

Чеповецкая С.В.^{1), 2)}, Мелкозёрова В.И.¹⁾, Шерстинова Т.Ю.¹⁾

«ФУ, КАК НЕКУЛЬТУРНО!»: ФАКТОРЫ, ВЛИЯЮЩИЕ
НА УПОТРЕБЛЕНИЕ ОБСЦЕННОЙ ЛЕКСИКИ В РУССКОМ ЯЗЫКЕ
(НА МАТЕРИАЛЕ КОРПУСА УСТНОЙ РЕЧИ МОЛОДЕЖИ)^{©, 1)}

¹⁾ Национальный исследовательский университет
«Высшая школа экономики»,

²⁾ Санкт-Петербургский государственный университет,
Россия, Санкт-Петербург, avchepovetskaya@edu.hse.ru,
vimelkozrova@edu.hse.ru, tsherstinova@hse.ru

Аннотация. В статье рассматривается влияние экстравергистических факторов, таких как социально-психологические характеристики информантов и особенности коммуникативных ситуаций, на частоту использования ненормативной лексики в устной речи. Материалом исследования стал корпус расшифровок устной речи объемом в 286 662 словоупотреблений. Исследование показало, что количество ненормативной лексики и балл, полученный информантом по шкале «эмоциональная нестабильность», характеризуются умеренной обратной взаимосвязью, в то время как частота употребления бранной лексики уменьшается с увеличением социальной дистанции между участниками речевого акта.

Ключевые слова: устный дискурс; ненормативная лексика; корреляционный анализ; русский мат; речевая ситуация.

Поступила: 16.09.2024

Принята к печати: 10.01.2025

© Чеповецкая С.В., Мелкозёрова В.И., Шерстинова Т.Ю., 2025

¹⁾ Публикация подготовлена в результате проведения исследования по проекту «Текст как Big Data: методы и модели работы с большими текстовыми данными» в рамках Программы фундаментальных исследований НИУ ВШЭ в 2024 г.

Chepovetskaya S.V.^{1), 2)}, Melkozerova V.I.¹⁾, Sherstinova T.Y.¹⁾

**Factors influencing the use of obscene vocabulary in the Russian language
(on the material of the youth speech corpus)^{©, 1)}**

¹⁾ National Research University Higher School of Economics,

²⁾ Saint-Petersburg State University, Russia, Saint-Petersburg,

avchepovetskaya@edu.hse.ru, vimelkozerova@edu.hse.ru, tsherstinova@hse.ru

Abstract. The paper considers the influence of extra-linguistic factors, such as socio-psychological characteristics of informants and the features of communicative situations, on the frequency of their use of obscene language in oral communication. The empirical material was accumulated from a corpus of oral speech transcripts, containing 286 662 words. The study showed that the scale of obscenities used and the score obtained by an informant on the scale “emotional instability” are characterized by moderate inverse relationship, while the frequency of the obscene vocabulary used depends on the type of communicative situation and decreases when the social distance between the participants of the speech act increases.

Keywords: oral discourse; obscene vocabulary; correlation analysis; Russian profanity; speech situation.

Received: 16.09.2024

Accepted: 10.01.2025

Введение

В настоящей статье анализируется частотность употребления ненормативной лексики информантами старше 18 лет в различных коммуникативных ситуациях, классифицированных по признаку «социальная дистанция между участниками речевого акта». Исследование направлено на выявление социально-психологических характеристик, влияющих на наблюдаемое количество нецензурных слов в речи информантов, записи устной речи которых послужили материалом для настоящей работы.

В ряде исследований, посвященных описанию функций обсценной лексики и контекстов ее употребления, предпринимаются попытки составить речевой и социокультурный портрет человека, речь которого характеризуется высокой долей ненормативной лексики. Так, основываясь на историко-политическом контексте, М. Ковалев и

© Chepovetskaya S.V., Melkozerova V.I., Sherstinova T.Y., 2025

¹⁾ The paper was prepared within the framework of the Basic Research Program at HSE University in 2024 (project “Text as big data: methods and models for working with large textual data”).

