

УДК 81'322.2

КОЛИЧЕСТВЕННЫЙ СРАВНИТЕЛЬНЫЙ АНАЛИЗ СЕНТИМЕНТА РЕЧИ АВТОРА И РЕЧИ ПЕРСОНАЖЕЙ НА МАТЕРИАЛАХ КОРПУСА РУССКОГО РАССКАЗА

Анастасия Романовна Сейнова

стажер-исследователь лаборатории языковой конвергенции

Национальный исследовательский университет

Высшая школа экономики – Санкт-Петербург

190068, Санкт-Петербург, наб. канала Грибоедова, 119–121. arseynova@edu.hse.ru

Статья посвящена измерению эмоциональной составляющей русской малой прозы начала XX в. На основе случайной выборки в 70 рассказов, написанных в 1900–1930 гг. с помощью библиотеки анализа настроений Dostoevsky было проанализировано употребление сентимента в художественном тексте. Рассказы были поделены на две группы: рассказы от 1-го лица и рассказы от 3-го лица. Каждый рассказ в свою очередь был сегментирован на фрагменты двух типов: речь персонажей и речь автора, которые анализировались отдельно. Результаты исследования показали, что независимо от типа фрагмента и лица повествования доля текстовых фрагментов с нейтральным сентиментом превышает негативный, а позитивные фрагменты встречаются реже всего. Что касается интенсивности сентимента, наблюдается положительная корреляция между негативными и позитивными показателями авторской речи, независимо от типа повествования в рассказе.

Ключевые слова: русский рассказ; сентимент анализ; квантитативный анализ; речь автора; речь персонажей.

Введение

С развитием машинного обучения появляется все больше возможностей применения искусственного интеллекта (ИИ) в разных сферах человеческой жизни. В том числе стал возможным автоматический анализ сентимента в тексте. Наиболее широко сентимент-анализ с помощью применения ИИ используется для таких задач, как сбор данных отзывов пользователей на различные продукты и образовательные программы [Кирина, Тельнина 2022; Панасенко, Стукалова, Мазунина 2018; Рубцова 2013; Шимохин 2021], оценка эмотивной лексики в социальных сетях [Богданов, Дуля 2019; Боровых, Брагина, Плотников 2021; Колмогорова 2019], анализ мнений и прогнозирование реакций в журналистской и политической сферах [Василькова, Легостаева 2019; Суходолов, Бычкова, Ованесян 2019; Третьяков и др. 2018]. Помимо такого «традиционного применения» сентимент-анализа, можно отметить и ряд исследований, посвященных анализу с его помощью художественных литературных произведений как для исследования индивидуального стиля автора и соотношения героев и сюжета с эмоциональной составляющей текста, так и для сопоставления различных произведений, жанров и т. п. [Шерстинова и др. 2023; Jacobs, Kinder 2019;

Min, Park 2019: электр. ресурс; Mohammad 2011; Sherstinova et al. 2023].

Эмоции являются не только естественной реакцией на тот или иной раздражитель и побудителем к действию, но и незаменимым средством в общении с другими социальными объектами, отображающим отношение индивида к той или иной ситуации. Лексическая составляющая языка и речевого акта во многом является отображением нашего эмоционального состояния или субъективного восприятия ситуации. Эмоция, заложенная в словах, наделяет их дополнительной семантикой, которая и позволяет нам считать то или иное высказывание эмоционально окрашенным [Величкова 2007; Маслечкина 2015; Стаценко 2011; Шаховский 2019].

Если в дискурсе за формирование речевого высказывания и передачу эмоций с помощью использования характерной лексики, синтаксиса и т. п. ответственен говорящий, то в художественных литературных произведениях, которые находятся в фокусе нашего внимания, за повествование, речевые приемы и диалоги отвечают автор и рассказчик. Несмотря на особый вид взаимодействия с читателем и на то, что диалог в художественном произведении не идентичен речевому акту в реальности, «художественная реальность»

является проекцией реальности на художественный мир произведения. Ситуации в художественном мире «моделируют вербальное и авербальное поведение человека» [Шаховский 2008: 188] и так же, как и в речевом дискурсе, содержат эмоционально наполненный текст [Баженова 2004; Стефанский 2007; Шаховский 2008]. Некоторые общие когнитивные и коммуникативные законы сохраняются при попытке воспроизвести их в вымышленной ситуации.

