

Песенные традиции XX века

УДК 398.86:004.42

DOI: 10.28995/2658-5294-2024-7-1-10-36

Квантитативный анализ частушек

Виктория Б. Новикова

*Российский государственный гуманитарный университет,
Москва, Россия, jranchik@gmail.com*

Маргарита А. Тихонова

*Российский государственный гуманитарный университет,
Москва, Россия, margarita.a.tikhonova@mail.ru*

Аннотация. В статье исследуется возможность применения метода кластеризации текстов для анализа фольклорного материала. С помощью программы Stylo анализируются два корпуса частушек из 239 и 645 текстов, собранных в фольклорной экспедиции ЦТСФ РГГУ в 2021–2023 гг. в Бирском районе Республики Башкортостан. В статье описаны принципы кластеризации текстов на графике метода главных компонент, проанализированы коллокации наиболее частотных слов. Показано, что они отличаются от коллокаций с этими же словами в обыденной речи (в сравнении с основным подкорпусом НКРЯ). Использование Stylo показало, что сформированные кластеры отражают наиболее характерную для определенного количества текстов лексико-синтаксическую структуру. Например, сочетания отрицательной частицы «не» с глаголами в частушечной речи преобладают над сочетаниями с именными формами, более характерными для обыденной речи, сочетания «а+я+глагол» и «а+я+местоимение» преобладают над сочетанием «а+я+служебные части речи», сочетания «на + существительное» или «на + прилагательное /местоимение + существительное» преобладают над сочетанием «на + местоимение», сочетание «в + существительное» преобладает над сочетанием «в + местоимение». Выделенные тенденции показывают свою устойчивость и прослеживаются при анализе корпуса как из 239 текстов, так и из 645. Наглядно показано разделение корпуса из 239 текстов на тексты с формами местоимения «я» и с обращениями. Выделены три распространенных варианта построения обращений в частушках:

© Новикова В.Б., Тихонова М.А., 2024

введение «нереального адресата», описание стереотипной ситуации, вовлечение слушающего.

Ключевые слова: частушка, Stylo, коллокации, частотные списки, обращения

Для цитирования: Новикова В.Б., Тихонова М.А. Квантитативный анализ частушек // Фольклор: структура, типология, семиотика. 2024. Т. 7. № 1. С. 10–36. DOI: 10.28995/2658-5294-2024-7-1-10-36

Quantitative analysis of chastushka

Victoria B. Novikova

*Russian State University for the Humanities,
Moscow, Russia, jranchik@gmail.com*

Margarita A. Tikhonova

*Russian State University for the Humanities,
Moscow, Russia, margarita.a.tikhonova@mail.ru*

Abstract. The article explores the possibility of applying text clustering methods for the analysis of folklore material. Using the Stylo program, two chastushka corpora of 239 and 645 texts are analyzed. These texts were collected during a folklore expedition by the Centre for Typological and Semiotic Folklore Studies of the Russian State Humanities University in 2021–2023 in the Birsky District of the Republic of Bashkortostan. The article describes the principles of text clustering, based on the graph of the principal component analysis; and the most frequent words collocations analysis. It is shown that these collocations differ from collocations with the same words in everyday speech (compared to the main subcorpus of the Russian National Corpus). The use of Stylo reveals that the formed clusters reflect the most characteristic lexical and syntactic structure for a specific quantity of texts. For example, combinations of the negative particle «не» (not) with verbs in the couplet speech prevail over combinations with nominal forms, which are more typical for everyday speech. Combinations like «а + я + глагол» (and + I + verb) and «а + я + местоимение» (and + I + pronoun) prevail over the combination «а + я + служебные части речи» (and + I + function words), combinations with «на + существительное» (on + noun) or «на + прилагательное / местоимение + существительное» (on + adjective/pronoun + noun) prevail over the combination «на + местоимение» (on + pronoun), and the combination «в + существительное» (in + noun) prevails over the combination «в + местоимение» (in + pronoun). The identified trends demonstrate their stability and are observed in the analysis of both the corpus of 239 texts and the corpus

of 645 texts. The article demonstrates the division of the corpus of 239 texts into texts with the pronoun «я» (I) and those with address forms. Three common variants of constructing address forms in couplets are highlighted: the introduction of an “unreal addressee,” the description of a stereotypical situation, and the involvement of the listener.

Keywords: chastushka, ditty, Stylo, vocative expression, collocations, frequency list

For citation: Novikova, V.B. and Tikhonova, M.A. (2023), “Quantitative analysis of chastushka”, *Folklore: Structure, Typology, Semiotics*, vol. 7, no. 1, pp. 10–36, DOI: 10.28995/2658-5294-2024-7-1-10-36

Преобладание коммуникативной функции частушки над всеми остальными – характерная черта этого фольклорного жанра. С.Б. Адоньева подчеркивает эту особенность, используя термины «частушечный коммуникативный акт», «частушечная речь» [Адоньева 2004, с. 161–162]. Для анализа частушек она использует инструментарий теории коммуникации, обращая внимание на коммуникативную ситуацию, намерение говорящего, модальность высказывания, его стилистический регистр. Авторы настоящей статьи разделяют ее точку зрения, однако исследуют не качественные параметры, а количественные. Мы считаем допустимым опустить контекст коммуникативной ситуации, так как анализируем корпус частушек, собранных не в живом бытовании, а исполненных в ответ на вопрос собирателя. Применяя методы компьютерной лингвистики, мы пытаемся выявить лексико-синтаксические особенности, характерные для этого фольклорного жанра.

Идея о том, что коммуникативные акты строятся по определенным моделям и число их ограничено и исчисляемо, для лингвистики не нова. Особенно плодотворно она развивалась во второй половине XX в., породив множество различных типологий, основывающихся на морфологических, синтаксических, семантических параметрах. В их числе, например, теория речевых актов Джона Сёрля [Сёрль 1986], теория «Смысл ↔ Текст» А.К. Жолковского и И.А. Мельчука [Жолковский, Мельчук 1965; Мельчук 1967], теория семантических примитивов Анны Вежицкой [Вежицкая 1999]. Еще в первой трети XX в. была опубликована работа В.Я. Проппа, посвященная морфологии волшебной сказки [Пропп 1928]. Во второй половине XX в. было выпущено еще два фундаментальных труда – работа Г.Л. Пермякова [Пермяков 1970] по систематике малых фольклорных жанров и «Типы народных сказаний» Б.П. Кербелите [Кербелите 2001], создавшей структурно-семантическую матрицу литовского повествовательного фольклора.

Материал и метод

В качестве пилотного корпуса мы выбрали 239 русских частушек, собранных в фольклорной экспедиции ЦТСФ РГГУ в 2022 г. в Республике Башкортостан. Корпус включает в себя около 3200 слов, при этом средняя длина каждого текста составляет 10–15 слов. Часть из них извлечена из глубинных интервью, часть записана в ходе уличных опросов. Яркой локальной специфики – диалектных, жанровых особенностей – тексты не имеют, что для нашего исследования является преимуществом, так как его результаты могут быть экстраполированы на другие корпуса.

Обработка текстов производилась программными средствами. В качестве инструмента для анализа мы выбрали программный пакет Stylo для языка программирования R [Eder and al. 2016]. Stylo активно применяется в стилеметрии для установления авторства текстов. Он позволяет работать с большими корпусами текстов, представлять результаты в виде наглядных графиков, при этом знание программирования необязательно. Мы пользовались основной функцией библиотеки «stylo()». Она создает список наиболее частотных слов (MFW – most frequent words) для всего корпуса, на основе которого строится первоначальная матрица стилистической близости всех текстов корпуса: каждая ячейка содержит рассчитанную относительную частоту слов в каждом соответствующем тексте.