соавторы рассматривают обсценную лексику как маркер антисоциального поведения, поскольку ее употребление квалифицировалось советским законодательством как мелкое хулиганство [Kovalev, Platonov, Hutchings, 2014, p. 45] и расценивалось как проявление вербальной агрессии. Помимо этого авторы монографии формируют социальный портрет индивида, склонного к использованию ненормативной лексики, с точки зрения представителей власти: «...типичным хулиганом был молодой человек, как правило, бедный, основными характеристиками которого являлись использование нецензурной лексики, употребление алкоголя и вульгарное поведение, в частности, враждебное отношение к традиционным ценностям» [ibid., p. 44] (перевод наш – С.Ч., В.М., Т.Ю.).

Связь употребления обсценной лексики с гендерной идентичностью личности подтверждается результатами социологического исследования, описанными в работе Т.С. Киенко. Согласно опросу жителей провинциального города, 9,9% опрошенных, среди которых 65% мужчин, считают допустимым регулярное употребление мата. В числе 33,8% респондентов, не рассматривающих использование нецензурной лексики, 39,8% мужчин [Киенко, 2014, с. 127]. В статье М. Слюсаренко, направленной на выявление причин возникновения речевых актов, в которых используется ненормативная лексика, наряду с проявлением агрессии и конфронтации упоминаются потеря самоконтроля в стрессовых ситуациях, когда человек прибегает к не свойственным ему стратегиям речевого поведения [Slusarenko, 2023, p. 39], а также низкий уровень словарного запаса и отсутствие осведомленности о культуре речи [ibid., p. 40]. Сопоставимые данные приводит В. Собкин со ссылкой на работу С.М. Кеннисон и Р.Х. Мессер, в которой утверждается, что количество ненормативной лексики «положительно коррелирует с экстраверсией и враждебностью» [Kennison, Messer, 2017, p. 123; цит. по: Собкин, 2021, с. 9–10].

Среди исследований, направленных на получение количественных данных об употреблении обсценной лексики в зависимости от характеристик коммуникативной ситуации, таких как гендер информанта и коммуникантов, статус их отношений и интенция высказывания, выделим работу К.Б. Фэгерстен [Fägersten, 2012]. Анализ, проведенный методом включенного наблюдения за спонтанной устной речью, позволил установить разницу между типами коммуникативной ситуации с точки зрения приемлемости употребления ненормативной лексики. Рассмотрение подвыборки,

состоящей из 77 женщин, продемонстрировало, что в 58% случаев обсценные выражения использовались информантами, если они состояли в дружеских отношениях с коммуникантами. Для эпизодов формальной коммуникации (преподаватель / студент, начальник / подчиненный) частота случаев употребления не превышала 3% [Fägersten, 2012, p. 65].

Цель работы – изучение влияния приведенных факторов на употребление обсценной лексики с опорой на статистические данные, полученные при исследовании аннотированного корпуса текстовых расшифровок современной устной речи, описанию которого посвящен второй раздел статьи. В проведенном корреляционном анализе независимыми переменными послужили результаты пятифакторного опросника личности и теста на словарный запас. Был исследован также фактор типа коммуникативных ситуаций, нормированных в соответствии с отношениями между информантом и коммуникантом.

Материал и методика исследования

В исследовании использовались материалы «Корпуса устной речи молодежи» – нового ресурса, который наследует от корпуса ОРД¹ принцип непрерывной записи в естественных условиях, но концентрируется на молодежной, преимущественно студенческой аудитории [Шерстинова, Петрова, 2023]. Идея создания такого корпуса опирается на тезис, согласно которому речь молодых людей наименее консервативна, наиболее открыта экспериментам и больше подвержена изменениям.