Задачей предлагаемого исследования является сравнительный анализ эмотивности в речи автора и речи персонажей литературного текста на материале жанра рассказа. Кроме того, кажется целесообразным отдельно рассмотреть разные типы повествования (от 1-го и 3-го лица), чтобы определить, насколько повествование от лица героя художественного произведения отличается с эмоциональной точки зрения от нарратива от лица стороннего рассказчика. Главный вопрос, на который в ходе исследования предстояло получить ответ – зависит ли сентимент текста от типа повествования (от 1-го лица и от 3-го лица) и от типа речи (авторская речь и речь персонажей).

Метод и материал исследования

«Под сентимент-анализом понимается процедура идентификации и извлечения из текста эмоционально и оценочно окрашенного отношения говорящего или пишущего к предмету высказывания» [Колмогорова, Калинин, Маликова 2018: 140]. Основными подходами сентимент-анализа текстов с помощью компьютерных технологий являются словарный подход и машинное обучение [там же: 140–141]. При этом эти подходы в свою очередь делятся на словарный подход, основанный на правилах и на словарях, машинное обучение с учителем и без учителя. Словарный подход основан целиком на количественных показателях, т. е. результат зависит от того, какое количество негативных и позитивных слов было использовано в тексте. Словарный подход, при котором текст представляется как «мешок слов», не принимает во внимание контекст, что снижает релевантность результатов. На данных также может негативно сказаться отсутствие необходимых лексем в словаре, т. е. несоответствие словарного набора слов и лексического состава изучаемого текста. Если оценка отзывов в социальных сетях может быть проанализирована с помощью словарного подхода, то такой многосторонний и комплексный вид письменного текста, как художественное произведение, представляет большую трудность даже для моделей в машинном обучении. Тем не менее в настоящее время исследователи чаще делают выбор в пользу последнего.

В данном исследовании анализ тональности проводится с помощью модели Dostoevsky (<https://github.com/bureaucratic-labs/dostoevsky>) – библиотеки для сентимент-анализа в Python, для создания которой использовалась модель Fasttext и данные корпуса RuSentiment (<https://text-machine.cs.uml.edu/projects/rusentiment>) [Rogers et al. 2018]. RuSentiment – это текстовый аннотированный корпус, основанный на случайной выборке в 6 950 постов из социальной сети ВКонтакте. Благодаря разнообразию тем и разной длине публикаций на основе данных RuSentiment может быть проведен сентимент анализ различных типов текстов [там же]. Стоит признать, что лексическая составляющая социальных постов и художественных текстов совпадают не полностью, что является недостатком использования данной модели для анализа рассказов. Тем не менее по причине малого количества моделей, направленных на сентимент анализ именно художественных произведений, Dostoevsky стал вариантом, который больше всего соответствовал изначальным целям разделения эмотивных компонентов на негативные, позитивные и нейтральные.

В качестве материала была использована выборка в 70 рассказов, отобранных из Корпуса русского рассказа [Корпус русского рассказа: электр. ресурс; Sherstinova, Martynenko 2019] таким образом, что в половине рассказов (35) повествование ведется от 1-го лица, а в другой половине – от 3-го. В остальных рассказы были выбраны случайным образом. В данном корпусе собраны произведения 1900–1930-х гг., включающие в себя большое количество известных и малоизвестных писателей и произведения разнообразной тематики, позволяющие выявить как определенные сходства и направления в произведениях литературного творчества того периода, так и индивидуальные черты каждого из писателей [Мартыненко и др. 2018].

Текстовый материал каждой подвыборки исследовательского корпуса был в свою очередь был сегментирован на повествование от лица автора и прямую речь героев (140 текстовых фрагментов). Сегментация, разделяющая авторский нарратив и речь персонажей, была выполнена автоматически с помощью специальной компьютерной программы и затем отредактирована студентами образовательной программы «Филология» Высшей школы экономики в Санкт-Петербурге.