Далее функция нормализует повторения, выбирая слова из желаемого диапазона частот для анализа, и создает итоговый список слов для фактического анализа, который преобразуется в данные для кластерного анализа, многомерного шкалирования и анализа главных компонент. В результате можно получить наглядное представление расстояний между текстами в виде графика.

Мы выбрали метод главных компонент (PCA – principal component analysis), так как именно он позволяет визуализировать набор данных с большим количеством переменных (в нашем случае это наиболее частотные слова в каждом отдельном тексте), при этом потери значимой информации минимальны. Доверие к методу также было продиктовано тем, что, во-первых, он уже применялся для анализа корпусов на русском языке, во-вторых, использовался при анализе фольклорных текстов. Среди авторитетных примеров можно назвать работу А.В. Козьмина [Козьмин 2009] по статистическому анализу сказочных указателей. Он обнаружил, что по количеству сюжетов одной группы (волшебные сказки, сказки о животных, глупцах и т. д.) можно предсказать количество сюжетов другой, при этом формулы, описывающие эту корреляцию, универсальны для разных традиций, а также выяснил, что

длина сказочного фрагмента напрямую коррелирует с его семантической нагрузкой – наиболее важное излагается дольше. Метод главных компонент использовался и Ю.Е. Березкиным [Березкин 2021] при исследовании корреляции распространения мотивов с использованием аналитического каталога мотивов мирового фольклора с основными маршрутами ранних миграций.

Работа с корпусом 2022 г.

Наше экспериментальное исследование мы начали с анализа 239 текстов. Нам казалось, что характерные для них синтаксические структуры будут более очевидны при сохранении исходных форм слов и словосочетаний. Полученные результаты уже были частично опубликованы нами [Новикова, Тихонова 2022], в этой статье мы попытаемся добавить к имеющимся выводам новые, а также верифицировать результаты исследования, применив ту же методику анализа к более крупному корпусу.

Как мы уже упоминали, в качестве метода анализа мы выбрали метод главных компонент (РСА (cov.)¹ – анализ главных компонент с использованием ковариационной матрицы, `analysis.type = "PCV"`). Далее на основании полученного графика были выделены кластеры и проведен их анализ.

На рис. 1 представлен график, полученный после загрузки корпуса в Stylo. Расположение обозначенных номерами текстов относительно друг друга отражает степень их стилистической близости: чем ближе тексты, тем больше слов из их частотных списков совпадает. В верхней половине графика четко визуализируются 4 «точки притяжения», вокруг которых образовались наиболее крупные и плотные кластеры. Два из них, обозначенные нами как

¹ РСА – метод понижения размерности признакового пространства с минимальной потерей данных (см.: *Shlens J.* A tutorial on principal component analysis. URL: <http://www.brainmapping.org/NITP/PNA/Readings/pca.pdf> [дата обращения 29 марта 2023]). В качестве входных данных в Stylo используется список наиболее частотных слов, из которого рассчитывается ковариационная матрица. Из нее выражаются собственные векторы, показывающие параметры схожести текстов в порядке убывания их важности. Ортогональные векторы используются для нахождения евклидова расстояния между текстами с помощью проекции текстов друг на друга для вычисления евклидова расстояния между ними. Такая проекция – это число, показывающее расстояние между парами текстов с наименьшей дисперсией. Далее по вычисленным расстояниям строится граф (тексты – вершины графа, ребра – их схожесть).

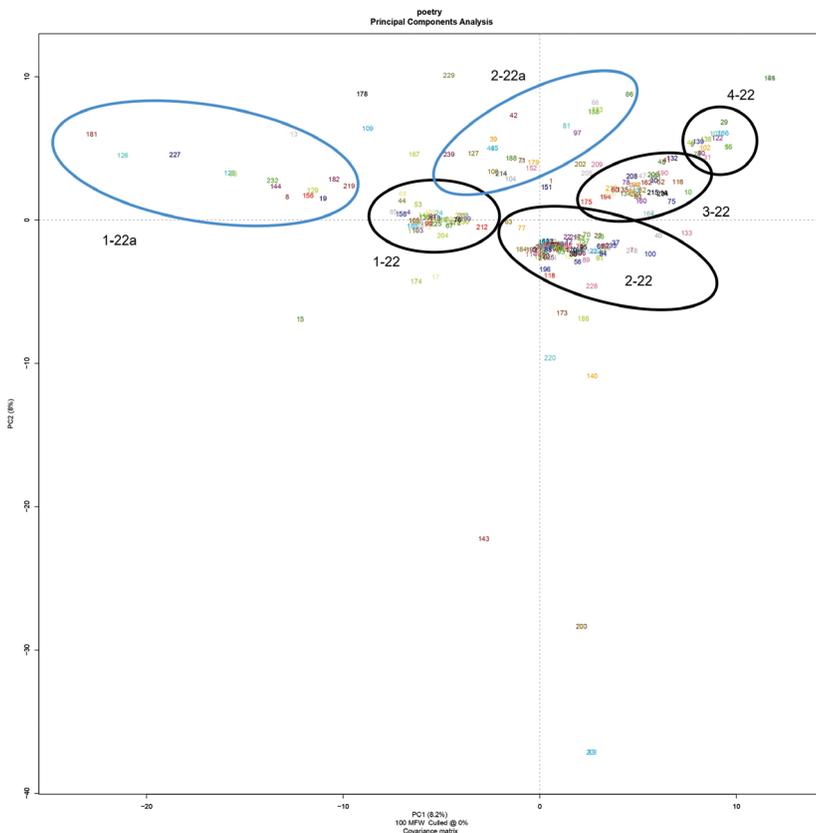


Рис. 1. Визуализация корпуса 2022 г. методом главных компонент

3-22 и 4-22, находятся в правой верхней четверти координатной плоскости, кластер 2-22 – в левой нижней, кластер 1-22 – на границе верхней и нижней левых четвертей. Также в левой верхней четверти и частично в правой верхней визуализируются две разнонаправленные оси, вокруг которых сгруппировались тексты, составляющие разреженные кластеры 1-22a и 2-22a (на графике они обведены синими овалами). Кроме этого, хорошо заметны тексты, находящиеся на периферии графика и не вошедшие ни в один из сформированных кластеров. Частотные списки этих текстов имеют наименьшее количество общих слов по сравнению с теми текстами, которые Stylo расположил близко друг к другу.

Поскольку график построен в результате многомерного математического расчета взаимного влияния частотности каждого

слова, необходимо интерпретировать результаты визуализации, опираясь на сами тексты и их частотные списки. Поэтому при анализе графика мы сравнивали интересующие нас тексты вручную.

В первую очередь мы решили рассмотреть периферийные тексты № 83, 84, 143, 200, 201. На фоне остальных они выделяются тем, что находятся максимально далеко от сформированных кластеров, однако довольно близко по отношению друг к другу. Как видно из табл. 1, их объединяет одинаковый зачин «Всё бы пела, всё бы пела», а также уникальный для этого корпуса набор частотных слов, среди которых очевидно преобладают наречие «всё» (19 словоупотреблений на 22 уникальных слова), частица «бы» (19 словоупотреблений) и глагол «пела» (10 словоупотреблений). Интересно, что среди самых частотных слов корпуса и кластеров (табл. 6) указанные лексемы не встречаются совсем.