В настоящий момент работа над «Корпусом устной речи молодежи» осуществляется в Лаборатории языковой конвергенции при НИУ ВШЭ (СПб). Пока в корпусе представлены преимущественно записи информантов женского пола, поэтому корпус сложно назвать сбалансированным по гендерному признаку. Средний возраст информанта составляет 21 год, но есть единичные «выбросы» [Шерстинова, Петрова, 2023].

Как было отмечено ранее, запись ведется непрерывным способом: информанты носят диктофон с собой в течение дня [The ORD speech corpus of Russian everyday communication “One speaker’s day”: Creation principles and annotation, 2009]. Они проходят психологические тесты и заполняют ряд анкет, в которых указывают информацию о себе, информацию о тех, с кем они общались

¹ Корпус «Один речевой день»: <https://ord.spbu.ru/>

во время записи. Полученные аудиофайлы передаются студентам-практикантам для обработки, которая состоит из трех основных этапов: нарезки эпизодов, автоматизированного распознавания звучащей речи, а также экспертного редактирования расшифровок. Каждый эпизод сопровождается аннотацией о говорящих, действиях и месте, что позволяет исследовать влияние фактора коммуникативных ситуаций на те или иные особенности речи.

Для выявления связи между частотностью ненормативной лексики и исследуемыми экстралингвистическими факторами были рассмотрены экспертные расшифровки от 19 информантов (286 622 словоупотреблений, около 65 часов записи). Тексты подверглись автоматической предобработке: очищены от знаков препинания, токенизированы с помощью нейросетевой модели токенизации razdel¹ и лемматизированы инструментами пакета ruystem³.

В качестве базового списка нецензурных слов использовался готовый список слов². Однако при автоматической лемматизации текстов некоторые обсценные словоформы были приведены к леммам некорректно (например, вопросительное слово *х**и* модель, по-видимому, интерпретировала как императив и заменила на инфинитив *хулить*). Словарь был дополнен автоматически лемматизированными обсценными формами в тех случаях, когда они не совпадали с базовыми словарными леммами.

Работа велась по двум направлениям: были сформированы выборки по информантам и по коммуникативным ситуациям. Основные этапы исследования: автоматическое извлечение контекстов с обсценными словами, составление частотных словарей, статистический подсчет и проведение корреляционного анализа.

Использование ненормативной лексики: характеристики информанта

Для оценки частоты использования ненормативной лексики информантами из корпуса были извлечены метаданные каждого эпизода, в частности информация о роли говорящих, что позволило составить 19 подкорпусов в соответствии с количеством информантов. Каждый из подкорпусов содержит от 361 до 16 455 слово-

¹ Razdel библиотека // Проект Natasha URL: <https://natasha.github.io/razdel>

² Словарь маты русского языка // Github URL: <https://github.com/nickname76/russian-swears>

употреблений. На следующем этапе были осуществлены подсчет количества нецензурных слов в подкорпусах и их перевод в относительные частоты (в ipm), благодаря чему стало возможным сопоставление полученных данных. После завершения участия в исследовании каждому информанту предлагалось пройти тест на словарный запас¹, разработанный Г. Головиным с опорой на частотный словарь Т.В. Ефремовой [Головин, 2015, с. 149], в рамках которого участники должны были отметить, знакомо ли им представленное слово. В тесте предусмотрены контрольные вопросы, предполагающие выбор определения, подходящего слову, если участник дает положительный ответ. В измерении объема пассивного словарного запаса принял участие 14 из 19 информантов, что привело к вынужденному сокращению подвыборки при проведении корреляционного анализа.

