

СОЦИОЛОГИЯ КУЛЬТУРЫ

ЧТО СКРЫВАЕТ РУССКИЙ РЭП? ТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ТЕКСТОВ РУССКОЯЗЫЧНОЙ ХИП-ХОП СЦЕНЫ

Антон Евгеньевич Бойченко (anton@boychenko.pro),
Светлана Васильевна Жучкова

Национальный исследовательский университет «Высшая школа экономики»,
Москва, Россия

Цитирование: Бойченко А.Е., Жучкова С.В. Что скрывает русский рэп? Тематическое моделирование текстов русскоязычной хип-хоп сцены. *Журнал социологии и социальной антропологии*, 23(2): 130–165. <https://doi.org/10.31119/jssa.2020.23.2.6>

Аннотация. Представлены результаты эксплораторного анализа русского рэпа на основе корпуса текстов русскоязычных песен этого жанра. Корпус содержит более 11 000 текстов, собранных авторами посредством автоматического извлечения данных с веб-страниц (веб-скрапинга) и покрывающих более 500 исполнителей, творчество которых приходится на разное время и отличается разной степенью известности. Отталкиваясь от идеи о том, что медиа и музыка могут выступать агентом социализации, исследование ставит своей целью поиск тех нарративов, которые представлены в русском рэпе и которые могут оказывать социализирующее действие на многомиллионную аудиторию жанра, особенно на молодых людей. С помощью модели аддитивной регуляризации BigARTM проводится тематическое моделирование, в результате которого выделяются 17 основных тем, встречающихся в русском рэпе. Анализ результатов тематического моделирования показывает: среди всех тем наиболее распространены в рэпе нарративы поиска и становления себя, несчастной любви и смерти, а реже других встречаются размышления о родине и нарратив успеха. Для определения тех тем, которые транслируются наибольшему числу слушателей, анализируется распространенность полученных тем в текстах трех ключевых исполнителей российской хип-хоп сцены: Басты, Тимати и Oxxxymiron. С содержательной точки зрения результаты исследования представляют русский рэп в неожиданном свете, показывают его особенности по сравнению с рэпом, возникшим в США, и могут быть использованы в качестве источника гипотез для будущих исследований русского рэпа. С методологической точки зрения исследование становится обширной иллюстрацией возможности применения тематического моделирования в исследованиях социальных наук.

Ключевые слова: веб-скрапинг, интеллектуальный анализ текстов, русский рэп, тематическое моделирование, хип-хоп, BigARTM, text mining.

Введение

Последние десятилетия с развитием Интернета значительно увеличилось многообразие и доступность различных форм медиа: журналов, книг, музыки, подкастов, видео и пр. Большинство из них можно рассматривать как агентов социализации. Несмотря на традицию восприятия социализации как активного процесса, предполагающего взаимодействие с агентом (Schieffelin 1986), большинство медиа транслируют некое восприятие мира, нормы и ценности, которые могут сказываться на миропонимании социализируемого индивида. Особое влияние медиа может иметь на молодую аудиторию — детей и подростков, которые все еще находятся на этапе восприятия ключевых социальных норм и ролей, формирования собственной идентичности. Медиа, в отличие от традиционных агентов социализации — семьи, школы, церкви, в которых человек оказывается императивно, предоставляет возможность выбора. Человек сам решает, что смотреть, читать или слушать. В связи с этим мы можем наблюдать популярность одних форм и жанров медиа по сравнению с другими.

Одной из самых потребляемых форм медиа становится музыка, на прослушивание которой в среднем уходит около 2,5 часов в день (IFPI 2018), а у молодых людей (до 24 лет) почти 3 часа (AudienceNet & Music Biz 2018). На текущий момент самым популярным жанром из существующих является рэп, который в 2017 г. превзошел по популярности рок, более 50 лет державший пальму первенства среди музыкальных жанров (Nielsen 2017). 2018-й год показал, что рэп не теряет в своей популярности и лишь увеличивает отрыв от рока, закрепляя за собой статус самого прослушиваемого жанра в мире (Nielsen 2018). Рэп оказывается самым «влиятельным» жанром внутри одной из самых потребляемых форм медиа. Транслируя различные идеи и нарративы, рэп выступает агентом социализации для миллионов своих слушателей. При этом аудитория хип-хопа* молода — в США больше половины слушающих рэп младше 24 лет и около двух третей моложе 34 (Statista 2018), что дополнительно позволяет говорить о его потенциальном влиянии на молодое поколение и значимости как агента социализации.

Объектом настоящего исследования выступает русскоязычный рэп. В России рэп не менее популярен, чем во всем мире. Здесь рэп-треки

* Несмотря на то что понятия «рэп» и «хип-хоп» изначально имеют разные значения, где «хип-хоп» предполагает общую культуру, а «рэп» именно музыку, в данном исследовании они будут использованы как синонимичные, потому что треки принадлежат и к жанру рэп, и к хип-хоп культуре одновременно.

стабильно занимают лидирующие позиции в музыкальных чартах (iTunes 2019; Яндекс.Музыка 2019) и широко обсуждаются в СМИ (Комогорцев 2018; Ведомости 2018; Горбачев 2018). Помимо популярности рэпа в России важно заметить его политизацию (Фролова 2015). Рэп стал значимым политическим феноменом и инструментом политических элит: в текстах популярных исполнителей встречается как критика современной власти, которая выливается в запреты концертов и удаление клипов (Ver-shidsky 2018), так и ее поддержка и попытки дискредитации оппозиции (YouTube, канал «ЦАО Records»; YouTube, канал «BlackStar»). Заметно, что рэп действительно транслирует определенные идеи и взгляды, в частности политические, что позволяет убедиться в его социализирующей функции.

Тем не менее политическую повестку сложно назвать лейтмотивом российских рэп-треков — рэп состоит из большого количества поджанров (гангста-рэп, клауд-рэп, мамбл-рэп, рэп-кор, ньюскул и пр.) и включает сотни исполнителей, каждый из которых предлагает аудитории свое уникальное видение мира, свои идеи и взгляды. Трансляция этих идей аудитории, особенно молодой, формирует ее мировоззрение и социализирует ее. Цель настоящего исследования состоит в выявлении совокупности тем, которые присутствуют в русском рэпе, оценке их распространенности и описании исполнителей через призму полученных тем. Достижение цели становится возможным благодаря тематическому моделированию — эксплораторному методу, направленному на выявление латентных признаков (тем) внутри большого корпуса текстов.

Такой анализ интересен по ряду причин. Во-первых, выявление тем внутри корпуса текстов русского рэпа позволит обнаружить нарративы, которые русскоязычные рэперы предлагают своей обширной аудитории. Выявляемые темы, по сути, и являются воплощением существующих в рэпе нарративов, которые выступают одним из возможных инструментов социализации (Miller 1989). Понимание нарративов внутри русского рэпа позволит понять, как он может определять мировоззрение миллионов своих слушателей и какова его роль как агента социализации.

Во-вторых, тематическое моделирование текстов русского рэпа позволяет решить практическую исследовательскую задачу. База данных исследования достаточно велика и, по нашим оценкам, охватывает абсолютное большинство артистов жанра, включает тысячи текстов и сотни исполнителей, поэтому построение тематической модели позволит в дальнейших исследованиях глубже анализировать отдельные темы и нарративы внутри русского рэпа. Настоящее исследование дает возможность снизить проблему субъективности при выборе текстов и исполнителей

для анализа. Субъективность в выборе текстов для анализа оказывается важной проблемой и уже была отмечена в работах, касающихся изучаемого жанра (Hollander 2016). Дополнительную ценность этому результату добавляет то, что корпус содержит ряд малоизвестных исполнителей, которые невольно могут оставаться вне фокуса исследователей.

В частности, результаты настоящей работы откроют возможности для более глубокого анализа специфической для жанра лексики — позволят лучше понимать смыслы текстов, даже не имея при этом соответствующего бэкграунда. В тематическом моделировании отдельная тема выражается набором слов, рассмотрение которых в совокупности позволит лучше понимать смысл жаргонных, специфических слов и выражений.

В-третьих, тематическое моделирование — относительно новый метод анализа данных, появившийся в конце 1990-х годов в компьютерных науках (Papadimitriou 2000), а конкретный метод, используемый в настоящей работе, — только в 2015 г. (Vorontsov 2015). Работа позволяет раскрыть на эмпирическом примере возможности использования этого метода в социальных науках — той области, в которой применение продвинутых методов анализа данных только зарождается, складываясь в отдельную область вычислительных социальных наук (computational social science) (Lazer 2009). Каждый новый пример открывает для других исследователей ценный опыт и мотивирует к использованию нестандартных для социальных наук методов, тем самым увеличивая познавательные способности этих наук.

Настоящее исследование использует в качестве базы данных более 11 000 текстов песен русскоязычного рэпа. Сбор данных осуществлялся методом веб-скрапинга, т.е. автоматизированного извлечения данных с веб-страниц. Принадлежность собранных текстов к жанру рэпа конвенционально определяется пользователями тех сайтов, с которых была собрана информация. С помощью метода тематического моделирования производится анализ собранных данных, продуктом которого является тематическая модель русского рэпа. Модель включает в себя 17 различных по смыслу тем, описывающих сразу ряд нарративов, транслируемых через русский рэп.

Работа структурирована следующим образом: сначала мы обращаемся к истории рэпа как жанра, пониманию рэпа как культурного и социального феномена и рассмотрению музыки и рэпа, в частности, как агентов социализации, затем в работе предлагается подробное описание методологии сбора и анализа информации, за которым следует разбор результатов и выводы.

История и специфика рэпа как социального феномена

Для того чтобы наиболее полно понять сущность феномена русского рэпа, необходимо обратиться к истории рэпа как жанра, так как российские исполнители на протяжении развития жанра в России заимствовали идеи и паттерны поведения с родины рэпа — США (Фролова 2015).

Хип-хоп культура родилась в Нью-Йорке, в Южном Бронксе, в 1970-х годах (Hollander 2016). Долгое время рэп оставался культурой бедных афроамериканцев и являл собой репрезентацию их личного жизненного опыта (Stephens 2005). На заре своего существования рэп повествует о тяготах районной жизни и криминале. Первые десятилетия рэп представлял ряд трагических нарративов, связанных с криминалом, и стал своего рода рупором, гласящим о проблемах афроамериканского населения (Powell 1991; Smitherman 1997). Смерть, убийства, употребление наркотиков, алкоголя (Stephens 2005; Kubrin 2005; Hunnicutt 2009; Herd 2005) — те нарративы, которые находят свое отражение в ранних текстах жанра.