Таблица 1

Частушки, наложившиеся
в нижней половине графика метода главных компонент

№ 143	№ 83	№ 201	№ 84	№ 200
Всё бы пела, всё бы пела, Кабы голос позволял, Всё бы времечко ревела, Кабы мил не унимал.	Всё бы пела, всё бы пела, Всё бы веселилася, Всё бы под низом лежала, Всё бы шевелилася.	Всё бы пела, всё бы пела, Всё бы веселилася. Всё бы под низом лежала, Всё бы шевелилася.	Всё бы пела, всё бы пела, Всё бы веселилася, Но одно меня сгубило: Молода влюбилася.	Всё бы пела, всё бы пела, Всё бы веселилася. Но одно меня сгубило, Молода влюбилася.

Следующей закономерностью, которую мы решили рассмотреть, стали случаи наложения. Среди группы периферийных текстов, о которых речь шла выше, нам встретилось 2 таких случая: наложившиеся тексты совпали дословно. Другие случаи наложения в верхней половине графика (табл. 2) также выявили тексты, либо полностью совпадающие (№ 164, 187), либо различающиеся одним-двумя словами: *такого* (№ 2) – *плохого* (№ 71); *повалил* (№ 16) – *положи* (№ 55); *маманя /мимо окон* (№ 87) – *мамаша /мим окошка* (№ 214); *разбила* (№ 141) – *отбила* (№ 176).

Таблица 2

Частушки, наложившиеся в верхней половине графика метода главных компонент

№ 2 Я, бывало, всем давала Сидя на скамеечке. Не подумайте такого, Из кармана семечки.	№ 16 Гармонист, гармонист Повалил меня на низ. А я встану, погляжу, Хорошо ли я лежу.	№ 87 Не ругай меня, маманя, Что сметану пролила, Мимо окон шёл Алёшка, Я без памяти была.	№ 141 Дробила я дробиночки, Разбила я ботиночки, И стою любуюся: Во что же я обуюся?	№ 164 Пела я, напелася, Ревела – наревелася. В ключике умылася, Кого люблю – лишилася.
№ 71 Я, бывало, всем давала, Сидя на скамеечке. Не подумайте плохого, Из кармана семечки.	№ 55 Гармонист, гармонист, Положи меня на низ. А я встану, погляжу, Хорошо ли я лежу.	№ 214 Не ругай меня, мамаша, Что сметану пролила. Мим окошка шел Алёшка, Я без памяти была.	№ 176 Дробила я дробиночки, Отбила я ботиночки. И стою, любуюся, Во что же я обуюся.	№ 187 Пела я, напелася, Ревела, наревелася. В ключике умылася, Кого люблю – лишилася.

Также на графике встречаются случаи, когда тексты не накладываются, но при этом располагаются очень близко друг к другу, объединяясь в группу по 2–3 (табл. 3). Несмотря на совсем небольшую разницу в расстоянии на графике, между случаями наложения и подобной группировки наблюдается одно значимое отличие: тексты совпадают уже не дословно, а лишь по одной-двум строчкам: *мне да не по карманчику* (№ 43) – *не по карману мне* (№ 45); *не ругай меня, маманя* (№ 87) – *не ругай меня, мамаша* (№ 108, 214); *я в вас не влюблялся* (№ 138) – *я и не влюблялся* (№ 223); *во мне всё шевелится* (№ 202) – *всё во мне шевелится* (№ 205, 209).

Таблица 3

Частушки, расположенные максимально близко друг к другу в верхней половине графика метода главных компонент

№ 43 Я пешком хожу По тrotуарчику, Жизнь такая мне Да не по карманчику.	№ 87 Не ругай меня, маманя, Что сметану пролила, Мимо окон шёл Алёшка, Я без памяти была.	№ 138 Мил измену предлагает, А я рассмеялася – Голубые ваши глазки, Я в вас не влюблялася.	№ 202 Говорят, что я – старуха, Что-то мне не верится. Ну, какая же я старуха, Во мне всё шевелится.
№ 45 Заплатить врачу Не по карману мне, Знать, останусь я Без зубов совсем.	№ 108 Гармониста я любила, Заругала меня мать. Не ругай меня, мамаша, Развеселый будет зять.	№ 2 23 Мил изменой угрожает, А я рассмеялася. Ну, какая, мил, измена? Я и не влюблялася.	№ 205 Ой, говорят, что я стара, Только мне не верится. Посмотрю я на себя – Да всё во мне шевелится.
	№ 214 Не ругай меня, мамаша, Что сметану пролила. Мим окошка шел Алёшка, Я без памяти была.		№ 209 Говорят, что я стара, Что-то мне не верится. Погляжу я на себя, Всё во мне шевелится.

Анализ текстов, расположившихся на графике близко друг к другу, но не образующих кластеры, показал некоторую противоречивость в работе Stylo: в нескольких случаях тексты, расположенные ближе к другу другу, различались по лексическому составу сильнее, чем тексты, разнесенные программой на более далекое расстояние. Так, например, пары текстов, приведенные в табл. 4, не имеющие ни одной общей фразы, были расположены ближе друг к другу, чем группа текстов № 202, 205, 209, совпадающих практически дословно. Следует отметить, что характерной особенностью

всех перечисленных текстов является то, что они расположены не только близко друг к другу, но и на некотором отдалении от остальных. Большинство из них не входит в выделенные нами кластеры. Это значит, что их лексический состав, а следовательно, и частотные списки достаточно специфичны по сравнению с исследуемым корпусом и кластерами. На этом этапе анализа мы предположили, что для текстов, сгруппированных более тесно (т. е. по кластерам), указанные параметры будут более единообразными.

Таблица 4

Частушки, расположенные близко друг к другу
в верхней половине графика метода главных компонент

№ 1 Меня милый не целует, Говорит: «Потом, потом». Я иду, а он на печке Тренируется с котом.	№ 7 Ох, тещу – грех, И невестку – грех. Я бы папину жену Через это шурану.	№ 65 Не ходи по коридору, Не шурши калошами. Все равно любить не буду, Рожа как у лошади.	№ 129 Вы не каркайте, вороны, Не чирикайте, скворцы. Дайте вымазать ворота И ещё вырвать огурцы.	№ 144 Ты, маганечка моя, Какая не сговорная – Не пустила ночевать, Ночь была холодная.
№ 151 Ты трепач, а я не знала, Ты откудава и чей, А теперь тебя узнала – Ты трепач из трепачей.	№ 218 Насолила я грибочков И капустки два бочка. На зимовку к этим бочкам Мне бы ещё старичка.	№ 128 Не ходите, девки, замуж, Замужем не до того. Не дадут четыре милого, Гляди на одного.	№ 156 Не ходите, девки, замуж, Замужем не с мёдом чай. Народятся голопузы, Только окай да качай.	№ 232 Полюбила гармониста, Не попала за него. Не хватило капитала У папаши моего.

Сопоставлять вошедшие в кластеры тексты вручную, как мы делали это с периферийными текстами, не кажется целесообразным, так как, во-первых, статистически значимые закономерности при таком анализе выявить гораздо сложнее, во-вторых, распо-

ложение текстов на графике указывает на их взаимное влияние и подсказывает исследовательскую логику: необходимо рассматривать тексты по группам, уделяя основное внимание их частотным спискам.

В табл. 5 представлены верхушки частотных списков корпуса, четырех основных кластеров (1-22-4-22) и двух разреженных (1-22а-2-22а). Для формирования списков частотных слов и их коллокаций мы использовали программу AntConc, она позволяет посмотреть не только самые частотные слова, но и количество их употреблений.

Таблица 5

Верхушка частотного списка кластеров корпуса 2022 г.