Каждый текст выборки с учетом сегментации на речь автора и речь персонажей был обработан с помощью библиотеки Dostoevsky. Анализ эмотивностей проводился по предложениям, то есть на выходе было получено предсказание модели для

каждого предложения. Библиотека Dostoevsky может присвоить один из 5 имеющихся тегов: *neutral*, *positive*, *negative*, *skip*, *speech*. Теги *positive* и *negative* указывают на содержание в предложении эмоциональной составляющей, а числовой показатель выражает вероятность. Вероятность можно сравнить с интенсивностью эмоции: чем сильнее выражена эмотивность предложения, тем выше будет вероятность соответствующего показателя (*neutral*, *positive*, *negative*). Например: *Зачем он думает об этом!* > {'neutral': 0.712, 'negative': 0.377}. Тег *speech* присваивается разговорным устоявшимся выражениям и не был рассмотрен в данном исследовании. Тег *skip* также не принимался во внимание, так как зачастую он означает, что программа Dostoevsky не смогла обнаружить в данном предложении вторую составляющую: *Что с тобой?* > {'neutral': 0.997, 'skip': 0.019} (в таких случаях значение тега часто оказывалось приближено к нулю). Помимо этого, тег *skip* присваивался предложению в том случае, если модель не смогла определить позитивность, негативность или нейтральность с высокой вероятностью: *Пана!* > {'positive': 0.665, 'skip': 0.484}. Так как тег *skip* не несет в себе информации о сентименте, он был исключен при дальнейшем рассмотрении результатов. В работу брались только теги *neutral*, *positive* и *negative*.

Таким образом, каждое предложение получило по два сентимент-тега с приписанной им вероятностью, выраженной числом от 0 до 1, т. е. предложение могло заключать в себе одновременно и негативный, и позитивный сентимент, или, например, и негативный, и нейтральный. В таком случае, оба сентимента учитывались при анализе результатов. Значения каждого сентимента для отдельных предложений были усреднены для каждого текстового фрагмента, таким образом, каждому фрагменту было сопоставлено три оценки (от 0 до 1) по шкалам *positive*, *negative* и *neutral*. В дальнейшей работе сравнивались именно эти средние значения сентимента по фрагментам, а не по отдельным предложениям.

На основании результатов сентимент-анализа была составлена статистика для каждого из проанализированных текстовых фрагментов и в целом для следующих четырех подвыборок: 1) рассказы от 1-го лица, речь автора; 2) рассказы от 1-го лица, речь персонажей; 3) рассказы от 3-го лица, речь автора; 4) рассказы от 3-го лица, речь персонажей. В Таблице 1 представлена часть сводной таблицы усредненных значений тегов для сегментов из слов автора рассказов от 1-го лица, где «количество» – число предложений с соответствующим тегом, «значение» – среднее значение (вероятность) сентимента.

Таблица 1

Усредненные значения тегов для рассказов от 1-го лица, слова автора (фрагмент)

Авторы	Количество			Значение		
	neutral	negative	positive	neutral	negative	positive
Алексеев Г.В	81	65	15	0,549	0,242	0,269
Давыдов Н.Н.	88	55	24	0,389	0,389	0,322
Дорошевич В.М.	78	80	33	0,546	0,369	0,360
Елпатьевский С.Я.	96	70	21	0,474	0,293	0,279
Фурманов Д.А.	86	75	20	0,526	0,311	0,289
Гарин-Михайловский Н.Г.	95	70	34	0,613	0,333	0,424
Амфитеатров А.В.	19	10	6	0,505	0,299	0,271

Результаты сентимент-анализа

Чтобы ответить на поставленные вопросы о закономерностях в эмотивности рассказов, были посчитаны корреляции для значений (вероятностей) сентимента и его «количестве» в текстах. Усредненный показатель *neutral* составил 0.6, *negative* и *positive* по 0.3, что в равной степени верно для слов автора и речи персонажей и для рассказов от 1-го и 3-го лица. Что касается «количества» сентимента, были рассмотрены как абсолютные значения, т. е. число предложений, так и средние доли, т. е. нормализованные по длине фрагмента значения с тем или иным сентиментом.

Наблюдается тенденция к преобладанию нейтральных предложений над негативными и

негативных над позитивными. Несмотря на то что в среднем модель присваивала наибольшую вероятность нейтральным предложениям и одинаковую – позитивным и негативным, предложений с негативным сентиментом оказалось больше, чем предложений с позитивным сентиментом.