Попытка систематизации проблем, отраженных в рэпе, представлена у Дж. Холландера (Hollander 2016). На основе количественного контент-анализа автор приходит к выводу, что наиболее часто описываемая тема в американском рэпе — это проблема отчаяния, которая выражается в отражении тягостей жизни, лишений и неудач. За ней следуют проблемы насилия, смерти и наркотиков. В качестве же самых часто встречающихся акторов в текстах выступают Бог, знакомые и полиция. Проблемы, связанные с нарушением закона, в рэпе тесно переплетаются с репрезентацией городской и районной жизни. Жизнь «на районе» предполагает «правила улиц», которые также отражаются в текстах жанра (Hunnicutt 2009). Отчасти поэтому в рэпе зачастую встречаются указания на конкретные улицы, переулки, районы (Forman 2002).

Несмотря на ныне существующее мнение о том, что рэп перестал представлять действительный жизненный опыт и артикулировать проблемы населения (Hunnicutt, Andrews 2009), эта часть истории жанра позволяет видеть его как инструмент выражения мнения и отражения существующих проблем. Учитывая содержание обозначенных источников, мы можем предположить, что русский рэп также включает в себя трагические нарративы, темы употребления наркотиков, насилия и смерти.

Мы нарочно ограничиваем описание американского рэпа началом XXI в. Подобный экскурс в историю начала рэпа и его первичного развития позволяет понять, каким рэп пришел в Россию, где он появился в конце 1980-х годов. Тогда русский рэп продолжил традиции, заложенные

западными «коллегами»: в нем также фигурируют и активно артикулируются нарративы районной жизни, нищеты и криминала (Лапина 2015), которые мы увидели в истоках рэпа как жанра и которые особенно часто встречаются в гангста-рэпе («бандитском» рэпе) (Кошелева, Маслова 2018).

Для понимания того, какой русский рэп сейчас, обратимся к исследованию компании «Яндекс» (Яндекс.Исследования 2018), которое опирается на весь доступный компании корпус песен разных жанров и стремится определить наиболее и наименее характерные для жанра и его исполнителей слова. Характерность слов определялась сравнением частотности их использования в рэпе и в других жанрах. Например, исследователи показывают, что наиболее характерными для русского рэпа на момент проведения исследования являются существительные «рэпер», «флоу», «блант», «рэп», «братик». В остальных частях речи в основном встречаются слова с нецензурными корнями, что позволяет говорить о специфике используемой рэперами лексики. Наименее характерными оказываются «сирень», «синева», «разлука», «ромашка», «луг» — они встречаются значительно реже, чем в песнях других жанров. Преимуществом исследования компании является доступ к лицензионным материалам исполнителей, а также ко времени создания каждого трека, что позволяет проводить анализ в динамике и выявлять тенденции в развитии жанра через призму используемых в разное время слов.

Помимо содержания русского рэпа важно понимать его распространенность и исключительную популярность. Рэп в России (как и в мире) — один из самых популярных жанров музыки, аудитория его слушателей исчисляется сотнями тысяч (Nielsen 2017). В пользу особой популярности жанра говорят и лидирующие позиции рэп-треков в музыкальных чартах крупнейших онлайн-платформ для прослушивания музыки, таких как iTunes и «Яндекс.Музыка» (iTunes 2019; Яндекс.Музыка 2019). Популярность жанра делает его эффективным инструментом выражения взглядов и идей артистов и позволяет рассматривать как значимого агента социализации.

Таким образом, обратившись к истокам рэпа как жанра и его особенностям в России, можно сделать ряд выводов. Во-первых, рэп долгие годы является инструментом репрезентации личного опыта и выражения существующих социальных проблем. Несмотря на то что сейчас, предположительно, рэп теряет эту особенность, это его важная черта как жанра. Именно ее и перенимает русский рэп на заре своего существования, в конце 1980-х — начале 1990-х годов. Во-вторых, обнаруживается сходство в нарративах русских и американских рэперов. Проблемы нищеты, криминала и районной жизни, будучи основополагающими для жанра,

занимают важную роль и в русском рэпе. И наконец, в-третьих, русский рэп благодаря популярности сегодня можно рассматривать как важный и влиятельный инструмент трансляции идей артистов и воспринимать его как агента социализации.

Социализация через медиа

Настоящее исследование рассматривает рэп как агента социализации, поэтому особую важность приобретает разработка той теоретической перспективы, которая касается социализации. Несмотря на то что многие годы в социологии существует традиция восприятия социализации как активного взаимодействия двух агентов, восходящая к символическому интеракционизму (Mead 1934; Schieffelin 1986), здесь социализируемый индивид будет восприниматься скорее как пассивно воспринимающий идеи и нормы через потребление контента. Его агентность в данном случае будет сводиться исключительно к выбору медиа, которое он или она потребляет. В этом и заключается отличие медиа от традиционных агентов социализации, таких как семья, школа или церковь. Особую значимость эта специфика медиа имеет для молодых людей, которые еще не до конца сформировали свой взгляд на мир и на которых медиа может оказывать большее влияние. При этом выбор медиа расширяется, а влияние семьи в рамках процесса социализации уменьшается, что ставит подростка или ребенка перед самостоятельным выбором агента социализации. В частности, поэтому некоторые авторы апеллируют к этому процессу как к «самосоциализации» (Arnett 1995; Heinz 2002; Anderson 2012).

Учитывая возрастающее медиапотребление, связанное с развитием и распространением Интернета, все больше подростков и детей оказываются перед обширным выбором различных медиа, которые могут оказывать на них социализирующее действие. Особую нишу здесь занимает музыка, потреблению которой отводится значительная часть дня — в среднем около 2,5 часов в день (IFPI 2018). Причем молодые люди потребляют значимо больше музыки, чем более старшая аудитория (AudienceNet & Music Biz 2018). Таким образом, музыка становится одним из значимых агентов социализации молодого поколения.

Дж. Арнетт с опорой на опыт эмпирических работ в области потребления медиа выделил пять ключевых использований медиа подростками (Arnett 1995): развлечение, формирование идентичности, получение впечатлений, преодоление трудностей и идентификация с молодежной культурой. В контексте обсуждения социализации особенно важно обратить внимание на формирование идентичности. Например, благодаря

телевидению молодые люди воспринимают нормы романтических и сексуальных отношений (Brown 1990), а через журналы могут осваиваться гендерные роли (Evans 1991). Не меньшую роль в процессе социализации и формировании идентичностей играет и музыка (North 1999; Lloyd 2000; Tarrant 2002). В частности, рэп может определять восприятие подростками гендерных ролей и отношений между женщинами и мужчинами (Ward 2003; Bryant 2008).

Одним из инструментов социализации становится нарратив, с помощью которого агент социализации может доносить социализируемому индивиду какие-то идеи и нормы. П. Миллер и Б. Мур отмечают, что проще всего увидеть социализирующую функцию нарративов в мифах и сказаниях, «которые воплощают коллективную мудрость с целью просвещения и наставления молодых» (Miller, Moore 1989). Под нарративом здесь мы понимаем репрезентацию пережитого опыта (в случае рэпа — артиста или лирического героя) через повествование, историю (Richardson 1990). Рэп, как мы уже увидели, также создает свои нарративы и, как и вся музыка, обладает социализирующей функцией. Таким образом, обнаружив нарративы внутри рэпа, нам удастся осознать, каково то социализирующее действие, которое рэп может оказывать на свою аудиторию.

Безусловно, нельзя говорить о том, что медиа в целом или рэп в частности могут полностью определять мировоззрение индивида. Человек начинает слушать музыку, уже пройдя как минимум первичную социализацию, имея в сознании представления о социальных нормах и ценностях. Однако мы видим, что музыка, идеи и взгляды, через нее транслируемые, могут оставлять свой след в мировосприятии молодого поколения.

Методология исследования

Сбор и подготовка данных. Основным методом сбора данных (текстов песен) в исследовании стал веб-скрапинг — процедура автоматизированного извлечения данных с веб-страниц. Всего было собрано 11 396 текстов русских рэп-исполнителей. Источниками данных стали два сайта: «рэп-текст.рф»* (10 196 текстов) и genius.com (1200 текстов). Выбор такого метода сбора данных и конкретных сайтов для него обусловлен несколькими причинами. Во-первых, в масштабе тысяч потенциальных наблю-

* Сбор данных осуществлялся в 2019 г., и к моменту публикации работы сайт «рэп-текст.рф» прекратил работу. Тем не менее в открытом доступе находятся архивные версии страниц этого сайта за разные годы, их можно увидеть, например, на сервисах <http://web-arhive.ru/> или <https://web.archive.org/>.

дений использование других вариантов приписывания текста определенному жанру проблематично: изучение биографии, интервью и понимание самоидентификации сотен артистов требует досконального экспертного анализа, автоматизация которого невозможна. Выбранные же сайты предлагают наборы текстов, которые уже отнесены пользователями к жанру рэпа. Границы между жанрами в целом видятся достаточно расплывчатыми и частично субъективными, поэтому принятие пользователями ресурсов текста как принадлежащего к жанру позволяет делать вывод о некоей конвенции, на основании которой текст действительно может быть определен как принадлежащий к жанру рэпа. Во-вторых, объем данных на сайте «рэп-текст.рф» достаточно велик, содержит тысячи текстов и включает как относительно старые и непопулярные треки, так и достаточно известные. Несмотря на это, у сайта есть недостатки: на момент сбора данных в нем отсутствовали несколько важных артистов с большой аудиторией и большое количество текстов песен, выпущенных позднее 2016 г. Учитывая проблему, дополнительные 1200 текстов были собраны с сайта genius.com, который содержит тексты тех исполнителей, которые не были учтены на первом сайте, что позволило сделать базу данных более наполненной и репрезентативной. Помимо самих текстов песен были извлечены их названия и исполнители.