	Корпус	1-22	2-22	3-22	4-22	1-22а	2-22а
Кол-во текстов	239	41	85	36	14	14	21
Кол-во уникальных слов	1405	331	656	288	117	136	166
1	не (121)	не (42)	на (40)	я (37)	я (26)	не (36)	я (28)
2	я (118)	а (13)	а (20)	на (12)	а (7)	вы (4)	не (22)
3	на (80)	меня (12)	в (20)	а (8)	в (6)	у (4)	на (6)
4	а (56)	и (10)	ты (20)	за (8)	его (5)	что (4)	по (6)
5	и (49)	у (8)	и (16)	и (8)	мильный (5)	в (4)	и (5)
6	в (45)	да (7)	меня (15)	в (7)	гармонист (4)	жалко (3)	меня (5)
7	меня (45)	на (7)	с (14)	его (7)	жу (4)	замуж (3)	давала (5)
8	у (36)	в (6)	мильный (13)	у (7)	на (4)	только (3)	то (4)
9	ты (35)	вы (6)	у (11)	меня (5)	и (3)	девки (2)	что (4)
10	с (32)	мильный (5)	по (10)	он (5)	изменил (3)	женить (2)	а (3)

Таблица 6

Наиболее устойчивые коллокации
с самыми частотными словами корпуса 2022 г.

	не (121)	я (118)	на (80)	а (56)	и (49)	в (45)
1	не верит-ся (3)	я любила (6)	на меня (4)	а я (16)	и в (2)	в городе (2)
2	не же-нись (3)	я его (5)	на мель-ницу (3)	а теперь (3)	и вся (2)	в ключике (2)
3	не подумайте (3)	я на (4)	на низ (3)	а на (3)	и жалко (2)	в лесу (2)
4	не ругай (3)	я свою (4)	на свою (3)	а был (2)	и не (2)	в мазуте (2)
5	не ходите (3)	я, бывало (4)	на скаме-ечке (3)	а где (2)	и нигде (2)	в муку (2)
6	не будет (2)	я ему (3)	на столе (3)	а мой (2)	и никто (2)	в огороде (2)
7	не влюб-лялася (2)	я не (3)	на аэро-дром (2)	а мы (2)	и стою (2)	в пруде (2)
8	не жалко (2)	я с (3)	на берегу (2)	а нам (2)	и так (2)	в санато-рии (2)
9	не знаете (2)	я без (2)	на кой (2)	а он (2)	и я (2)	в армию (1)
10	не могу (2)	я боти-ночки (2)	на ней (2)	а потом (2)	и видела (1)	в брюхо (1)

График корпуса 2022 г. показал, что тексты группируются достаточно плотно в верхней половине графика. Тексты, попавшие в нижнюю половину и расположенные далеко относительно остальных, как мы показали выше, имеют отличающийся от корпуса набор частотных слов. Плотность расположения точек в верхней половине обусловлена тем, что разница в употреблении слов из верхушки частотного списка небольшая: все они употребляются сопоставимое количество раз – от 121 до 32. Такая же ситуация наблюдается и в основных кластерах: частотный список самого крупного и плотного кластера 2-22 содержит слова практически с равной частотой употребления – от 40 до 10 случаев. Остальные кластеры (1-22, 3-22, 4-22) являются менее плотными, потому что разница в количестве употреблений

самого частотного слова и следующих за ним более значительна. Наконец, наибольшая разница между верхним и последующими частотными словами наблюдается в разреженных кластерах 1-22а и 2-22а. Можно предположить, что слова, входящие в частотные списки кластеров, также зависят друг от друга и составляют коллокации, характерные для определенного типа частушек. Примеры, которые мы приведем далее, это доказывают.

Так, отрицательная частица «не» является самым употребительным словом в целом по корпусу, а также в кластерах 1-22, 1-22а, 2-22а. Можно предположить, что именно на этом основании данные кластеры находятся ближе друг к другу относительно остальных и лежат в верхней левой четверти графика. Любопытно, что «не» встречается практически с одинаковой частотностью во всех строках частушек, кроме первой (такой случай встретился всего однажды). Почти всегда за «не» следует глагол или существительное. Эта же тенденция прослеживается по всему корпусу (121 употребление): «не» следует за глаголом в 88%, на остальные 12% выпадают сочетания с именными формами: «не жалко», «не надо», «не стыдно», «не по карману», «не моя пизда», «не до того», «не с медом чай», «не рожь», «не с вашего села». Для сравнения: в Национальном корпусе русского языка на 100 самых частотных случаев совместного употребления «не» с различными частями речи только 68 приходится на «не» с глаголами. Остальные случаи выпадают на сочетания с наречиями (*надо, нужно, очень, менее, совсем, слишком, сразу* и др.), местоимениями (*то, тот, такой, все* и др.), предлогами, категорией состояния (*надо, нужно*), модальными словами (*должен*). Такая высокая распространенность сочетания «не» с глаголами в нашем корпусе могла бы быть объяснена сравнимой частотой этого сочетания в НКРЯ, однако в нем довольно частотны сочетания с частями речи, которые не встречаются в нашем корпусе вообще. Это наводит на мысль о специфических моделях употребления «не» в частушках нашего корпуса и кластеров 1-22, 1-22а, 2-22а.

Кластеры 3-22 и 4-22 располагаются наиболее близко на основании лидирующих позиций в их частотных списках личного местоимения «я» и сочинительного союза «а». При этом, как показывает анализ их словоупотреблений, эти лексемы практически всегда встречаются в текстах рядом, составляя триграммы «а я + глагол», «а я + местоимение». Если обратиться к НКРЯ, то мы увидим, что подобные конструкции имеют гораздо меньший показатель *ipm* (*instances per million words* – количество употреблений на миллион слов), чем сочетания «а я + не», «а я + в», «а я + и» и т. д., практически не встречающиеся в кластерах 3-22 и 4-22. Очевидно, это оказывается особенностью вошедших в них частушек.

Наконец, обращают на себя внимание служебные части речи. Предлог «на» занимает первое место по частотности употребления в кластере 2-22, второе – в 3-22 и третье – в целом по корпусу. Из 80 употреблений по корпусу в 56 случаях он входит в сочетания «на + существительное» (*на столе, на автобусе, на берегу, на чердак*), в 10 – в сочетания «на + прилагательное /местоимение + существительное» (*на худой кобыле, на скорой помощи, на Кольском полуострове, на самом дне*). Эти сочетания в подавляющем большинстве являются обстоятельством места. Остальные случаи употребления выпадают на сочетания «на + местоимение» (7) и «на + числительное» (1) (*на одного*). Сопоставление с основным корпусом НКРЯ показывает, что сочетание «на + местоимение» в текстах на русском языке гораздо более распространено, чем наиболее частотные для нашего корпуса сочетания «на + существительное» или «на + прилагательное /местоимение + существительное».

Нельзя также не упомянуть предлог «в». Он занимает третье место по частотности употребления в кластерах 2-22 и 4-22 и шестое – в кластере 3-22 и по корпусу в целом. Практически все коллокации с «в» включают в себя существительное, в единичных случаях за предлогом следуют прилагательное, местоимение или глагол. Большинство коллокаций (25 из 45 по корпусу) с «в» выполняют в предложении роль обстоятельства места, около трети составляют дополнения (10) и определения (8), 2 раза встречается обстоятельство места. Схожее распределение синтаксических ролей наблюдается и в отдельных кластерах – практически всегда сочетания «в + существительное» являются обстоятельством места. Интересно, что в основном корпусе НКРЯ половина из 10 самых частотных случаев употребления «в» с другими частями речи выпадает на местоимения, существительные стоят на втором месте по частотности (3 из 10 случаев).