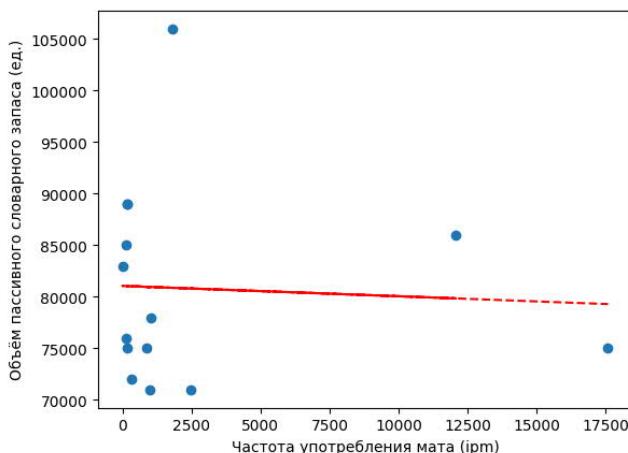


Рис. 1. Зависимость частоты употребления обсценной лексики от объема словарного запаса

Средний объем словарного запаса информанта составил 80 786 слов, медиана – 77 000. Средняя частота употребления обсценной лексики участниками – 2492 ipm. При оценке выбросов было установлено верхнее пороговое значение, равное 17 277 ipm, что, в свою очередь, равно сумме среднего значения и трех стан-

¹ Тест словарного запаса (русский язык): <https://www.myvocab.info/ru>

дартных отклонений. Установлено, что данные содержат одно значение, превышающее критическое (17 578 *ipm*) и полученное в результате подсчета частоты ненормативной лексики в речи единственного информанта мужского пола в исследуемой выборке.

Для оценки взаимосвязи параметров применялась формула линейной корреляции Пирсона. Коэффициент корреляции, рассчитанный для переменных *частота ненормативной лексики* и *объем словарного запаса* составил $-0,05$ (рис. 1), что, согласно шкале Чеддока, не является значимым уровнем силы взаимосвязи. Кроме того, об отсутствии связи между словарным запасом и количеством ненормативной лексики в речи свидетельствует результат статистического теста по критерию хи-квадрат (χ^2 , *p-value* $\approx 0,85$).

С целью проверки гипотезы о влиянии индивидуальных психологических характеристик на частоту употребления обсценной лексики были использованы результаты пятифакторного опросника личности (5PFQ), предложенного информантам. Тест разработан Н. Tsuji (1996), адаптирован А.Б. Хромовым (2000)¹ и включает в себя 75 вопросов, направленных на определение балла участника в рамках пяти шкал, по каждой из которых максимальный возможный балл составляет 75: «интроверсия – экстраверсия», «обособленность – привязанность», «импульсивность – самоконтроль», «эмоциональная устойчивость – эмоциональная неустойчивость», «практичность – экспрессивность». Шкалы представляют собой континуум, полюса которого соответствуют противоположным личностным характеристикам. Таким образом, чем больше баллов набрано участником по шкале, тем ближе к правому полюсу континуума он находится. Это означает, что участник, набравший высокий балл по шкале, будет демонстрировать более выраженные черты, связанные с ее правым полюсом. Респондент, набравший более низкий балл, с большей вероятностью будет демонстрировать черты, связанные с левым полюсом континуума. Предполагается, что полученные результаты могут быть связаны с характеристиками речи информанта, в том числе с частотностью употребления им обсценной лексики.

В настоящем исследовании устанавливалась связь частоты использования нецензурной лексики с баллами, полученными 18 информантами, прошедшими тест, по всем указанным шкалам. Для этого проводилось попарное сравнение переменных в соответствии

¹ Пятифакторный опросник личности, 5PFQ: <https://psytests.org/big5/5pfq.html>

с количеством шкал. Результат корреляционного анализа продемонстрировал умеренную обратную связь между баллом, набранным по шкале «эмоциональная стабильность – эмоциональная нестабильность», и частотой употребления ненормативной лексики ($r \approx -0,4$, рис. 2), в то время как коэффициент корреляции с результатами, полученными по другим шкалам, варьировался в пределах от $-0,22$ до $-0,28$, что свидетельствует о слабой взаимосвязи признаков, однако статистическая значимость полученных данных не дает права подтвердить или опровергнуть альтернативную гипотезу ($p\text{-value} \approx 0,1$).