Предобработка собранных текстовых данных включала в себя три этапа, рекомендуемых при автоматизированном анализе текстов (Grimmer, Stewart 2013):

1) *токенизацию*, т.е. разбиение исходных текстов на наборы слов. Токенизация в нашем исследовании достигалась за счет приведения текстов к единому регистру, удалению небуквенных символов (знаков препинания и цифр), двойных пробелов. Кроме того, были удалены все буквы, не относящиеся к кириллице. Такая особенность предобработки обусловлена первичным анализом и построением тематических моделей, на основании которых удалось понять, что иностранные слова в текстах (1) встречаются достаточно редко и лишь искажают анализ, (2) не несут содержательного смысла в большинстве случаев, (3) объединяются в одну тему, интерпретация которой сводится к тому, что в нее включены иностранные слова.

2) *лемматизацию*, т.е. приведение слов к их начальной форме. Для осуществления этой задачи использовался пакет компании Yandex rumystem3, предназначенный для морфологического анализа русского языка. Преимущество этого инструмента состоит в способности с определенной точностью учитывать контекст, в котором находится слово (т.е. окружающие его слова), при выборе начальной формы.

3) *удаление стоп-слов*, т.е. слов, встречающихся в текстах достаточно часто, но не несущих при этом содержательного смысла. К таковым относятся предлоги, союзы, междометия, местоимения, числительные и т.п. В качестве источника стоп-слов использовался одноименный пакет на языке python — `stop_words`.

По завершении предобработки из базы были удалены дубликаты. Удаление дубликатов производится именно после предобработки, так как исходные тексты могут различаться технически (в регистре, пунктуации, количестве пробелов и пр.), но содержательно представлять идентичные наблюдения. Для программного обеспечения тексты даже при минимальных различиях будут считаться разными наблюдениями и входить в анализ несколько раз, искажая его результаты. Предобработка позволяет нивелировать подобные различия и выбрать из всей базы набор действительно уникальных текстов. Объем наблюдений после удаления дубликатов составил 11 103 песни. Своеобразные «дубликаты» были исправлены и в контексте исполнителей песен: поскольку тексты собирались с разных порталов, одни и те же исполнители могли быть записаны разным образом (например, Оксимирон и Охххуm1гon). Разные формы записи одних и тех же исполнителей были приведены к единому виду вручную. Таким образом, после предобработки в базе содержались тексты 507 исполнителей, а средняя длина текста составила 174 слова. Самые ранние тексты датируются началом 1990-х годов, а самые поздние — 2020 г. Таким образом, база данных охватывает почти 30 лет творчества российских рэп-исполнителей.

Важно отметить, что существует несколько ограничений предобработки в рамках настоящего исследования. Во-первых, поскольку тексты на сайты добавляются различными пользователями самостоятельно, в текстах могут встречаться орфографические ошибки, которые не учтены в используемых программных пакетах. Таким образом, два слова, несущие один и тот же смысл, будут восприниматься программой как разные, тоже искажая результаты. Несмотря на это, ошибки встречаются редко, так как зачастую тексты изначально собираются из официальных источников артистов. Во-вторых, некоторые слова обрабатываются таким образом, что искажается их изначальное контекстуальное значение. Например, одним из наиболее частых случаев является слово «бабки», имеющее два значения: множественное число слова «бабка» и разговорное слово «бабки» в значении «деньги». Поскольку используемый инструмент для лемматизации возвращает слова в единственном числе, он возвращает слово «бабка», однако значение «деньги» встречается гораздо чаще в контексте рэпа. Учитывая проблему, подобные случаи были вручную заменены с учетом первичного анализа контекста использования.

Отдельно стоит отметить, что после сбора и предобработки текстов припевы, которые повторяются несколько раз, сохранились в текстах. Припевы целенаправленно были оставлены в текстах, поскольку припев видится именно той составляющей текста песни, которая специально выбирается исполнителем для повторения и с большей вероятностью запоминается аудитории. Также именно припев может содержать «хук» — используемое несколько раз слово или выражение, которое чаще всего повторяется и запоминается (McLeod 2002).

Анализ данных. Как уже было упомянуто, основным методом анализа текстовых данных в исследовании стало тематическое моделирование. Тематическое моделирование направлено на выявление скрытых тем в большом корпусе текстов, где под темой при этом понимается набор связанных друг с другом слов. Концептуально тематическое моделирование схоже с факторизацией, часто используемой в социальных науках: если при факторизации стоит задача объединить множество исходных индикаторов в меньшее число латентных факторов, то в тематическом моделировании происходит то же самое, только в качестве индикаторов выступают отдельные слова или другие единицы текста (например, биграммы — сочетания из двух слов), а в качестве наблюдений — отдельные тексты. При этом данные, идущие «на вход» тематическому моделированию, могут быть представлены по-разному: они могут отражать наличие анализируемых единиц текст в каждом тексте (бинарные переменные), частоту встречаемости этих единиц в тексте (дискретные переменные) либо более сложные метрики, отражающие относительный вес слова в тексте и корпусе (например, метрика TF-IDF; непрерывные переменные). В нашем случае использовались только униграммы (отдельные слова, а не сочетания слов), а данные были представлены как частота встречаемости этих слов в текстах песен.

В качестве конкретной модели для проведения тематического моделирования была выбрана модель с аддитивной регуляризацией, реализованная в пакете BigARTM на языке python. Эта модель реализована как «конструктор»: она позволяет получить одно базовое решение с заданным количеством тем, к которому затем применимы дополнительные настройки — регуляризаторы, позволяющие увеличить контрастность тем, улучшить их интерпретируемость и т.д. (Aplishev et al. 2016). Следует отметить, что мы сознательно отказались от использования наиболее популярной модели тематического моделирования — латентного размещения Дирихле (Latent Dirichlet Allocation, LDA), так как эта модель не позволяет получить устойчивые результаты (т.е. одинаковые результаты при каждом перезапуске модели) из-за стохастического характера используемых в ней алгоритмов (Belford et al. 2018).

Результатом применения модели с аддитивной регуляризацией BigARTM становятся две матрицы: матрица, содержащая вероятности принадлежности каждого слова к каждой теме (матрица Φ), и матрица, содержащая вероятности наличия каждой темы в каждом тексте (матрица Θ). Первая используется для интерпретации тем (тема интерпретируется на основе слов, имеющих высокую вероятность принадлежности к ней), вторая — преимущественно для применения результатов тематического моделирования в последующем анализе, т.е. для использования полученных тем как новых содержательных переменных, хотя зачастую тексты, имеющие высокую вероятность наличия в них определенной темы, также привлекаются для ее интерпретации и иллюстрации.

Как и при факторизации, при тематическом моделировании необходимо заранее задать число тем, на которые модель разложит исходные данные. При этом, в отличие от факторизации, в большинстве современных методов тематического моделирования нет критерия для автоматического выбора этого числа (как, например, критерий Кайзера в факторизации). Поскольку никаких теоретических предпосылок к выбору конкретного числа тем в корпусе русского рэпа также не было обнаружено, для получения итогового решения были построены и проанализированы модели с количеством тем от 14 до 20 включительно, внутри каждой был применен ряд регуляризаторов (а именно регуляризаторы SparsePhi и SparseTheta, повышающие разреженность, или контрастность, матриц Φ и Θ соответственно). Критерием выбора итоговой модели стало качество интерпретации: было выбрано такое решение, в котором содержится минимальное количество трудно интерпретируемых тем и при этом наибольшее количество уникальных тем — таковой стала модель с 17 темами с применением регуляризатора SparsePhi, который делает полученные темы менее похожими друг на друга.

Матрица Θ содержит вероятности присутствия каждой темы в каждом тексте и была использована для оценки распространенности тем в корпусе. Матрица Φ , содержащая вероятности принадлежности слов к выделенным латентным темам, использовалась для интерпретации полученных тем. Темы интерпретировались в несколько этапов: вначале отбирались 30 слов с самыми высокими вероятностями принадлежности к теме (с помощью матрицы Φ), а также 5 текстов с наивысшей вероятностью встретить текущую тему в тексте (с помощью матрицы Θ). На основе этого набора слов и текстов, иллюстрирующих одну и ту же тему, а также имеющегося опыта прослушивания рэп-треков авторами проставлялись названия к каждой теме. Составленные таким образом названия затем выносились на обсуждение

относящихся к экспертам либо в области изучения рэп-культуры, либо в области применения тематического моделирования участников научного коллектива, к которому принадлежат авторы. В общей сложности в процедуре интерпретации тем участвовало шесть экспертов, включая авторов исследования.

Результаты

Тематическое моделирование. Итак, интерпретация тем в тематическом моделировании происходит на основе слов, входящих в нее с высокой вероятностью. В таблице 1 представлены 30 наиболее вероятных слов для каждой выявленной темы. Темы при этом упорядочены в порядке убывания распространенности, под которой мы понимаем среднюю вероятность встретить тему в тексте (т.е. среднее по всем вероятностям в матрице Theta), переведенную в проценты. Следует понимать, что выбранная модель с аддитивной регуляризацией исходно направлена на то, чтобы получаемые в итоге матрицы были довольно разрежены, т.е. содержали большое количество нулевых вероятностей. Именно поэтому представленные в таблице проценты достаточно низкие. В таблице также присутствуют доверительные интервалы для этих процентов, так как используемая база, скорее всего, не покрывает абсолютно весь существующий объем русского рэпа, т.е. представляет собой лишь выборочную совокупность. В тех случаях, где наблюдается пересечение доверительных интервалов, нельзя говорить о том, что какая-то тема более распространена в контексте всего русского рэпа, чем другая.