«Ручной» анализ самых частотных слов корпуса и отдельных кластеров, а также коллокаций, в которые они входят, подтвердил применимость функционала Stylo для работы с частушками. Построенный график является хорошей отправной точкой для выделения групп текстов и анализа текстов внутри них.

Были сделаны следующие выводы.

1. Есть корреляция между плотностью или разреженностью кластеров и разницей в количестве употреблений самого частотного слова (и слов, следующих за ним): чем больше разница, тем более разреженный кластер.

2. Кластеры, выделенные вручную на графике, основанном на списке частотных слов, не однородны и позволяют наглядно увидеть как схожие тексты, так и отличающиеся. Анализ коллокаций

служебных слов (которые являются наиболее частотными) позволяет выявить устойчивые синтаксические структуры, не совпадающие со структурами обыденной речи и характерные для речи частушечной.

3. Верхушки частотных списков как корпуса, так и основных кластеров в основном составляют разные конфигурации одних и тех же слов (см. табл. 5). Отличающийся набор частотных слов характерен для разреженных кластеров, их частотные списки различаются как между собой, так и в сравнении с частотными списками корпуса и плотных кластеров. По всей видимости, разреженные кластеры стянуты друг к другу на графике частицей «не», но разделены по наличию (2-22а) или отсутствию (1-22а) «я».

4. В каждом кластере есть уникальные слова, которые не встречаются в верхушках частотных списков других кластеров, например: «да», «ты», «за», «его», «жалко», «что» и «давала». Из них только слово «ты» входит в самые частотные слова корпуса (35 употреблений), остальные встречаются в кластерах довольно редко (3–8 употреблений в кластерах).

Особенности обращений в частушках

Рассматривая полученные с помощью программной обработки группы текстов, мы заметили, что выделенные Stylo синтаксические структуры, очевидно, порождают отличия и на уровне семантики. Так, большинство частушек из кластера 4-22 построены как повествования от первого лица, например: *Не ругай меня, мамаша, // Что сметану пролила. // Мим окошка шёл Алёшка, // Я без памяти была* (№ 214). Тем не менее очень схожие частушки повествования, но без местоимения «я» были отнесены к третьему кластеру: *Мама чаю, мама чаю, // Мама чаю не хочу // Во дворе стоит мальчишка, // Познакомиться хочу* (№ 61).

В обеих приведенных выше частушках есть частица «не», которая не является частотной для кластера 3-22, но ключевая для 1-22. Данный пример хорошо показывает ограничения Stylo – при отсутствии ключевого частотного слова текст не попадает в кластер со схожими по лексике текстами.

Нас удивило, что в частотном списке второго и третьего кластеров «я» не встречается совсем. Более пристальный взгляд на тексты позволил заметить схожесть построения частушечного высказывания на уровне семантики. Многие тексты этих кластеров содержат обращения или выполняют апеллятивную функцию, т. е. подразумевают реакцию слушающего на текст. Важно, что все частушки записаны в ходе глубинных интервью, поэтому ни обра-

щения, ни апеллятивы (как правило, глаголы в повелительном наклонении) не выполняют своей прямой функции. Их функцию скорее можно назвать эмотивной: обращения и апеллятивы позволяют ярко и компактно описать ситуацию, которую легко представить. Такие тексты выстраиваются одним из трех способов:

1. Введение «нереального» адресата. В этом случае, как правило, обращаются к предметам:

Вы потопайте, ботиночки,
Вам больше не плясать.
Выйду замуж – буду плакать,
Вы – на полочке лежать (№ 121).

2. Описание стереотипной ситуации. Обращение позволяет сэкономить языковые средства и емко обрисовать ситуацию, знакомую слушающему. Следующая частушка отлично иллюстрирует потенциал такого способа: в ней не используются обращения, а апеллятивная функция реализуется за счет одного глагола в повелительном наклонении («пой»):

Песни пой, пока поётся,
Выйдешь замуж – не певать.
Попадёт свекровка злая,
Будешь песни забывать (№ 105).

3. Вовлечение слушающего. Использование обращений направлено на сокращение дистанции между адресантом и адресатом высказывания.

Не любите военных,
А любите шоферов.
От военных толку мало,
А шофёр навозит дров (№ 235).

Последний способ реализуется с помощью использования устойчивых обращений: *гармонист, подружка /подружки*. Примечательно, что пол слушающего не имеет значения во время произнесения текста, частушку про подружку исполняли и собирателям мужского пола:

А вы, подружки, друг у дружки
Зачем отбиваете?
Нелегко ведь расставаться,
Разве вы не знаете? (№ 115)

Работа с более крупным корпусом

Для верификации полученных выводов мы расширили корпус, добавив в него полевые материалы, собранные на той же территории в 2021 (272 текста) и 2023 (133 текста) гг. Итоговый корпус составил 645 текстов. На рис. 2 представлен график, полученный после его обработки в Stylo методом PCA. Тексты распределились в 4 плотных кластера и 3 разреженных, на периферии осталось небольшое количество текстов. Примечательно, что практически все периферийные тексты – добавленные: тексты, находившиеся на первом графике на периферии, на графике нового корпуса были включены в состав новых плотных кластеров. Кластеры корпуса из 239 текстов сохранили свои очертания, дополнившись новыми

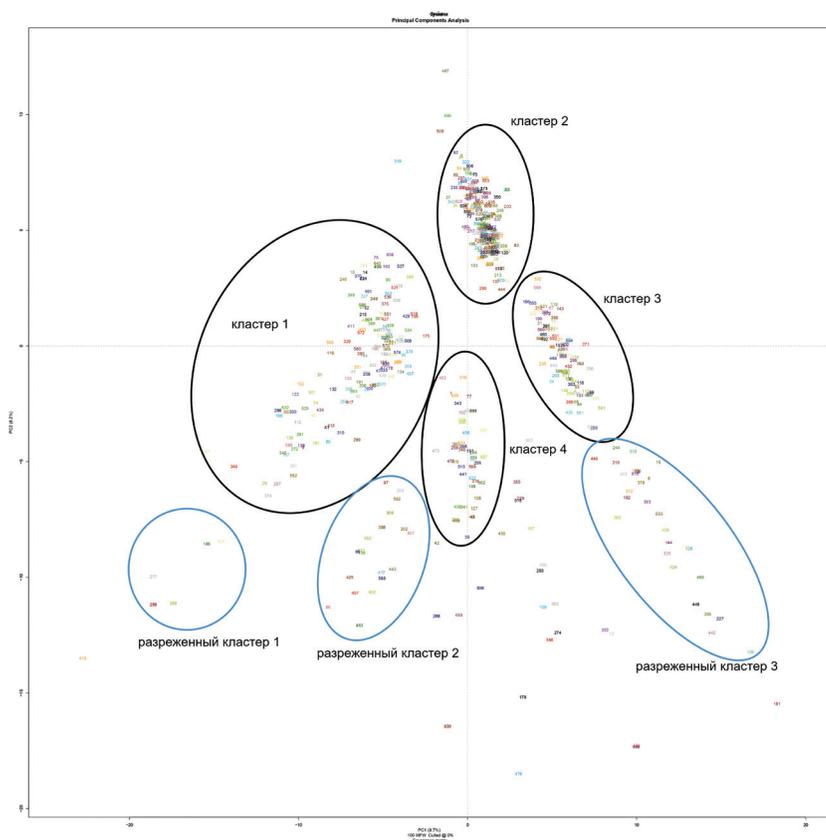


Рис. 2. Визуализация более крупного корпуса частушек методом главных компонент

текстами. Так, самый плотный и объемный кластер 2-22 полностью вошел в состав 2-го кластера нового графика (также самого плотного и крупного), 1-22 вошел в 3-й кластер, 3-22 и 4-22 вошли в 1-й кластер, 4-й кластер включил в себя тексты разреженного кластера 2-22а, 1-й разреженный кластер – тексты разреженного 1-22а.