Стоит отметить, что 7 из 140 фрагментов проанализированной выборки выбиваются из представленной выше тенденции и нарушают количественную закономерность «нейтральное > негативное > позитивное». После рассмотрения каждого из случаев несоответствия, были выявлены причины, по которым данная закономерность могла быть нарушена. Одной из причин мог стать ма-

льный объем рассказа, вследствие чего среднее значение по нескольким предложениям отражает особенность конкретных предложений, а не общей тенденции. Было замечено, что несоответствия чаще встречаются в отрывках с речью персонажей в малых текстах, так как количество предложений прямой речи значительно уступает количеству предложений во фрагментах от лица рассказчика.

Интересно отметить, что для фрагментов от лица рассказчика характерно значительное преобладание количества негативных предложений над позитивными. В рассказах от 1-го лица в авторской речи в 26 из 35 фрагментов количество негативных предложений преобладает над позитивными более чем в 2 раза, из них в 10 фрагментах – более чем в 3 раза. В рассказах от 3-го лица в авторской речи в 30 из 35 фрагментов количество негативных предложений преобладает над позитивными более чем в 2 раза, из них в 16 фрагментах – более чем в 3 раза. В то время как в прямой речи, несмотря на сохранение тенденции (negative > positive), встречаются рассказы, где разрыв между количеством позитивных и негативных предложений не так велик. Так, в рассказах от 1-го лица в прямой речи количество негативных предложений преобладает над позитивными более чем в 2 раза в 19 фрагментах, из них в 8 фрагментах – более чем в 3 раза. В рассказах от 3-го лица в прямой речи количество негативных предложений преобладает над позитивными более чем в 2 раза также в 19 фрагментах, из них в 6 фрагментах – более чем в 3 раза. Вероятно, это связано с тем, что в авторской речи эмоции выражаются более открыто, чем в прямой речи. Оценочной лексики в авторском повествовании присутствует больше, так как автор – не важно, является он сам участником событий или только сторонним наблюдателем, – занимает позицию оценивающего ситуацию и происходящее вокруг. Благодаря постоянной оценке и диалогу с читателем, слова автора оказались более насыщенными оценочной лексикой, чем прямая речь, в которой присутствуют характерные для речевого дискурса информативные высказывания: *Идемте на открытие клуба* {'neutral': 0,989, 'negative': 0,022} или *Не правда ли?* {'neutral': 0,998, 'skip': 0,025}.

Тем не менее один текстовый фрагмент не соответствует описанной выше тенденции: в нем

доля негативного сентимента (0,828) значительно превосходит долю нейтрального (0,463), причем это не связано с его объемом. Для сравнения, средняя доля нейтрального сентимента для фрагментов речи автора в рассказах от 3-го лица составляет 0,719, а средняя доля негативного сентимента – 0,507. Таким исключением оказалась повесть Л.Н. Андреева «Рассказ о семи повешенных», в которой описываются последние дни семи арестантов, приговоренных к смертной казни. Количество негативного сентимента значительно превышает количество нейтрального и это непосредственно связано с тематикой произведения. Опираясь на результаты частотного словаря лексем данного рассказа, можно заметить, что десять самых частотных нарицательных существительных связаны с темой смерти, жизни, казни или описывают части тела человека: *смерть, человек, рука, глаз, жизнь, день, цыганка, лицо, казнь, голова*. В самые частотные глаголы входят следующие: *быть, сказать, становиться, мочь, говорить, отвечать, думать, знать, вешать, начинать*. Здесь присутствуют только два активных действия – это *вешать* и *начинать*. Все остальные связаны или с мыслительными процессами (*думать, знать* и т. д.) или с общением (*говорить, отвечать*). В авторской речи присутствует подробное описание мыслей каждого из героев. Мрачная тематика произведения, выраженная на лексическом уровне, приводит к нарушению установленной нами тенденции, характерной для большинства рассказов эпохи.