Таблица 1

Интерпретация и распространенность тем, выявленных в ходе тематического моделирования

Тема	Средняя вероятность встретить тему в тексте (доверительный интервал), %	Слова, характеризующие тему
Поиск и «становление» себя	12,07 (11,8–12,33)	знать, свой, верить, жить, друг, понимать, видеть, давать, становиться, никто, мир, идти, сила, говорить, делать, слово, думать, хороший, оставаться, сделать, душа, правда, бог, дело, путь, судьба, деньги, ничто, брат, место

Продолжение таблицы 1

Тема	Средняя вероятность встретить тему в тексте (доверительный интервал), %	Слова, характеризующие тему
(Несчастливая) любовь	11,31 (10,99–11,62)	любовь, любить, сердце, знать, глаз, слово, забывать, ночь, прощать, уходить, слеза, друг, сон, боль, чувство, душа, понимать, вместе, рука, оставаться, просить, становиться, весна, ла, последний, нужный, ждать, говорить, помнить, оставлять
Природа	8,43 (8,2–8,65)	небо, свет, солнце, лететь, ветер, звезда, мир, земля, огонь, белый, рука, глаз, гореть, снег, идти, вода, падать, видеть, сон, путь, море, волна, облако, птица, пусть, летать, давать, лето, свой, бежать
Смерть	6,74 (6,52–6,96)	кровь, тело, смерть, душа, бог, глаз, лицо, боль, черный, страх, видеть, земля, голова, война, ад, мертвый, брат, сила, вера, последний, крест, рука, свой, пыль, белый, мир, чей, свет, сердце, живой
Житейские истории	6,53 (6,32–6,74)	утро, друг, дело, спать, знать, минута, думать, завтра, ночь, звонить, работа, пара, вроде, вообще, самый, вечер, неделя, вчера, ладно, телефон, давать, начинать, забывать, час, становиться, писать, идти, пойти, сосед, просыпаться
Исполнение музыки	5,70 (5,52–5,88)	новый, музыка, игра, слово, свой, голос, оставаться, становиться, строка, мысль, рука, писать, играть, сцена, песня, самый, делать, зал, роль, стих, текст, лист, голова, главный, звук, сердце, куплет, слушать, путь, дело
Размышления о мире	5,55 (5,37–5,66)	мир, давать, искать, либо, дом, становиться, дно, свой, среди, ноль, ум, поиск, пусть, тысяча, сеть, находить, крутить, понимать, знать, заново, видеть, вдох, глаз, система, внутри, пустой, стена, воздух, менять, путь

Продолжение таблицы 1

Тема	Средняя вероятность встретить тему в тексте (доверительный интервал), %	Слова, характеризующие тему
Создание и чтение рэпа	5,44 (5,22–5,66)	рэп, хип, хоп, трек, давать, бит, читать, делать, свой, качать, мс, писать, микрофон, рифма, стиль, альбом, голова, текст, слово, слушать, говорить, становиться, знать, понимать, пара, дело, куплет, рука, парень, готовый
Жизнь на районе	4,95 (4,74–5,17)	че, пацан, брат, курить, давать, ак, рука, говорить, братан, витя, взять, мусор, план, бл*, хулить, дело, делать, хата, х*й, вася, район, дым, любить, голова, идти, ноггано, братик, минимум, карман, мутить
Город	4,88 (4,7–5,05)	город, улица, Москва, район, дом, дым, дорога, свой, знать, ехать, двор, ночь, машина, окно, центр, старый, столица, метро, большой, идти, поезд, спать, качать, дело, Питер, знакомый, вагон, домой, городской, круг
Вечеринки и секс	4,83 (4,63–5,03)	давать, любить, клуб, танцевать, девочка, знать, свой, тело, черный, парень, секс, танец, детка, подруга, рука, ночь, девчонка, дама, хороший, нравиться, бар, пить, смотреть, пьяный, делать, видеть, деньги, играть, отдыхать, утро
(Тяжелое) детство	4,74 (4,57–4,9)	мама, ребенок, помнить, сын, становиться, умирать, отец, мать, детство, уходить, папа, дом, жить, плакать, прощать, старый, школа, родной, смотреть, семья, вернуться, слеза, пить, сидеть, оставаться, бог, приходить, память, окно, вспомнить
Мат	4,73 (4,5–4,95)	рэп, *бать, х*й, сука, нах*й, бл*дь, рот, читать, свой, русский, жопа, бл*ть, говорить, знать, пох*й, лох, че, идти, парень, п*здеть, х*йня, гавно, давать, мудака, п*зда, крутой, бл*, п*дор, любить, рэпер

Окончание таблицы 1

Тема	Средняя вероятность встретить тему в тексте (доверительный интервал), %	Слова, характеризующие тему
Разборки	4,01 (3,84–4,19)	бить, рука, пуля, бой, закон, слово, стиль, дон, свой, подымать, готовый, нужный, за*бывать, пацан, голова, злой, враг, удар, давать, ствол, сила, убивать, держать, выстрел, бум, стрелять, патрон, грязный, сын, игра
Рэперские атрибуты	4,01 (3,77–4,25)	эй, сука, деньги, ха, нах*й, курить, сук, ай, знать, делать, нужный, дерьмо, дело, забирать, давать, видеть, пау, сделать, парень, самый, дно, пох*й, стиль, молодой, окей, новый, свой, бл*, па, бл*ть
Размышления о родине	3,06 (2,9–3,22)	россия, страна, русский, родина, власть, народ, война, красный, становится, тип, сдохнуть, америка, политика, обычный, голова, стадо, продавать, работать, смотреть, армия, экран, делать, пам, орел, думать, любой, нафиг, золото, автомат, душа
Успех	3,03 (2,88–3,19)	песня, переть, петь, самый, оу, слово, ой, кайф, чудо, стоп, голова, становится, мистер, дурак, ля, оп, говорить, пойти, шоу, палец, весело, дерьмо, тупой, джа, звать, гоп, жить, называть, смех, мода

Прежде чем мы перейдем к пониманию заключенных в темах идей и смыслов, стоит отметить особенность некоторых из них. Помимо тех тем, которые определяют некий нарратив, излагаемый в тексте, например любовь или детство, существуют такие темы, которые не дают понимания содержания того текста, который характеризуют, а описывают скорее его стиль. Это прежде всего темы «Мат» и «Природа». Описание текста через одну из этих тем не означает, что текст «рассказывает» про мат или повествует о природе. Так, в тексте с некоторой (ненулевой) вероятностью будет присутствовать тема «Мат», даже если матерные слова в нем встречаются в минимальном количестве, и также будет с некоторой (ненулевой)

вероятностью наличествовать тема «Природа», если, например, текст заключает в себе строку «Тихо, как падал снег, падал весь квартал». Слово «снег» в данном случае используется метафорично и не определяет тематику всего текста. При этом стоит отметить, что в теме «Мат», помимо непосредственно мата, встречаются такие слова, как «русский», «рэп», «читать», что может говорить о том, что тексты, характеризующиеся этой темой, содержат оскорбление других исполнителей и унижение их творчества, что характерно для культуры баттл-рэпа.

Перейдем к темам, которые возможно содержательно интерпретировать и выявить конкретный нарратив, заложенный в тексты, в которых эти темы содержатся. Самая распространенная тема в имеющемся корпусе (а учитывая разницу в доверительных интервалах с ближайшей к ней теме, и во всем рэпе) — это «Поиск и “становление” себя». Подобная интерпретации темы определяется набором слов, описывающих жизненный путь артиста/лирического героя, например словами «становиться», «путь», «судьба», «место», которые в совокупности с другими словами темы позволяют видеть в теме идею самоопределения. Для большей наглядности здесь и далее для каждой темы будет представлен отрывок текста песни, в котором вероятность обнаружить эту тему высока (вероятность из матрицы Theta). Например, иллюстрацией темы «Поиск и “становление” себя» может служить следующий отрывок песни «Хватит» коллектива «Marry Jane», в которой с вероятностью 99 % присутствует указанная тема:

*Каждый пытается мне что-то сказать
В надежде на то, что его начну слушать.
Я не вижу наперед и, может, так много
Совершаю ошибок на своей дороге,
Но это не значит, что можно исправлять меня.*

Если же при первичной интерпретации можно было лишь предполагать о смысловой нагрузке темы и тех текстов, которые она описывает, то рассмотрение конкретных текстов, вероятность встретить тему в которых очень высока, позволяет с большей уверенностью сказать о том, что она действительно посвящена поиску себя и жизненному пути исполнителя или лирического героя. Таким образом, самым распространенным нарративом русского рэпа становится нарратив жизненного пути и сюжет становления. Учитывая социализирующую функцию рэпа, такой нарратив может формировать у аудитории (особенно молодой) идею важности поиска себя, своего пути и места в жизни.

Следующая по распространенности тема — «(Несчастливая) любовь». Описание чувства любви видно достаточно отчетливо через слова «любовь» и «любить». Одновременно с этим в теме обнаруживаются слова «боль», «слеза» и «уходить», которые позволяют заметить сюжеты расставания и переживания негативного романтического опыта артиста и/или лирического героя. Рассмотрим отрывок текста, в котором с высокой вероятностью (94 %) присутствует тема несчастной любви:

*Я сжег все твои фото, пепел выбросил в окно,
И с кем ты теперь, мне абсолютно все равно.
Все нити наших отношений давно порвались между нами,
Я никого не виню в этом, мы виноваты в этом сами.
Помнишь наши обещанья, сотни теплых слов?
Мы когда-то называли это словом «любовь»*

(Noize MC, «Навсегда»).

Как и предполагалось на этапе анализа полученной совокупности слов, описывающей тему, действительно замечен нарратив расставания, любви, окрашенной скорее негативно, нежели позитивно. Принимая во внимание, что тема любви часто поднимается в различных формах искусства: живописи, скульптуре, литературе, поэзии, такая ее распространенность в текстах русских рэперов неудивительна. Вместе с этим важно понимать, что нарратив несчастной любви означает для социализирующей функции русского рэпа нормализацию выражения чувств, с одной стороны, и формирование представлений о романтических и сексуальных отношениях — с другой. Причем последнее уже отмечалось другими исследователями как часть социализации через музыку и рэп, в частности (Ward 2003; Bryant 2008).

Третья по распространенности тема — «Природа», о которой было сказано выше. Высокая распространенность этой темы также понятна, потому что использование слов, каким-то образом описывающих окружающий мир, может быть свойственно любому повествованию: такие слова могут как описывать обстановку, так и выступать в качестве метафор. О второстепенности этой темы по сравнению с другими говорит и относительно низкая максимальная вероятность присутствия этой темы в тексте по всему имеющемуся корпусу, а учитывая, что средняя вероятность встречаемости этой темы по корпусу довольно высокая, можно сделать вывод, что тема встречается часто, но не является основной в тех текстах, в которых присутствует. Рассмотрим отрывок с вероятностью 86 % встретить тему «Природа»:

*Чистого неба поровну,
Ведь все равно,
Когда уйду,
Небо грома раскатом ухнет,
И если звезды гвозди —
Снимите, пусть небо рухнет*
(Loc-Dog, «Даже если»).