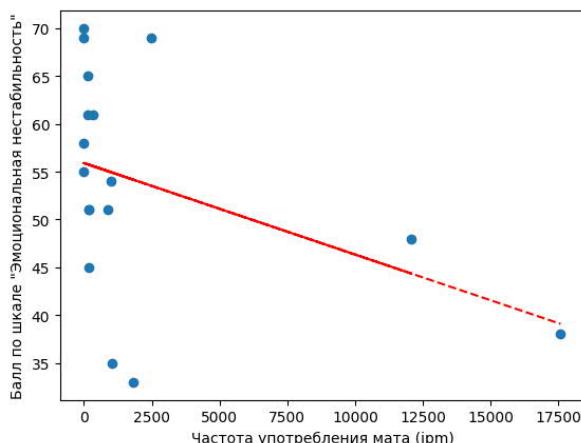


Рис. 2. Зависимость частоты употребления ненормативной лексики от балла, набранного по шкале «эмоциональная стабильность – эмоциональная нестабильность»

Употребление обсценной лексики: тип речевой ситуации

В зависимости от степени нормированности отношений между информантом и коммуникантами исследуемые эпизоды были поделены на три группы.

1. «Вертикальная коммуникация» включает эпизоды, состоящие из общения между клиентами и сотрудниками сервисов, а также взаимодействия в академической среде между студентами и преподавателями. Коммуникативная ситуация характеризуется психо-

логической дистанцией между собеседниками, формализованным стилем речи и запретом использования нецензурной лексики.

2. Коммуникация с членами семьи и домашними животными. Общение между старшими и младшими родственниками также имеет ограничения на использование ненормативной лексики, однако эти правила менее строгие, чем в группе № 1, и могут нарушаться.

3. «Горизонтальная коммуникация»: общение с друзьями, однокурсниками. К отличительным чертам относится иерархическое равенство собеседников и неформальная обстановка.

Поскольку в этой части исследования фокус внимания направлен на влияние фактора коммуникативной ситуации, личностные особенности не представлялись важными, в связи с чем были рассмотрены не только реплики информантов, но и все остальные реплики, принадлежащие коммуникантам. Описание получившихся выборок и результат поиска обсценной лексики представлены в таблице. Хотя выборки получились не равными по количеству токенов, это не помешало анализу, потому что учитывалась не абсолютная, а относительная частотность (ipm).

Частотность обсценных слов в трех выборках по речевым ситуациям

№	Количество эпизодов	Длительность эпизодов, часы	Количество токенов	Количество обсценных слов, токены	Относительная частотность, ipm
1	46	15,4	55704	4	0,7
2	64	21,6	78930	15	1,9
3	82	28,4	140048	579	41,3

Полученные результаты свидетельствуют о том, что частотность ненормативной лексики в речи возрастает с понижением степени формальности коммуникативной ситуации, особенно в коммуникации с равными собеседниками. Несмотря на то, что рассчитать коэффициент корреляции между порядковой и числовой переменными не представляется возможным, следует обозначить тенденцию: наблюдается существенная разница в частотности обсценных слов между текстами первой, второй и третьей групп (0,7 ipm, 1,9 ipm и 41,3 ipm, соответственно). Как следствие, в третьей группе обнаруживается значительная вариативность нецензурных форм (на 579 вхождений найдено 44 типа). При этом наиболее частотные вхождения совпадают во всех трех группах (*б**ть, б*я, х**ня*).

Заключение

Проведенный анализ продемонстрировал результаты, частично подтверждающие данные, представленные в других работах (см., например: [Комалова, Голощапова, 2021]). Однако взаимосвязь объема словарного запаса и частоты обсценной лексики, установленная в настоящей работе, позволяет опровергнуть гипотезу об обратной взаимосвязи признаков, представленную в более ранних исследованиях (см., например: [Бакунев 2018; Клейберг 2022]). Неопределенный уровень значимости, наблюдаемый для частоты употребления нецензурной лексики и результатов 5PFQ, может быть объяснен недостаточным объемом материала, в связи с чем предполагается, что увеличение количества записей в корпусе позволит прийти к более однозначным выводам.