Далее было проведено сравнение значений сентимента, который мы понимаем как интенсивность соответствующей эмоции. Сравнились корреляции по возможным комбинациям из трех типов сентимента (negative vs positive, neutral vs positive, neutral vs negative) для каждого сегмента: авторская речь в рассказах от 1-го лица, авторская речь в рассказах от 3-го лица, прямая речь в рассказах от 1-го лица и прямая речь в рассказах от 3-го лица. Наибольшая положительная корреляция, хотя и незначительная, была обнаружена между интенсивностью позитивности и негативности в авторской речи, независимо от типа повествования в рассказе – и в произведениях от 1-го лица, и в произведениях от 3-го лица (см. Табл. 2).

Таблица 2

Корреляции значения разных типов сентимента в каждом сегменте

Тип сегмента	Negative vs. positive	Neutral vs. negative	Neutral vs. positive
Авторская речь, 1 л.	0,638	0,317	0,482
Прямая речь, 1 л.	0,289	-0,129	0,049
Авторская речь, 3 л.	0,633	0,204	-0,010
Прямая речь, 3 л.	0,374	0,090	-0,070

Стоит обратить внимание прежде всего на то, что авторская речь объединяет повествование, в котором доминирует один нарратор. Прямая речь же подразумевает зачастую не только диалог между двумя героями, но и полилог, т. е. реплики принадлежат не одному «нарратору», а нескольким героям, каждый из которых поддерживает дискуссию. Так как речевые высказывания не были поделены на принадлежность опреде-

ленному герою, их смысловое содержание и эмоциональная наполненность являются разнообразными и не коррелируют друг с другом.

Кроме того, было проведено сравнение корреляции одинакового сентимента 1) в словах автора и прямой речи в рассказах от 1-го лица и 2) в словах автора и прямой речи в рассказах от 3-го лица. Однако оказалось, в данном случае никакой значимой корреляции не наблюдается. (см. Табл. 3).

Таблица 3

Корреляции количества сентимента одинакового типа в рассказах от 1-го и от 3-го лица

Тип повествования	Neutral	Negative	Positive
От 1-го лица	0,355	0,265	0,186
От 3-го лица	0,480	0,286	0,05

Было также принято решение проанализировать корреляцию между одинаковым сентиментом 1) в словах автора рассказов от 1-го лица и словах автора рассказов от 3-го лица и 2) в пря-

мой речи рассказов от 1-го лица и прямой речи рассказов от 3-го лица. Результаты показали, что и в этом случае корреляции почти не наблюдаются (см. Табл. 4).

Таблица 4

Корреляции значения сентимента в речи автора и персонажей

Тип речи	Neutral	Negative	Positive
Речь автора	-0,10	-0,10	0,22
Речь персонажей	-0,27	0,33	-0,01

Это указывает на то, что в диапазоне представленных средних значений для негативного и позитивного сентимента (0.2–0.4) существуют значительные колебания, которые зависят лишь от особенностей конкретного произведения и предпочтений автора.

Заключение

Проведенное исследование показало, что в русском рассказе исследуемого периода, как правило, соблюдается следующая тенденция: большинство предложений не маркированы с точки зрения эмоциональности, их можно считать нейтральными по сентименту, на втором месте по частоте – негативные предложения, в то время как позитивные встречаются в текстах реже всего. Нарушают обнаруженную закономерность лишь рассказы малого объема, в которых наблюдается уменьшение доли нейтральных предложений, а также рассказы с ярко выраженной негативной тематикой.

Интересное наблюдение заключается в том, что в авторской речи от 1-го и 3-го лица прослеживается корреляция интенсивности позитивности и негативности. Для речи персонажей такой корреляции не наблюдается. Это может объясняться тем, что из-за неоднородности высказываний персонажей, прямая речь является более разнообразной и менее сбалансированной по эмоциональности, чем авторское повествование. Соответственно, интенсивность негативного и

позитивного сентимента сильно варьирует в предложениях в отличие от авторского повествования, где единственным субъектом речи является сам рассказчик.

Корреляции между показателями одинакового сентимента, т. е. между значением нейтрального сентимента в прямой речи и значением нейтрального сентимента в авторской речи или между значением нейтрального сентимента в рассказах от 1-го лица и значением нейтрального сентимента в рассказах от 3-го лица, обнаружено не было, что говорит о том, что эмоции одного типа проявляются в каждом рассказе по-разному и не зависят от типа повествования.