Для подтверждения неслучайности полученных результатов мы провели анализ методом Bootstrap Consensus Tree (консенсусное дерево, рис. 3), который показывает среднее поведение корпуса для заданного диапазона частот (в нашем случае от 100 до 1000 наиболее частотных слов) и фильтрует локальные помехи [Eder 2017], что подтвердило наблюдения о перераспределении текстов по кластерам.

Для анализа нового корпуса мы применяли уже опробованный алгоритм: вручную рассматривали сначала периферийные тексты, затем случаи наложения, затем различные группы текстов, не вошедших в кластеры, и, наконец, анализировали синтаксические структуры кластеров.

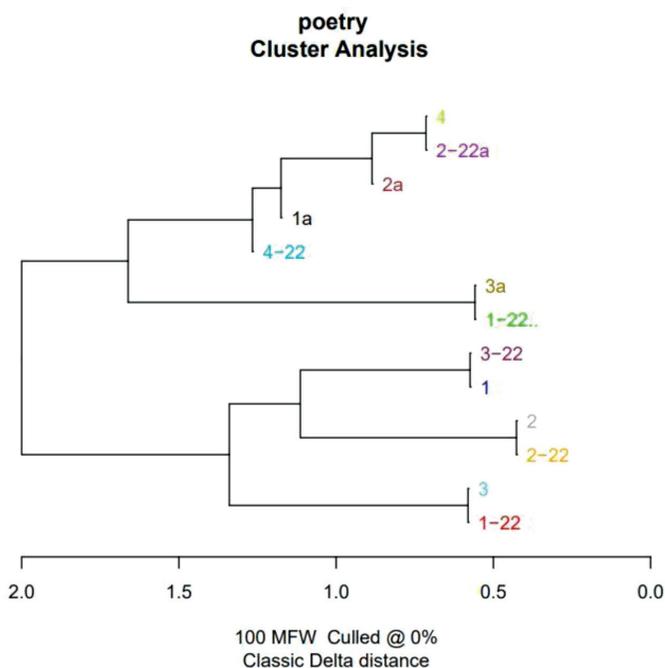


Рис. 3. Дендрограмма кластеров итогового корпуса (1 – 4, разреженные 1а – 4а) и кластеров 2022 г. (1-22 – 4-22)

В верхней половине графика несложно заметить 4 периферийных текста (табл. 7).

Таблица 7

Периферийные тексты в верхней половине нового графика

№ 319 У меня милёнок есть На все четыре букочки: На Т, на О, на Л, на Я, Милёнок Толя у меня.	№ 487 У меня милёнка есть На четыре букочки: На М, на И, на Ш, на А, Милёнка Миша у меня.	№ 490 Сидит ёжик на берёзе – Белая рубашка, На головке сапожок, На ноге фуражка.	№ 509 Меня милый уговаривал На ю, на ю, на ю. А теперь он уговариват Подруженьку мою.
---	---	---	--

Как видим, 2 из них (№ 319 и 487) совпадают практически дословно и отличаются от остальных текстов корпуса уникальной концовкой – перечислением букв имени. Каждая буква (как и каждое слово) считается за отдельный токен, поэтому они рассматриваются как тексты с четырьмя уникальными токенами (25% текста), которые не встречаются в других текстах корпуса. Этим объясняется их периферийное положение. В тексте № 509 также присутствует многократный повтор отдельной буквы, что делает его уникальным на 16%. Таким образом, основное лексическое ядро частушек довольно однородно, и для дистанцирования отдельного текста вверх от основной массы в нашем корпусе достаточно уникальности в 15–43 %. Уникальность токенов в тексте не является единственным необходимым условием для удаления текста от кластеров – наличие уникального слова (вместо уже встречаемого ранее в корпусе) лишь снижает количество связей с другими текстами. Это важно для анализа текстов, не вошедших в кластер на нижней половине графика (табл. 8). Уникальность текста № 410 там составляет 0%, однако ближайшие к нему совпадающие тексты № 250, 278:

Дура я, да дура я,
Дура из картошки,
Дура, я тебе дала,
Протянула ножки.

содержат многократные повторы «дура», не встречаемые в других текстах корпуса. Уникальность остальных периферийных текстов в нижней половине графика составляет 7–46 %.

Таблица 8

Периферийные тексты в нижней половине нового графика

№ 167 Я давно старуха стала, Да старику давать не стала. Повесила на сучок, Не достанешь, старичок.	№ 178 Не пошла я нынче замуж, Не пошла, не каюсь я. С хорошим мужем наживуся, С пьяницей – намаюся.	№ 181 Не топись, не топись, В огороде баня. Не женись, не женись, Мой милёнок Ваня.	№ 410 Дура я, дура я, Дура я проклятая. У него четыре дуры, А я дура пятая.	№ 355 Я любила, не грубила, Он ведь сам любить не стал. На другого посмотрела – Он опять ко мне пристал.
№ 453 Я не буду из- под дубу Ключевую воду пить. Я не буду из- за милого Скандалы заводить.	№ 455 Ты любовь, ты любовь, Я тебя не знала, Сколько раз ты меня Плакать заставляла.	№ 478 Не глядите на меня, Что я худоватая. Мама салом не кормила, Я не виноватая.	№ 502 Я у мамы дочь одна, Ничего не делаю. Только знаю: завлекаю Да измену делаю.	№ 506 Ты почём меня ласкала, Коли я тебе не мил? Ты бы с осени сказала, Я бы зиму не ходил.

В нижней половине графика отчетливо видны случаи наложения (табл. 9). Как и при анализе корпуса из 239 текстов, наложившиеся тексты в более крупном корпусе представляют собой либо идентичные тексты, отличающиеся лишь синтаксическим оформлением (тексты № 246 и 273, 262 и 498), либо тексты, различающиеся одной строкой. Это утверждение справедливо и для случаев наложения, вошедших в основные кластеры и составляющих точки «сгущения».

Устойчивое выделение 4 кластеров в обоих корпусах позволяет сравнить их результаты. Интересно, что с расширением корпуса состав самых частотных слов в плотных кластерах и по корпусу в целом остался практически неизменным, некоторые из них лишь немного изменили свои позиции. Мы видим, что по-прежнему самыми частотными оказываются служебные слова и местоимения (табл. 10). Одни из немногих самостоятельных частей речи, оказавшихся в верхушке частотных списков, –

существительные «гармонист», «соперница», прилагательное «милый» и формы глагола «любить» и «изменить» – встречаются примерно в 5–10 раз реже. Сохраняется и выявленная на корпусе из 239 текстов корреляция между плотностью расположения текстов на графике и количеством употреблений самого частотного слова и следующих за ним: чем разница в количестве употреблений ниже, тем плотнее располагаются тексты.