Третьей именно содержательной темой является тема «Смерть». Помимо непосредственно слова «смерть» тема описывается такими словами, как «мертвый», «тело», «кровь», которые также показывают связь текстов этой темы со смертью. Текст песни «Милый прадед» исполнителя Loqiemean описывается названной темой на 87 % и позволяет ярко увидеть освещение в нем темы смерти и связанных с ней атрибутов: могилы, червей и пр.

*Милый, милый, милый прадед, как тебе там в могиле?
Роят туннели ли черви?
Я рад, что ты сдох, захлебнись в этой гнили,
Я собрал бы медали твои,
Нагрел и залил прямо в ямы глазниц, чтоб потухли огни,
А в аду, кроме трупов детей, тебя ярко слепили все ужасы тьмы*
(Loqiemean, «Милый прадед»).

Таким образом, в русском рэпе можно увидеть смерть как через ее материальные атрибуты, так и через описание состояний и связей с чем-то сверхъестественным, что видно через слова «бог», «ад», «вера». Нарратив смерти может сочетаться с нарративом поиска себя и своего места в жизни, дополнительно указывая на конечность жизни и необходимость осознания себя.

Появление таких нарративов, как несчастная любовь, смерть и поиск себя, могут быть частично объяснены связью русского рэпа с поэзией. Как поэзия полна размышлений о смерти, любви, своего места в мире, так и в рэпе, как мы видим, эти темы встречаются исключительно часто. Помимо этого, у русского рэпа и поэзии зачастую находят общие как содержательные, так и стилистические черты (Визель 2018; Молоков 2016), а сами рэперы зачастую цитируют классиков в своих строках (Яндекс 2019).

Четвертая по распространенности содержательная тема — «Житейские истории» — требует особенно внимательного рассмотрения. На первый взгляд кажется, что она, равно как и «Мат» и «Природа», представляет

собой скорее набор базовых слов, описывающих действия, и содержательной не является. При этом дальнейший анализ и выявление текстов, описываемых этой темой, позволили заключить, что совокупность слов описывает некую повседневность, рутину и житейские истории. В качестве подтверждения предложенной идеи обратимся к тексту, в котором вероятность присутствия житейских историй составляет 91 %:

*Когда я покупаю продукты, всегда беру кого-то туда —
Мне лень таскать все одному там.
И если это делаю, то только ночью,
Потому что ненавижу стоять в огромной очереди.
Я просыпаюсь ближе к вечеру, лень, конечно,
Но делать нечего, надо пойти поесть чего-нибудь...*

(Guf, «О лени»).

Как видно из текста, автор действительно говорит о некотором переживании повседневной рутины и ее описании, что позволяет говорить о верной содержательной интерпретации представленной темы. Таким образом, рэп в значительной мере становится инструментом репрезентации повседневности. Если изначально для жанра эта повседневность была наполнена криминалом, то теперь рэперы повествуют о бытовой рутине. Рэп, помимо «гласа народа», говорящего о социальных проблемах, является и чем-то повседневным, тем, что позволяет описать жизнь артиста или лирического героя в ее обыденности. Этот нарратив также важен и в контексте социализации, так как именно через него могут транслироваться нормы, следование которым характерно для повседневной жизни.

Дальнейшее описание каждой из тем по отдельности видится излишним, и мы лишь постараемся выделить основные идеи и нарративы, которые содержатся в русском рэпе. Отрывки текстов, в которых с наибольшей вероятностью обнаруживаются остальные темы, приведены в приложении 1.

Во-первых, стоит отметить те темы, которые описывают осознание себя и мира, а именно уже названная «Поиск и “становление” себя», а также темы «Размышления о мире», «Смерть» и «(Трудное) детство». Существование этих нарративов в корпусе позволяет говорить о том, что рэп, равно как и другие формы искусства, обращается к классическим темам смерти, души, семьи, своего места в мире и целей существования, определяя тем самым их восприятие для своей аудитории.

Здесь же отметим тему «Размышление о родине», которая также укладывается в идею (само)-рефлексии. Стоит отметить ряд ее особенностей.

Во-первых, она говорит не только о России, но и об ее отношениях с другими странами, как мы видим по словам «война» и «Америка». Во-вторых, такие слова? как «народ», «война», «сдохнуть», «политика», «власть», «стадо»? позволяют предположить наличие в текстах критики власти и народа. В ряде источников, обращающихся к хип-хоп культуре, отмечается, что рэп становится политическим инструментом (Perry 2004; Stapleton 1998; Фролова 2015), поэтому появление такой темы и в русском рэпе кажется вполне закономерным. Вопрос вызывает лишь то, что эта тема оказывается относительно мало распространена. Вероятнее всего, такая низкая распространенность объясняется тем, что политизация рэпа, как уже было отмечено, явление 2000-х годов, тогда как русский рэп существует еще с 1980-х. Роль также может играть то, что политическая обстановка в стране намного более ситуативна и темпоральна, нежели смерть, любовь или человеческое развитие, осмысление которых актуально вне времени.

В отдельную категорию хотелось бы выделить темы «Создание и чтение рэпа» и «Исполнение музыки», так как они описывают именно жизнь артистов внутри индустрии. Существование таких тем позволяет судить о важности для артистов идентификации себя как принадлежащих к рэп-культуре. В пользу этого тезиса также может говорить существование темы «Рэперские атрибуты», которая непосредственно показывает через текст понимание и использование специфических для жанра атрибутов или приемов. Говоря о теме «Рэперские атрибуты», стоит уточнить, что рэп — многогранное и гетерогенное явление, сводить которое под названную совокупность слов по меньшей мере затруднительно, а по большей — неправильно. Однако те слова, практики и приемы, которые указаны в этой теме, кажутся настолько стереотипно связанными с рэпом, что их обозначение как «рэперских» здесь будет вполне справедливым. Учитывая уже отмеченную Дж. Арнетт практику использования различных медиа подростками для идентификации с определенными культурами (Arnett 1995), появление этой темы важно в контексте социализации подростков, которые могут найти отражение субкультурных атрибутов и норм внутри творчества артистов. Обнаруживая эти культурные образцы, с которыми они могут себя идентифицировать, молодые люди могут с большей вероятностью принимать транслируемые через русский рэп нормы и ценности.

Описываемые ранее проблемы нищеты и криминала в текстах русских хип-хоп исполнителей объясняет появление в корпусе тем «Разборки» и «Жизнь на районе». Первая тема позволяет увидеть в текстах описание жестокости и физического насилия, вторая же — отсылки к криминальной

активности в текстах рэперов. Например, слова «курить», «мусор» (жаргонное слово для обозначения полицейского), «план» (жаргонное слово для обозначения марихуаны), «мутить», «дело» могут служить для описания употребления, хранения, продажи наркотиков и отношений с правоохранительными органами. Появление этой темы в итоговой модели также неудивительно, так как рэп долгий период своего существования поднимал темы криминала, особенно в рамках гангста-рэпа.

С темой нищеты ярко контрастирует тема «Успех». Из слов, характеризующих тему, видно, что она скорее всего повествует о материальных достижениях исполнителя/лирического героя. Успех здесь, вероятно, является следствием того жизненного пути, который проходит рэпер (в частности, того пути, который проходит через криминал). Нарратив успеха исключительно важен в рамках социализации, так как напрямую задает терминальные ценности в виде богатства, статуса и власти. Примечательно, что эта тема реже всех остальных встречается в нашем корпусе.

В продолжение идеи восприятия гендерных ролей и отношений между мужчинами и женщинами через музыку отметим тему «Вечеринки и секс», которая показывает важность демонстрации активного сексуального поведения, а также проявления высокой социальной активности и употребления алкоголя. Именно через эти маркеры могут задаваться границы мужского и может определяться роль мужчины.

Нарративы исполнителей. Так как исследование исходит из предпосылки о возможном социализирующем влиянии рэп-артистов на свою аудиторию, описание наиболее значимых и прослушиваемых из них позволит частично осознать, какие идеи транслируются наиболее широкой аудитории. Для решения этой задачи все вероятности встречаемости тем в текстах были агрегированы по отдельным исполнителям, т.е. для каждого исполнителя была рассчитана средняя вероятность встретить ту или иную тему в его песнях.

В качестве наиболее популярных и условно влиятельных исполнителей были выбраны трое: Баста, Тимати и Охxxуmгoп. Выбор именно этих артистов обусловлен рядом причин. Во-первых, они популярны. Под популярностью здесь мы будем понимать комплексный параметр, включающий в себя подписчиков в социальных сетях, количество прослушиваний на стриминговых сервисах («Яндекс.Музыка» или Apple Music), количество слушателей на концертах. Каждый из названных исполнителей имеет несколько миллионов подписчиков в социальных сетях (если брать наибольшее количество из нескольких социальных сетей: Instagram, VK, YouTube, Twitter), треки каждого из них стабильно занимают лидирующие позиции по прослушиванию среди рэп-исполнителей в названных сер-

висах. Каждый из них организовывал концерты на одной из самых больших концертных площадок в России, СК «Олимпийский», собирая более 20 тысяч человек, и все трое находятся в списке семи самых успешных русских рэперов по версии «Forbes» (Forbes 2019). Во-вторых, выбранные исполнители популярны уже достаточно продолжительное время, что позволяет говорить не об их сиюминутном (недельном, сезонном) влиянии, а об устойчивом многолетнем доминировании на рэп-сцене. И в-третьих, каждый из названных исполнителей является главой своего лейбла (объединения исполнителей), т.е., помимо того, что артист может иметь влияние на свою аудиторию непосредственно через свои тексты, он также может иметь влияние через своих подопечных. Стоит отметить, что возглавляемые ими лейблы также являются одними из самых популярных в стране (под популярностью лейбла здесь будем понимать то, насколько известны члены названных лейблов и как совокупно пользователями признается значимость таких объединений), что дополнительно подкрепляет выбор именно этих исполнителей.

Как видно из полученных данных (рис.), каждый из выбранных артистов характеризуется рядом ключевых тем. Видно, что с наибольшей вероятностью в текстах Басты встречается тема любви, в текстах Тимати — условно рэперские атрибуты, а в текстах Oxxxymiron — мат и оскорбления.