Благодарности

Публикация подготовлена в результате проведения исследования по проекту «Текст как Big Data: моделирование конвергентных процессов в языке и речи цифровыми методами», выполненного в рамках Программы фундаментальных исследований НИУ ВШЭ в 2023 г.

Список литературы

- Баженова И.С.* Обозначения эмоций в художественном тексте (прагматический аспект): автореф. дис. ... д-ра филол. наук. М., 2004. 48 с.
Богданов А.Л., Дуля И.С. Сентимент-анализ коротких русскоязычных текстов в социальных медиа // Вестник Томского государственного университета. Экономика. 2019. № 47. С. 220–241.

- Боровых К.О., Брагина Д.С., Плотников А.В. Сентимент-анализ постов сообществ ВКонтакте // Актуальные научные исследования в современном мире. 2021. № 1–2(69). С. 66–69.
- Василькова В.В., Легостаева Н.И. Социальные боты в политической коммуникации. Вестник Российского университета дружбы народов. Социология. 2019. № 19(1). С. 121–133.
- Величкова Л.В. Психолингвистическая основа исследования эмоциональности звучащей речи // Вопросы психолингвистики. 2007. № 5. С. 20–26.
- Колмогорова А.В. Использование текстов жанра «Интернет-откровение» в контексте решения задач сентимент-анализа // Вестник Новосибирского государственного университета. Лингвистика и межкультурная коммуникация. 2019. № 17(3). С. 71–82.
- Колмогорова А.В., Калинин А.А., Маликова А.В. Лингвистические принципы и методы компьютерной лингвистики для решения задач сентимент-анализа русскоязычных текстов // Актуальные проблемы филологии и педагогической лингвистики. 2018. № 1. С. 139–148.
- Корпус русского рассказа 1900 – 1930 гг. [Электронный ресурс]. URL: <https://russian-shortstories.ru> (дата обращения: 22.10.2022).
- Кирина М.А., Тельнина Л.Д. Автоматическая оценка впечатлений обучающихся методами анализа тональности (на материале отзывов на онлайн-курсы на русском и английском) // Цифровая гуманитаристика и технологии в образовании (ДНТЕ 2022): сб. статей III Всерос. науч.-практич. конф. с междунар. участием / под ред. В.В. Рубцова, М.Г. Сороковой, Н.П. Радчиковой; Моск. гос. психол.-пед. ун-т. М., 2022. С. 355–374.
- Мартыненко Г.Я. и др. О принципах создания корпуса русского рассказа первой трети XX века / Г.Я. Мартыненко, Т.Ю. Шерстинова, Т.И. Попова, А.Г. Мельник, Е.В. Замирайлова // Труды XV Междунар. конф. по компьютерной и когнитивной лингвистике «TEL 2018». Казань, 2018. С. 180–197.
- Маслечкина С.В. Выражение эмоций в языке и речи // Вестник Брянского государственного университета. 2015. № 3. С. 231–236.
- Панасенко С.В., Стукалова И.Б., Мазунина Т.А. Направления развития современной розничной торговли. Российское предпринимательство. 2018. № 19(3). С. 599–608.
- Рубцова Ю.В. Метод построения и анализа корпуса коротких текстов для задачи классификации отзывов // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: тр. XV Всерос. науч. конф. RCDL'2013 / Ярославск. гос. ун-т им. П.Г. Демидова. Ярославль, 2013. С. 269–275.
- Стаценко А.С. Эмоционально-оценочная лексика как средство реализации речевой интенции. М.: Прометей, 2011. 118 с.
- Стефанский Е.Е. Культурные сценарии реализации эмоциональных концептов в художественном дискурсе // Вестник Самарской гуманитарной академии. Философия. Филология. 2007. № 1. С. 172–182.
- Суходолов А.П., Бычкова А.М., Ованесян С.С. Журналистика с искусственным интеллектом // Вопросы теории и практики журналистики. 2019. № 8(4). С. 647–667.
- Третьяков А.О. и др. Метод определения русскоязычных фейковых новостей с использованием элементов искусственного интеллекта / А.О. Третьяков, О.Г. Филатова, Д.В. Жук, Н.Н. Горлушкина, А.А. Пучковская // International Journal of Open Information Technologies. 2018. № 6(12). Pp. 99–105.
- Шаховский В.И. Лингвистическая теория эмоций. М.: Гнозис, 2008. 414 с.
- Шаховский В.И. Обоснование лингвистической теории эмоций // Вопросы психолингвистики. 2019. № 1(39). С. 22–37.
- Шерстинова Т.Ю. и др. Несчастливы по своему: как измерить тональность литературного текста? / Т.Ю. Шерстинова, А.Д. Москвина, М.А. Кирина, А.С. Карышева, Е.О. Колпащикова, П.И. Максименко, Р.А. Родионов, А.Р. Сейнова // Матер. XI Междунар. науч. конф. «Корпусная лингвистика 2023» (SPb Corpora 2023). (В печати.)
- Шимохин А.В. Семантический анализ отзывов о поставщиках на основе применения нейросетевой технологии // Фундаментальные исследования. 2021. № 5. С. 117–121.
- Jacobs A.M., Kinder A. Computing the Affective-Aesthetic Potential of Literary Texts // Artificial Intelligence. 2019. Vol. 1(1). Pp. 11–27.
- Min S., Park J. Modelling Narrative Structure and Dynamics with Networks, Sentiment Analysis, and Topic Modeling // PLoS ONE. 2019. Vol. 14(12). [Электронный ресурс]. URL: <https://journals.plos.org/plosone/article/file?type=printable&id=10.1371/journal.pone.0226025> (дата обращения: 11.09.2023).
- Mohammad S. From Once Upon a Time to Happily Ever After: Tracking Emotions in Novels and Fairy Tales // Proceedings of the ACL Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities / eds. K. Zervanou, P. Lendvai; Association for Computational Linguistics. Portland, OR, 2011. Pp. 105–114.
- Rogers A. et al. RuSentiment: An Enriched Sentiment Analysis Dataset for Social Media in Russian / A. Rogers, A. Romanov, A. Rumshisky, S. Volkova, M. Gronas, A. Gribov // Proceedings of the 27th International Conference on Computational Linguis-