Таблица 9

Случаи наложения текстов
в нижней половине нового графика

№ 262 Ты играй, ты играй, Ты играй, не бойся, Я тебя не завлеку, Ты не беспокойся.	№ 484 Ося, ося, Ты меня не бойся! Я тебя кусать не стану, Ты не беспокойся.	№ 498 Ты играй, ты играй, Ты играй, не бойся – Я тебя не завлеку, Ты не беспокойся.	№ 246 Милый мой, а я твоя, Прикрой полой, озябла я. Озябла – не согреюся, Люблю и не надеюсь.	№ 273 Милый мой, а я твоя. Прикрой полой – озябла я. Озябла – не согреюсь, Люблю и не надеюсь.
--	---	---	--	---

Особенных изменений с увеличением объема корпуса не произошло и на уровне коллокаций. Так, одно из самых частотных слов «я» наиболее часто встречается в сочетании с глаголом «любила» (18 случаев на 359 употреблений по корпусу), местоимением «его» (10) и служебными частями речи: «в» (11), «не» (11), «на» (15). Наиболее частотные для корпуса из 645 текстов коллокации с «я» не совпадают с коллокациями в НКРЯ: в последнем случае сочетания «я + глагол» встречаются чаще, также немного отличается набор служебных слов, с которыми употребляется «я», самыми распространенными коллокатами выступают «не», «что», «и», «в», «а».

Лидирующие позиции в частотных списках корпуса и нескольких кластеров сохраняет отрицательная частица «не». Ее самыми частотными коллокатами, как и в корпусе из 239 текстов, выступают глаголы. Предлоги «на» и «в» также устойчиво сочетаются с существительными и в предложении выполняют роль обстоятельства места. Союз «а» по-прежнему наибольшее количество раз употребляется с «я», вместе с тем прослеживается явная тенденция объединяться в коллокации и с другими личными местоимениями (*он, ты, мы, мне, меня, другого*) и наречиями (*где, теперь*). В отличие от «а», выделить часть речи, устойчиво сочетающуюся

с союзом «и», сложно: практически в равной степени характерны коллокации с падежными формами существительных, глаголами, местоимениями, предлогами.

Таким образом, синтаксические конструкции, выделенные нами в корпусе из 239 текстов, показали свою устойчивость и в корпусе, превышающем объем первоначального более чем в 2,5 раза.

Таблица 10

Верхушка частотного списка кластеров
корпуса 2021–2023 гг.

	Корпус	Кл. 4	Кл. 3	Кл. 2	Кл. 1	Разр. 1	Разр. 2	Разр. 3
Кол-во текстов	645	52	94	199	147	7	22	31
Кол-во уник. сл.	2938	390	675	1300	915	65	150	238
1	я (359)	я (52)	не (95)	на (73)	я (184)	я (16)	я (45)	не (72)
2	не (330)	не (51)	меня (29)	в (51)	на (52)	дура (9)	не (22)	надо (7)
3	на (221)	а (17)	на (25)	и (47)	а (49)	а (2)	мне (11)	что (7)
4	а (166)	на (16)	а (22)	у (47)	в (43)	да (2)	что (10)	а (6)
5	в (138)	меня (14)	и (22)	а (44)	и (38)	топ-ну (2)	на (7)	вы (6)
6	и (134)	ты (9)	в (20)	меня (33)	за (26)	что (2)	то (7)	запе-вай (6)
7	меня (134)	в (8)	у (18)	с (30)	у (25)		а (5)	мне (6)
8	у (122)	да (8)	ты (15)	за (24)	меня (23)		ве-рится (5)	толь-ко (6)
9	ты (99)	и (8)	да (14)	ты (21)	милый (22)		во (5)	ты (6)
10	милый (73)	что (8)	мы (14)	ми-лый (20)	ты (21)			

Окончание табл. 10

11	с (73)	по (7)	буду (12)	мы (18)	его (20)			
12	за (70)	у (7)	будет (10)	он (18)	он (16)			
13	да (57)	мне (6)	мне (10)	по (18)	с (14)			
14	что (55)	как (5)	где (9)	как (17)	ой (13)			
15	бы (52)	любила (5)	как (9)	мой (17)	моя (12)			
16	он (51)	милый (5)	милый (9)	да (16)	любила (11)			
17	мне (50)	моя (5)	бы (8)	моя (16)	мой (11)			
18	мой (50)	ой (5)	гармонист (8)	вот (14)	свою (11)			
19	по (50)	он (5)	за (8)	из (14)	соперницу (10)			
20	как (46)	тебя (5)	моя (8)	два (12)	изменил (9)			

Таблица 11

Наиболее устойчивые коллокации
с самыми частотными словами корпуса 2021–2023 гг.

	я (359)	не (330)	на (221)	а (166)	в (138)	и (134)
1	я любила (18)	не буду (11)	на горе (10)	а я (48)	в муку (7)	и брови (6)
2	я на (15)	не надо (9)	на мельницу (9)	а теперь (10)	в армию (6)	и не (6)
3	я в (1)	не подумайте (8)	на меня (9)	а он (8)	в лесу (4)	и надо (5)
4	я не (11)	не ругай (7)	на ней (6)	а ты (7)	в городе (3)	и я (5)
5	я его (10)	не верится (6)	на скамеечке (6)	а мы (6)	в коридоре (3)	и в (4)
6	я свою (10)	не стали (6)	на мосту (5)	а мне (5)	в огороде (3)	и полагается (4)

Окончание табл. 11

7	я, бывало (9)	не сумел (6)	на свою (5)	а где (4)	в поле (3)	и на (3)
8	я ему (7)	не подают (5)	на запятки (4)	а другого (4)	в татарах (3)	и стою (3)
9	я тебя (7)	не женись (4)	на низ (4)	а меня (4)	в америку (2)	и вся (2)
10	я по (6)	не знает (4)	на окошке (4)	а на (4)	в говне (2)	и дошла (2)

Выводы

Методы компьютерной лингвистики вообще и программа Stylo в частности кажутся нам продуктивными для анализа фольклорных текстов. Ограничение, связанное с минимальным объемом текстов, обеспечивающим «надежность» результатов, на которое указывают разработчики программы, не кажется нам релевантным при работе с большим количеством текстов примерно равной длины. В статье мы продемонстрировали, что использование Stylo позволяет упростить анализ частушечных корпусов. Построенные графики могут быть использованы для выделения групп текстов на основе их синтаксического строения, что сложно сделать вручную. Дальнейший анализ выделенных кластеров позволяет выделить устойчивые, характерные для частушечной речи лексико-синтаксические модели.

Выделенные с помощью Stylo кластеры позволяют рассмотреть корпус частушечных текстов на нескольких уровнях. На уровне текста выделяются тексты с формами местоимения «я», тексты с обращениями или апелляциями и все остальные. Тексты с «я» составляют большую часть корпуса. Обращения же в частушках используются для наиболее яркого эмоционального воздействия при использовании минимального количества средств. Для этого используют механизмы «нереального адресата», описание стереотипной ситуации, вовлечение слушающего. На уровне предложений и словосочетаний оба рассмотренных нами корпуса делятся на группы, выделенные на основе различных комбинаций частотных слов с другими частями речи. Для исследования локаций при работе с более крупными корпусами представляется обоснованным прибегнуть к методам корпусной лингвистики. Исследование наших корпусов позволило выявить следующие устойчивые закономерности частушечной речи: преобладание сочетания отрицательной частицы «не» с глаголами над сочетаниями с именными формами, преобладание сочетаний «а+я+глагол»

и «а+я+местоимение» над сочетанием «а+я+служебные части речи», преобладание сочетаний «на + существительное» или «на + прилагательное /местоимение + существительное» над сочетанием «на + местоимение», преобладание сочетания «в + существительное» над сочетанием «в + местоимение». Второй компонент сравнения во всех случаях является характерной чертой обычной речи, представленной в современной литературе, публицистике, интернет-коммуникации (примеры взяты из основного корпуса НКРЯ, интересной исследовательской задачей представляется сравнение выделенных нами особенностей частушечной речи с другими корпусами).

