <https://doi.org/10.4324/9781315798806>.

Tomkins, S. S. (1962). *Affect Imagery Consciousness*. Vol. 1: *The Positive Affects*. New York, Springer. 588 p.

Vasilyuk, F. E. (2016). Semiotics and the Technique of Empathy. In *Journal of Russian & East European Psychology*. No. 53 (2), pp. 56–79. DOI: <https://doi.org/10.1080/10610405.2016.1230994>.

Данные об авторе

Колмогорова Анастасия Владимировна – доктор филологических наук, профессор, заведующий кафедрой романских языков и прикладной лингвистики, Сибирский федеральный университет (Красноярск, Россия).

Адрес: 660041, Россия, г. Красноярск, Свободный пр., 82А.

E-mail: nastiakol@mail.ru.

Author's information

Kolmogorova Anastasia Vladimirovna – Doctor of Philology, Professor, Head of Department of Romance Languages and Applied Linguistics, Siberian Federal University (Krasnoyarsk, Russia).