Рис. Распределение частотности тем по ключевым исполнителям

Интерпретация этих тем будет различаться. Доминирование нарратива любви в треках Басты подразумевает то, что с большой вероятностью он пишет/размышляет о любви. При этом же доминирование темы «Рэперские атрибуты» в текстах Тимати подразумевает скорее то, что в его текстах встречается набор условных маркеров, практик, которые характерны для хип-хоп культуры: употребление наркотиков, активное сексуальное поведение, материальная успешность — соответственно, «курить», «сука», «деньги». В пользу предположения о маркерах культуры также говорит слово «дерьмо», которым рэперы обозначают свое творчество (Instagram аккаунт @porimuxxxo). Наличие и высокая распространенность темы «Мат» в творчестве Oxxxymiron говорит о частоте именно этих слов и выражений. Также, как уже было сказано выше, тема «Мат» может описывать тексты, в которых оскорбляются другие исполнители и их творчество, что действительно часто встречается в творчестве этого исполнителя. Например, в треке «Я хейтер», который с 95%-ной вероятностью описывается темой «Мат», артист читает такие строки: «Я стакан, ты — стопка. Мы не одна тусовка. Твой рэп — подтасовка, мой рэп — потасовка». Из текста видно, что Oxxxymiron оскорбляет оппонента, в том числе указывая на его рэп. Подобные обращения к оппонентам, обобщенным или конкретным, часто встречаются в творчестве артиста.

Говоря о последнем исполнителе, отметим, что сравнение сразу трех артистов позволяет увидеть, что Oxxxymiron тематически отличается от двух остальных. Если у и Басты доминируют нарративы любви и поиска себя, то наиболее распространенные темы Oxxxymiron не пересекаются ни с одним из них. В текстах артиста поднимаются темы размышлений о мире и политике (тема «Размышления о родине»), а одной из самых распространенных является (редкая для всего корпуса) тема политики. Эту тему в творчестве артиста можно описать как политическую риторику, которая касается российских реалий, так и написанным им альбомом «Горгород», представляющим слушателям город-антиутопию и описание его политической жизни.

Подводя итог описанию выбранных популярных исполнителей, мы можем заключить, что темы любви и поиска себя действительно оказываются значимыми для русского рэпа. Помимо того что они оказываются самыми распространенными во всем корпусе текстов, они доминируют и в текстах ключевых исполнителей для русской рэп-сцены на данный момент, что позволяет говорить о том, что русский рэп может мотивировать размышления большей части своих слушателей о своей самости, месте в жизни и определять восприятие романтических отношений.

Заключение

Настоящее исследование представляет собой эксплораторный анализ доступного корпуса текстов русского рэпа, созданных с момента зарождения жанра. Рэп в последние годы становится одним из самых, а местами и самым популярным жанром музыки. Отталкиваясь от идеи о том, что медиа и музыка, в частности, может выступать агентом социализации, исследование ставит своей целью поиск тех нарративов, которые представлены в русском рэпе и которые могут оказывать социализирующее действие на многомиллионную аудиторию жанра, особенно на молодых людей.

Для достижения цели был применен метод тематического моделирования. В работе была использована модель BigARTM, претендующая на решение недостатков более популярных моделей. Базой для исследования послужил корпус текстов, собранный методом веб-скрапинга с сайтов «рэп-текст.рф» и genius.com. Использование обоих сайтов позволило собрать более 11 000 текстов, покрывающих значительную часть всей генеральной совокупности текстов рэп-треков, так как корпус включал в себя как новые и старые тексты, так и популярные и относительно неизвестные.

В качестве основного результата исследования была получена тематическая модель всего русского рэпа, включающая 17 тем, большинство из которых помогли обнаружить нарративы, транслируемые через русский рэп. Во всем корпусе были обнаружены (в порядке убывания распространенности) следующие темы: Поиск и «становление» себя, (Несчастливая) любовь, Природа, Смерть, Житейские истории, Исполнение музыки, Размышления о мире, Создание и чтение рэпа, Жизнь на районе, Город, Вечеринки и секс, (Тяжелое) детство, Мат, Разборки, Рэперские атрибуты, Размышления о родине, Успех. Как видно, самыми популярными оказываются темы поиска себя и любви. В условиях того, что рэп зачастую ассоциируется с проявлениями жесткости, нарративами нищеты или преступности, появление тем любви, размышления о самости и житейских историях кажется парадоксальным и представляет весь жанр в совсем новом свете. Также важно заметить, что многие из обнаруженных тем представляют нарративы, которые могут иметь социализирующую функцию. Так, нарратив несчастной любви позволяет передать восприятие романтических и сексуальных отношений, а тема успеха — задать определенные терминальные ценности.

Выявление общего набора тем и наиболее распространенных из них позволило сформировать понимание того, какие нарративы транслиру-

ются через русский рэп. При этом стоит понимать, что такая картина получена за счет всех найденных текстов, в том числе малоизвестных. Чтобы выяснить, какие идеи рэп транслирует наибольшему числу слушателей, в исследование включены наиболее характерные темы для трех самых популярных артистов российской хип-хоп сцены: Баста, Тимати и Оxxxymiron. Как показывает анализ, наиболее распространенными у Басты являются темы любви, поиска и становления себя и природы, у Тимати часто фигурируют рэперские атрибуты, тема становления себя и любви. Оxxxymiron же отличается от двух названных артистов: в его текстах с большой вероятностью встречается мат, размышления о родине и о мире в целом. Анализ самых популярных исполнителей еще раз позволяет понять, что нарративы любви и поиска, становления себя действительно оказываются ключевыми для русского рэпа.

Итогом исследования стало понимание того, что русский рэп в своем большом многообразии заключает внутри себя ряд различных, иногда нетривиальных, нарративов. Являясь самым популярным музыкальным жанром, рэп становится значимым агентом социализации молодого поколения, который через обнаруженные нарративы может (не всецело, но) формировать представление о гендерных ролях, жизненных целях, отношениях между женщинами и мужчинами, границах нормального.

Стоит отметить и ряд ограничений проведенного исследования. Во-первых, база данных сформирована интернет-пользователями, а не самими исполнителями, что порой может вызывать ошибки в текстах, как орфографические, так и те, которые связанные с отклонением от оригинального текста исполнителя. Этот же фактор позволяет предположить, что в базе могут быть не все опубликованные тексты песен русского рэпа. Во-вторых, учитывая, что качество тематической модели в основном определялось интерпретируемостью тем, а не математическими метриками, мы предполагаем существование и других валидных интерпретируемых моделей, которые могут раскрыть иные темы внутри русского рэпа. Тем не менее, как показывает сравнение моделей, построенных различными методами с использованием разных значений регуляризаторов и количества тем, содержательно модели различаются не настолько, чтобы поставить под сомнения результаты, полученные в рамках настоящего исследования.

В продолжение исследования мы предполагаем изучить динамику выявленных тем по годам. Настоящее исследование не представляет подобных результатов в силу отсутствия на текущий момент необходимых данных (о годе выпуска треков), которые предполагается в будущем найти и использовать. При этом анализ динамики тем позволит обнаружить

«эпохи» в развитии русского рэпа, что кажется исключительно интересным, учитывая изначально нероссийские истоки жанра и изменение политической и социальной ситуации в России. Также собранные данные открывают возможность проведения кластеризации исполнителей на основе распространенности выявленных тем в их творчестве, однако такой анализ выходит за рамки поставленных исследовательских задач работы.

Настоящее исследование носит разведывательный характер не только с содержательной, но и с методологической точки зрения. Исследования, использующие методы тематического моделирования в социальных науках, встречаются исключительно редко как в российском, так и в мировом академическом сообществе. Дальнейшие исследования также предполагают изучение вопроса в его методологических аспектах, например поиск более объективных метрик качества модели и предложение иных алгоритмов встраивания результатов тематического моделирования в последующий анализ.

Выражение благодарности

Исследование выполнено за счет гранта Российского научного фонда (проект № 19-78-10080).

Литература

Кошелева О.Н., Маслова А.Н. (2018) Из насмешек, угроз... *Вестник Челябинского государственного университета*, 6: 83–88.

Лапина Е.В. (2015) Опыт Анализа Поэтических Текстов Рока и Рэпа. *Глобализация науки: проблемы и перспективы: сборник статей Международной научно-практической конференции (3 апреля 2015 г, г. Уфа)*. Уфа: РИО МЦИИ Омега Сайнс: 79-83.

Фролова Е.В. (2015) *Рэп как форма социально-политической рефлексии в современной российской культуре (2009–2013 гг.)*. М.: Изд. Дом Высшей школы экономики.

Anderson L., McCabe D.B. (2012) A coconstructed world: Adolescent self-socialization on the internet. *Journal of Public Policy & Marketing*, 31(2): 240–253.

Apishev M., Koltcov S., Koltsova O., Nikolenko S., Vorontsov K. (2017) Additive Regularization for Topic Modeling in Sociological Studies of User-Generated Texts. In: Sidorov G., Herrera-Alcántara O. (eds.) *Advances in Computational Intelligence. MICAI 2016. Lecture Notes in Computer Science*, vol. 10061. Springer, Cham.

Arnett J.J. (1995) Adolescents' uses of media for self-socialization. *Journal of youth and adolescence*, 24(5): 519–533.

Belford M., Mac Namee B., Greene D. (2018) Stability of topic modeling via matrix factorization. *Expert Systems with Applications*, 91: 159–169. Doi: 10.1016/j.eswa.2017.08.047.

Brown J.D., Childers K.W., Waszak C.S. (1990) Television and adolescent sexuality. *Journal of Adolescent health care*, 11(1): 62–70.

Bryant Y. (2008) Relationships between exposure to rap music videos and attitudes toward relationships among African American youth. *Journal of Black Psychology*, 34(3): 356–380.

Evans E.D., Rutberg J., Sather C., Turner C. (1991) Content analysis of contemporary teen magazines for adolescent females. *Youth & society*, 23(1): 99–120.

Forman M. (2002) *The 'Hood Comes First: Race, Space, and Place in Rap and Hip-Hop*. Middletown, CT: Wesleyan University Press.

Grimmer J., Stewart B.M. (2013) Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, 21(03): 267–297. Doi: 10.1093/pan/mps028.

Heinz W.R. (2002) Self-socialization and post traditional society. *Advances in life course research*, 7(1): 41–64.

Herd D. (2005) Changes in the prevalence of alcohol use in rap song lyrics, 1979–97. *Addiction*, 100(9): 1258–1269.

Hollander J., Quinn J. (2016) More than noise: Employing hip-hop music to inform community development practice. *Community Development*, 47(5): 652–669.