Благодарности

Работа выполнена в рамках проекта РГГУ «Песня в русской культуре: поэтика, историческая динамика, социальный контекст» (конкурс «Студенческие проектные научные коллективы РГГУ»).

Авторы выражают благодарность С.Ю. Неклюдову, поддерживавшему и направлявшему наше исследование на всех его этапах, В.А. Воробьеву, обратившему внимание на особенности языка частушки, и К.М. Новикову, писавшему скрипты для облегчения работы с корпусом. Кроме этого, мы сердечно благодарим всех собеседников, уделивших нам время.

Acknowledgements

The work was carried out within the framework of the RSUH project “Song in Russian culture: poetics, historical dynamics, social context” (competition “Student project research teams of RSUH”).

The authors express their gratitude to S.Yu. Neklyudov, who supports and guides our research at all its stages, V.A. Vorobyov, who drew attention to the features of the chastushechny language and K.M. Novikov, who wrote scripts to facilitate work with the corpus.

Литература

- Адоньева 2004 – Адоньева С.Б. Прагматика частушки // Антропологический форум. 2004. № 1. С. 156–178.
- Березкин 2021 – Березкин Ю.Е. Границы и контакты: модели распределения фольклорно-мифологических мотивов в Евразии // Этнография. 2021. № 4 (14). С. 6–23.
- Вежбицкая 1999 – Вежбицкая А. Семантические универсалии и описание языков. М.: Языки русской культуры, 1999. 780 с.
- Жолковский, Мельчук 1965 – Жолковский А.К., Мельчук И.А. О семантическом синтезе // Проблемы кибернетики. Вып. 19 / под ред. А.А. Ляпунова. М.: Наука, 1967. С. 177–238.

- Кербелите 2001 – *Кербелите Б.П.* Типы народных сказаний: структурно-семантическая классификация литовских этиологических, мифологических сказаний и преданий. СПб.: Европейский дом, 2001. 605 с.
- Козьмин 2009 – *Козьмин А.В.* Сюжетный фонд сказок: структура и система. М.: РГГУ, 2009. 109 с.
- Мельчук 1967 – *Мельчук И.А.* К построению действующей модели языка // Проблемы языкознания / отв. ред. К.П. Филин. М.: Наука, 1967. С. 82–89.
- Новикова, Тихонова 2022 – *Новикова В.Б., Тихонова М.А.* Квантитативный анализ частушек с использованием пакета Stylo для R // Песенный фольклор: тексты, традиция, современность: материалы круглого стола (Москва, РГГУ, 10 ноября 2022 г.) / сост.: С.Ю. Неклюдов, В.А. Воробьев, Я.А. Савицкая. М.: РГГУ, 2022. С. 20–27.
- Пермяков 1970 – *Пермяков Г.Л.* От поговорки до сказки (Заметки по общей теории клише). М.: Наука, 1970. 239 с.
- Пропп 1928 – *Пропп В.Я.* Морфология сказки. Л.: Academia, 1928. 152 с.
- Сёрль 1986 – *Сёрль Дж.Р.* Что такое речевой акт? // Новое в зарубежной лингвистике / сост. и вступ. ст. И.М. Кобозевой, В.З. Демьянкова. Вып. 17. М.: Прогресс, 1986. С. 151–169.
- Eder and al. 2016 – *Eder M., Rybicki J., Kestemont M.* Stylometry with R: A package for computational textanalysis // *The R Journal*. 2016. No. 1 (8). P. 107–121.
- Eder 2017 – *Eder M.* Visualization in stylometry: Cluster analysis using networks // *Digital scholarship in the humanities*. 2017. Vol. 32. No. 1. P. 50–64.

References

- Adonyeva, S.B. (2004), “The pragmatics of chastushka”, *Antropologicheskii forum*, vol. 1, pp. 156–178.
- Berezkin, Yu.E. (2021), “Boundaries and contacts: models of distribution of folklore-mythological motifs in Eurasia”, *Ethnografia*, vol. 14, no. 4, pp. 6–23.
- Eder, M. (2017), “Visualization in stylometry: Cluster analysis using networks”, *Digital scholarship in the humanities*, vol. 32, no. 1, pp. 50–64.
- Eder, M., Rybicki, J. and Kestemont, M. (2016), “Stylometry with R: A package for computational textanalysis”, *The R Journal*, vol. 8, no. 1, pp. 107–121.
- Kerbelite, B.P. (2001), *Tipy narodnykh skazanii: strukturno-semanticheskaya klassifikatsiya litovskikh etnologicheskikh, mifologicheskikh skazanii i predanii* [Types of folk tales: a structural-semantic classification of Lithuanian etiological, mythological tales and legends], *Evropeiskii dom*, Saint Petersburg, Russia.

- Koz'min, A.V. (2009), *Syuzhetnyi fond skazok: struktura i sistema* [Plot fund of fairy tales: structure and system], RGGU, Moscow, Russia.
- Mel'chuk, I.A. (1967), "On the construction of an operative model of language", in Filin, K.P., ed., *Problemy yazykoznaniya* [Problems of linguistics], Nauka, Moscow, Russia, pp. 82–89.
- Novikova, V.B. and Tikhonova, M.A. (2022), "Quantitative analysis of chastushkas using the Stylo package for R", in Neklyudov, S.Yu., Vorob'ev, V.A. and Savitskaya, Ya.A., comps., *Pesennyi fol'klor: teksty, traditsiya, sovremenost': materialy kruglogo stola (Moskva, RGGU, 10 noyabrya 2022 g.)* [Song folklore: Texts, tradition, modernity: proceedings of the conference (Moscow, RSUH, November 10, 2022) RGGU, Moscow, Russia, pp. 20–27.
- Permyakov, G.L. (1970), *Ot pogovorki do skazki (Zametki po obshchei teorii klishe)* [From proverb to fairy tale (notes on the general theory of cliché)], Nauka, Moscow, USSR.
- Propp, V.Ya. (1928), *Morfologiya skazki* [The morphology of the fairy tale], Academia, Leningrad, USSR.
- Searle, J.R. (1986), "What is a speech act?", in Kobozeva, I.M. and Dem'yankov, V.Z., comps., *Novoe v zarubezhnoi lingvistike* [New in foreign linguistics], vol. 17, Progress, Moscow, USSR, pp. 151–169.
- Vezhbickaya, A. (1999), *Semanticheskie universalii i opisaniye yazykov* [Semantic universals and language description], Yazyki russkoi kultury, Moscow, Russia.
- Zholkovskii, A.K. and Mel'chuk, I.A. (1965), "On semantic synthesis", in Lyapunov, A.A., ed., *Problemy kibernetiki* [Problems of cybernetics], vol. 19, Nauka, Moscow, USSR, pp. 177–238.

Информация об авторах

Виктория Б. Новикова, Российский государственный гуманитарный университет, Москва, Россия; 125047, Россия, Москва, Миусская пл., д. 6; jranchik@gmail.com

Маргарита А. Тихонова, Российский государственный гуманитарный университет, Москва, Россия; 125047, Россия, Москва, Миусская пл., д. 6; margarita.a.tikhonova@mail.ru

Information about the authors

Victoria B. Novikova, Russian State University for the Humanities, Moscow, Russia; 6, Miusskaya Sq., Moscow, Russia, 125047; jranchik@gmail.com

Margarita A. Tikhonova, Russian State University for the Humanities, Moscow, Russia; 6, Miusskaya Sq., Moscow, Russia, 125047; margarita.a.tikhonova@mail.ru