tics / eds. E.M. Bender, L. Derczynski, P. Isabelle. Santa Fe, New Mexico, 2018. Pp. 755–763.

Sherstinova T., Martynenko G. Linguistic and Stylistic Parameters for the Study of Literary Language in the Corpus of Russian Short Stories of the First Third of the 20th Century // R. Piotrowski's Readings in Language Engineering and Applied Linguistics: Proceedings of the III International Conference on Language Engineering and Applied Linguistics

(PRLEAL-2019). St-Petersburg, 2019. Vol. 2552. Pp. 105–120.

Sherstinova T. et al. Sentiment Analysis of Literary Texts vs Reader's Emotional Responses / T. Sherstinova, A. Moskvina, M. Kirina, A. Karysheva, E. Kolpashchikova, P. Maksimenko, A. Seinova, R. Rodionov // 33th Conference of Open Innovations Association (FRUCT). Zilina, Slovakia, 2023. Pp. 243–249.

QUANTITATIVE COMPARATIVE ANALYSIS OF THE SENTIMENT OF CHARACTERS' AND AUTHOR'S SPEECH BASED ON THE MATERIALS OF THE RUSSIAN SHORT STORY CORPUS

Anastasia R. Seinova

Research Intern, Linguistic Convergence Laboratory

National Research University Higher School of Economics – St Petersburg

The article is devoted to measuring the emotional component of Russian small prose of the early XX century. On the basis of a random sample of 70 stories written in 1900–1930, with the help of the Dostoevsky mood analysis library, the manifestation of sentiment in the literary text was analyzed. The stories were divided into two groups: stories from the 1st person and stories from the 3rd person. Each story, in turn, was segmented into fragments of two types: the speech of the characters and the speech of the author, which were analyzed separately. The results of the study showed that regardless of the type of fragment and the face of the narrative, the proportion of text fragments with a neutral sentiment exceeds the negative one, and positive fragments are the least common. As for the intensity of sentiment, there is a positive correlation between negative and positive indicators of the author's speech, regardless of the type of narration in the story.

Keywords: Russian story; sentiment analysis; quantitative analysis; author's speech; characters' speech.