Hunnicutt G., Andrews K.H. (2009) Tragic Narratives in Popular Culture: Depictions of Homicide in Rap Music. *Sociological Forum*, 24(3): 611–636. Doi: 10.1111/j.1573-7861.2009.01122.x.

Kubrin C.E. (2005) “I see death around the corner”: Nihilism in rap music. *Sociological Perspectives*, 48(4): 433–459.

Kubrin C.E. (2005) Gangstas, thugs, and hustlas: Identity and the code of the street in rap music. *Social problems*, 52(3): 360–378.

Lazer D., Pentland A., Adamic L., Aral S., Barabási A.L., Brewer D., ... Jebara T. (2009) Computational social science. *Science*, 323(5915): 721–723.

Lloyd B.T.N. (2000) Media influence on identity formation and social competence: Does music video impact adolescent development? *Dissertations available from ProQuest*. AAI9965521. [<https://repository.upenn.edu/dissertations/AAI9965521>] (дата обращения: 30.03.2020).

McLeod K. (2002) How copyright law changed hip hop: An interview with Public Enemy’s Chuck D and Hank Shocklee. *Stay Free Magazine*, 20 [<https://pdfs.semanticscholar.org/e227/2ef5ed9277c0ee63024cd811c9ec7b38eddc.pdf>] (дата обращения: 30.03.2020).

Mead G.H. (1962) [1934] *Mind, Self, and Society: From the Standpoint of a Social Behaviorist*. Chicago: University of Chicago Press.

Miller P.J., Moore B.B. (1989) Narrative conjunctions of caregiver and child: A comparative perspective on socialization through stories. *Ethos*, 17(4): 428–449.

North A.C., Hargreaves D.J. (1999) Music and adolescent identity. *Music education research*, 1(1): 75–92.

Papadimitriou C.H., Raghavan P., Tamaki H., Vempala S. (2000) Latent semantic indexing: A probabilistic analysis. *Journal of Computer and System Sciences*, 61(2): 217–235.

Perry I. (2004) *Prophets of the hood: Politics and poetics in hip hop*. Durham: Duke University Press.

Richardson L. (1990) Narrative and sociology. *Journal of contemporary ethnography*, 19(1): 116–135.

Schieffelin B.B., Ochs E. (1986) Language socialization. *Annual review of anthropology*, 15(1): 163–191.

Smitherman G. (1997) The Chain Remain the Same: Communicative Practices in the Hip Hop Nation. *Journal of Black Studies*, 28: 3–25.

Stapleton K.R. (1998) From the margins to mainstream: the political power of hip-hop. *Media, culture & society*, 20(2): 219–234.

Stephens R.J., Wright E. (2000). Beyond bitches, niggers, and ho's: some suggestions for including rap music as a qualitative data source. *Race and Society*, 3(1): 23–40.

Tarrant M., North A.C., Hargreaves D.J. (2002) Youth identity and music. *Musical identities*, 13: 134–150.

Vorontsov K., Frei O., Apishev M., Romov P., Dudarenko M. (2015) Bigartm: Open source library for regularized multimodal topic modeling of large collections. In: *International Conference on Analysis of Images, Social Networks and Texts*. Springer, Cham: 370–381.

Ward L.M. (2003) Understanding the role of entertainment media in the sexual socialization of American youth: A review of empirical research. *Developmental review*, 23(3): 347–388.

Источники

Визель М. История хабаровской школьницы Насти, рэпера Оксимирона и поэта Мандельштама, «взорвавшая» Рунет, ставит вопрос ребром: рэп — это поэзия или нет? *ГодЛитературы 2020*. [<https://godliterature.ru/public-post/гуер-роуезиуа-2>] (дата обращения: 30.03.2020).

Герасименко О., Фохт Е. «Тимати не переобувается»: история рэпера, верного Путину. *Би-би-си*. [<https://www.bbc.com/russian/features-50811886>] (дата обращения: 30.03.2020).

Горбачев А. Баста, Оксимирон и Noize MC дали концерт в поддержку Хаски. Его финал войдет в историю — посмотрите, как это было. *Meduza*. [<https://meduza.io/feature/2018/11/27/basta-oksimiron-i-noize-mc-dali-kontsert-v-podderzhku-haski-final-voydet-v-istoriyu-posmotrite-ego-zdes>] (дата обращения: 30.03.2020).

Комогорцев И. Прости, Окси. Лавров рассказал, что не любит рэп. *Life*. [https://life.ru/t/новости/1179017/prosti_oksi_lavrov_rasskazal_что_nie_liubit_rep] (дата обращения: 30.03.2020).

Молоков К. (2018) Рэп как альтернативная форма современной поэзии. *Новая Юность*, № 1. [https://magazines.gorky.media/nov_yun/2018/1/rep-kak-alternativnaya-forma-sovremennoj-poezii.html] (дата обращения: 30.03.2020).

Пост Instagram аккаунт @norimuxxo от 15.01.2020. *Instagram*. [<https://www.instagram.com/p/B7WECsGFg30/>] (дата обращения: 30.03.2020).

Почитай старших — кого цитируют русские рэперы. *Яндекс*. [<https://yandex.ru/promo/gigi-za-mozgi>] (дата обращения: 30.03.2020).

Путин призвал с осторожностью подходить к запретам рэп-концертов. *Ведомости*. [<https://www.vedomosti.ru/politics/news/2018/12/15/789383-putin-rep-kontsertov>] (дата обращения: 30.03.2020).

Русский рэп как набор слов. *Яндекс.Исследования*. [<https://yandex.ru/company/researches/2018/rap/>] (дата обращения: 30.03.2020).

Самые успешные рэперы России — 2019. Рейтинг Forbes. *Forbes*. [<https://www.forbes.ru/biznes-photogallery/380107-samye-uspeshnye-reper-y-rossii-2019-reyting-forbes?photo=6>] (дата обращения: 30.03.2020).

Треки, популярные на Яндекс.Музыке прямо сейчас. *Яндекс.Музыка*. [<https://music.yandex.ru/chart>] (дата обращения: 30.03.2020).

Чарты iTunes. *iTunes*. [<https://www.apple.com/ru/itunes/charts/songs/>] (дата обращения: 30.03.2020).

Audio Monitor US. THE OVERALL MUSIC LANDSCAPE 2018. *AudienceNet & Music Biz*. [https://musicbiz.org/wp-content/uploads/2018/09/AM_US_2018_V5.pdf] (дата обращения: 30.03.2020).

Bershidsky L. (2018) Russian Rappers Give Up on Putin. *Bloomberg*. [<https://www.bloomberg.com/opinion/articles/2018-12-14/russian-rappers-give-up-on-putin?srnd=premium-europe>] (дата обращения: 30.03.2020).

Favorite music genres among consumers in the United States as of July 2018, by age group. *Statista*. [<https://www.statista.com/statistics/253915/favorite-music-genres-in-the-us/>] (дата обращения: 30.03.2020).

IFPI Music Consumer Insight Report 2018. *IFPI*. [<https://ifpi.org/downloads/Music-Consumer-Insight-Report-2018.pdf>] (дата обращения: 30.03.2020).

U.S. Music Mid-Year Report 2018. *Nielsen*. [<https://www.nielsen.com/wp-content/uploads/sites/3/2019/04/us-midyear-music-report-2018.pdf>] (дата обращения: 30.03.2020).

2017 U.S. Music Year-End Report. *Nielsen*. [<https://www.nielsen.com/us/en/insights/reports/2018/2017-music-us-year-end-report.html>] (дата обращения: 30.03.2020).

Приложение 1

**Фрагменты песен с наибольшей вероятностью проявления тем
(для тех тем, которые не были рассмотрены в основном тексте)**

Тема	Отрывок текста	Выражен- ность темы в тексте, %
Исполне- ние музыки	Музыка по-настоящему только тогда бывает чистая, Когда, закрыв глаза, ты чувствуешь мысли. Ложится медленно на лист поэзия искристая, А воздухе, над головой, строки зависли. Мы от них зависимы, они словно выстрели. Подскажите мне, как можно управлять ими... <i>Fist, «В забвении»</i>	82,7
Размыш- ления о мире	Привет со дна, ты где-то там, где Свет и электрофанк, смех, Левый электорат, обед во фраках, светский раут, там нет меня, Здесь приём стеклотары, пакет отрав Едких, как винегрет с утра, нет контакта и недотрах; сто лет назад, Летом, нам так хотелось весь свет дотла нах*й сжечь, всю планету, Став прахом, пеплом и ветром — Как мне к тебе взять билет обратно без предоплат? <i>Oxxхuтигоп, «Привет со дна»</i>	85,7
Создание и чтение рэпа	Забив на все, люди бегут на запись, А мы, коты хорошие, им не покажем задницу. Пара нажатий клавиш, сведение, кэш... Делаем не ради денег, но за деньги, конечно. <i>Johnyboy, «Лапы вверх»</i>	90,2
Жизнь на районе	Брат роднуля, брат внатуре, Брат прикурит, как Везувий, Glock при зуме, И со сладенькой, как в изюме, Брат в костюме <i>Словесткий, «Брат»</i>	97,8

Тема	Отрывок текста	Выражен- ность темы в тексте, %
Город	<p>Ночь. Москва. Горят три вокзала. Мягкий вагон качает поезд Аврора. Снова я в соло под водкой спиртного, Нет разговора, я из питерского дома. Пролетаю пироны, листаю свой дневник, В мутном окне блик гонит проводник.</p> <p style="text-align: right;"><i>ШеFF, «Питер»</i></p>	83,5
Вечерин- ки и секс	<p>Танцуй со мною, детка, Блондинка и брюнетка. Танцуй со мною, детка, Блондинка и брюнетка. Танцуй со мною, детка, Гнись, как ветка, как конфетка, Липни ко мне, детка.</p> <p style="text-align: right;"><i>Dessar, «Детка»</i></p>	94,1
(Тяжелое) детство	<p>Помнишь, батя, о том, что есть сын? Я вот часто вспоминаю о тебе, отец. Помнишь, батя, о том, что есть сын? Хоть ты пил, батя, все равно ты молодец.</p> <p style="text-align: right;"><i>Лёша Маэстро, «Батя»</i></p>	72,1
Мат	<p>Дмитрий Бамберг, жидок *банный, ненавидит свою маму. На зоне его в туза *бали, как Кен *бал Барби. На хер насадили Димана нормально и сорвали резьбу. И пусть он лыс, как Мэтью Бут, но здесь я его мать *бу.</p> <p style="text-align: right;"><i>Babangida, «Мамбет»</i></p>	96,6
Разборки	<p>Верно, убийцы, плена, железных пчёл R'n'B. Попеременно, эй, слабачок, держись, Мы бледнолицы, словно, демоны, Из слов пускаем стрелы гнева.</p> <p style="text-align: right;"><i>Кажэ Обойма, «Два метода»</i></p>	89,5

Тема	Отрывок текста	Выраженность темы в тексте, %
Рэперские атрибуты	<p>Moonwalk-dance, мой флекс — это Ашер, Boy, мой Гуччи-бэг пахнет гашем (alright, what, alright, what) <...> Эй, эй, эй, эй, продал наркоту, эй Все, что я люблю, эй — Brand New Джимми Чу, эй Бой, мой флекс — это Ашер (Ашер, ха, Ашер, ха, е) Варим-варим воду; мы не варим эту кашу <i>Big Baby Tape, «Usher»</i></p>	99,9
Размышления о родине	<p>Могила, на могиле эпитафия: «Я умер, но бессмертна мафия». В кубышке капитала ГБУ Потные подмышки капитана ГРУ, Новые дворяне при погонах ФСБ, Сопли кокаина на казённых ДСП <i>Хаски (Husky), «Седьмое октября»</i></p>	85,1
Успех	<p>Количество писем — 2040, Личное звание — «Отец хардкора», Откуда родом — USA, Детройт, Подпись — «Опупенно крутой». Офигенно крутой? Обалденно крутой, Опупенно крутой <i>Мистер Малой, «Конфа»</i></p>	98,1

WHAT IS INSIDE RUSSIAN RAP? TOPIC MODELING OF THE TEXTS OF THE RUSSIAN-SPEAKING HIP-HOP STAGE

Anton Boichenko (anton@boychenko.pro),
Svetlana Zhuchkova

National Research University "Higher School of Economics", Moscow, Russia

Citation: Boichenko A., Zhuchkova S. (2020) Chto skryvayet russkiy rep? Tematicheskoye modelirovaniye tekstov russkoyazychnoy khup-khop stseny [What is inside Russian rap? Topic modeling of the texts of the Russian-speaking hip-hop stage]. *Zhurnal sotsiologii i sotsialnoy antropologii* [The Journal of Sociology and Social Anthropology], 23(2): 130–165 (in Russian). <https://doi.org/10.31119/jssa.2020.23.2.6>

Abstract. The study presents an attempt of the complex exploratory analysis of Russian rap based on the corpus of texts of the Russian-language songs of this genre. The corpus contains more than 11,000 texts that vary in their date of creation and popularity by more than 500 artists collected by automatically extracting data from web pages (web scraping). Basing on the idea that media and music, in particular, can act as an agent of socialization, the research aims to search for those narratives that are represented in Russian rap and which can have a socializing effect on the multi-million audience of the genre, and especially on young people. The result of topic modeling using the BigARTM additive regularization model is an extraction of 17 main topics of Russian rap. The analysis of the results of the topic modeling shows that among the narratives searching for life path, sad love and death are the most prevalent and those dedicated to homeland and success are the least. To reveal the topics that are transmitted to the largest number of listeners, the prevalence of the topics in the texts of the three key artists (Basta, Timati and Oxxxymiron) of the Russian hip-hop stage is analyzed. From a substantive point of view, the research sheds an unexpected light on Russian rap, shows its features that distinguish it from the American rap, and can be used as a source of hypotheses for the future research on Russian rap. From a methodological point of view, the study becomes an extensive illustration of the possibilities of applying topic modeling in social science research.

Keywords: web scraping, text mining, Russian rap, topic modeling, hip-hop, BigARTM.

Acknowledgements

The study was carried out with a grant from the Russian Science Foundation (project no. 19-78-10080).

References

- Anderson L., McCabe D.B. (2012) A coconstructed world: Adolescent self-socialization on the internet. *Journal of Public Policy & Marketing*, 31(2): 240–253.
- Apishev M., Koltcov S., Koltsova O., Nikolenko S., Vorontsov K. (2017) Additive Regularization for Topic Modeling in Sociological Studies of User-Generated Texts. In:

Sidorov G., Herrera-Alcántara O. (eds.) *Advances in Computational Intelligence. MICAI 2016. Lecture Notes in Computer Science*, vol. 10061. Springer, Cham.

Arnett J.J. (1995) Adolescents' uses of media for self-socialization. *Journal of youth and adolescence*, 24(5): 519–533.

Belford M., Mac Namee B., Greene D. (2018) Stability of topic modeling via matrix factorization. *Expert Systems with Applications*, 91: 159–169. Doi: 10.1016/j.eswa.2017.08.047.

Brown J.D., Childers K.W., Waszak C.S. (1990) Television and adolescent sexuality. *Journal of Adolescent health care*, 11(1): 62–70.

Bryant Y. (2008) Relationships between exposure to rap music videos and attitudes toward relationships among African American youth. *Journal of Black Psychology*, 34(3): 356–380.

Evans E.D., Rutberg J., Sather C., Turner C. (1991) Content analysis of contemporary teen magazines for adolescent females. *Youth & society*, 23(1): 99–120.

Forman M. (2002) *The Hood Comes First: Race, Space, and Place in Rap and Hip-Hop*. Middletown, CT: Wesleyan University Press.

Frolova Y.V. (2015) *Rep kak forma sotsial'no-politicheskoy refleksii v sovremennoy rossiyskoy kul'ture (2009–2013 gg.)* [Rap as a form of socio-political reflection in contemporary Russian culture (2009–2013)] Moscow: Izd. Dom Vyshey shkoly ekonomiki (in Russian).

Grimmer J., Stewart B.M. (2013) Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, 21(03): 267–297. Doi: 10.1093/pan/mps028.

Heinz W.R. (2002) Self-socialization and post traditional society. *Advances in life course research*, 7(1): 41–64.

Herd D. (2005) Changes in the prevalence of alcohol use in rap song lyrics, 1979–97. *Addiction*, 100(9): 1258–1269.

Hollander J., Quinn J. (2016) More than noise: Employing hip-hop music to inform community development practice. *Community Development*, 47(5): 652–669.

Hunnicut G., Andrews K.H. (2009) Tragic Narratives in Popular Culture: Depictions of Homicide in Rap Music. *Sociological Forum*, 24(3): 611–636. Doi: 10.1111/j.1573-7861.2009.01122.x.

Kosheleva O.N., Maslova A.N. (2018) Iz nasmeshkek, ugroz... [From mockeries and threats...] *Vestnik Chelyabinskogo gosudarstvennogo universiteta* [Bulletin of the Chelyabinsk State University], 6: 83–88 (in Russian).

Kubrin C.E. (2005) “I see death around the corner”: Nihilism in rap music. *Sociological Perspectives*, 48(4): 433–459.

Kubrin C.E. (2005) Gangstas, thugs, and hustlas: Identity and the code of the street in rap music. *Social problems*, 52(3): 360–378.

Lapina E.V. (2015) Opyt Analiza Poeticheskikh Tekstov Roka i Repa [An Experience of the Analysis of Poetic Texts of Rock and Rap]. In: *Globalizatsiya nauki: problemy i perspektivy: sbornik statey Mezhdunarodnoy nauchno-prakticheskoy konferentsii (3 aprelya 2015 g. g. Ufa)*. [Globalization of science: problems and prospects: International scientific-practical conference proceedings]. Ufa: RIO MTSII Omega Sayns: 79–83 (in Russian).

Lazer D., Pentland A., Adamic L., Aral S., Barabási A.L., Brewer D., ... Jebara T. (2009) Computational social science. *Science*, 323(5915): 721–723.

Lloyd B.T.N. (2000) Media influence on identity formation and social competence: Does music video impact adolescent development? *Dissertations available from ProQuest*. AAI9965521. [<https://repository.upenn.edu/dissertations/AAI9965521>] (accessed: 30.03.202).

McLeod K. (2002) How copyright law changed hip hop: An interview with Public Enemy's Chuck D and Hank Shocklee. *Stay Free Magazine*, 20 [<https://pdfs.semanticscholar.org/e227/2ef5ed9277c0ee63024cd811c9ec7b38eddc.pdf>] (accessed: 30.03.202).

Mead G.H. (1962) [1934] *Mind, Self, and Society: From the Standpoint of a Social Behaviorist*. Chicago: University of Chicago Press.

Miller P.J., Moore B.B. (1989) Narrative conjunctions of caregiver and child: A comparative perspective on socialization through stories. *Ethos*, 17(4): 428–449.

North A.C., Hargreaves D.J. (1999) Music and adolescent identity. *Music education research*, 1(1): 75–92.

Papadimitriou C.H., Raghavan P., Tamaki H., Vempala S. (2000) Latent semantic indexing: A probabilistic analysis. *Journal of Computer and System Sciences*, 61(2): 217–235.

Perry I. (2004) *Prophets of the hood: Politics and poetics in hip hop*. Durham: Duke University Press.

Richardson L. (1990) Narrative and sociology. *Journal of contemporary ethnography*, 19(1): 116–135.

Schieffelin B.B., Ochs E. (1986) Language socialization. *Annual review of anthropology*, 15(1): 163–191.

Smitherman G. (1997) The Chain Remain the Same: Communicative Practices in the Hip Hop Nation. *Journal of Black Studies*, 28: 3–25.

Stapleton K.R. (1998) From the margins to mainstream: the political power of hip-hop. *Media, culture & society*, 20(2): 219–234.

Stephens R.J., Wright E. (2000). Beyond bitches, niggers, and ho's: some suggestions for including rap music as a qualitative data source. *Race and Society*, 3(1): 23–40.

Tarrant M., North A.C., Hargreaves D.J. (2002) Youth identity and music. *Musical identities*, 13: 134–150.

Vorontsov K., Frei O., Apishev M., Romov P., Dudarenko M. (2015) Bigartm: Open source library for regularized multimodal topic modeling of large collections. In: *International Conference on Analysis of Images, Social Networks and Texts*. Springer, Cham: 370–381.

Ward L.M. (2003) Understanding the role of entertainment media in the sexual socialization of American youth: A review of empirical research. *Developmental review*, 23(3): 347–388.