Что касается фактора речевых ситуаций, полученные результаты соотносятся с предшествующими исследованиями: наиболее высокая частотность обсценных слов зафиксирована в эпизодах «горизонтальной» коммуникации, преимущественно между друзьями.

Таким образом, исследование позволило доказать достоверность тезисов о влиянии экстралингвистических характеристик на частоту использования ненормативной лексики и подтвердить их количественными данными на материале современной устной речи. При проведении дальнейших исследований и расширении выборки особое внимание следует уделить гендерной репрезентативности. Это позволит учитывать фактор гендера информанта и его коммуникантов в ряду независимых переменных при проведении корреляционного анализа. В перспективе планируется обучение модели линейной регрессии, задача которой будет заключаться в предсказании количества обсценной лексики в речи информанта в зависимости как от его социально-психологических особенностей, так и от типа коммуникативной ситуации.

Список литературы

- Бакунев В.В. Об использовании обсценной лексики в онлайн-среде // Мировая журналистика: единство многообразия. – 2018. – С. 90–95.
- Головин Г.В. Измерение пассивного словарного запаса русского языка // Социо- и психолингвистические исследования. – 2015. – № 3. – С. 148–159.
- Киенко Т.С. Ненормативная лексика как социальный феномен (анализ опроса жителей провинциального российского города) // Социологические исследования. – 2016. – № 5. – С. 124–133.

- Клейберг Ю.А. Обсценная лексика в подростково-молодежной среде // Язык и мышление: психологические и лингвистические аспекты. – 2022. – С. 151–164.
- Комалова Л.Р., Голоцапова Т.И. Дифференцированный анализ речевого жанра «оскорбление» на материале сообщений социальной интернет-сети // Вестник Российской университета дружбы народов. Серия: Теория языка. Семиотика. Семантика = RUDN Journal of Language Studies, Semiotics and Semantics. – 2021. – № 12(3). – С. 619–631. DOI: 10.22363/2313-2299-2021-12-3-619-631
- Собкин В.С. Особенности использования нецензурной речи российскими подростками // Вопросы психологии. – 2021. – № 2. – С. 28–42.
- Шерстинова Т.Ю., Петрова И.А. Моделирование повседневного речевого поведения: корпус устной речи молодежи, или ОРД V. 2.0. // Социо- и психолингвистические исследования. – 2023. – Вып. 11. – С. 7–13.
- Fägersten K.B. Who's swearing now?: The social aspects of conversational swearing. – Engelska Newcastle : Södertörns högskola, 2012. – 328 p.
- Kennison S.M., Messer R.H. Cursing as a form of risk-taking // Current Psychology. – 2017. – Vol. 36., N 1. – P. 119–126.
- Kovalev M., Platonov R., Hutchings S. The function of Russian obscene language in late Soviet and Post-Soviet prose. – Manchester : University of Manchester, 2014. – 226 p.
- Sliusarenko M. Obscene response as an individually responsible act // Scientific studios on social and political psychology. – 2023. – Vol. 29, N 1. – P. 38–42.

References

- Bakunov, V.V. (2018). Ob ispol'zovanii obstsennoj leksiki v onlajn-srede. *Mirovaya zhurnalistika: edinstvo mnogoobraziya*, 90–95.
- Golovin, G.V. (2015). Izmerenie passivnogo slovarnogo zapasa russkogo jazyka. *Sotsio- i psiholingvisticheskie issledovaniya*, 3, 148–159.
- Kienko, T.S. (2016). Nenormativnaja leksika kak sotsial'nyj fenomen (analiz oprosa zhitelej provintsial'nogo rossiskogo goroda). *Sotsiologicheskie issledovaniya*, 5, 124–133.
- Klejberg, Ju.A. (2022). Obstsennaya leksika v podrostkovо-molodyozhnoj srede. *Yazyk i myshlenie: psihologicheskie i lingvisticheskie aspekty*, 151–164.
- Komalova, L.R., Goloschapova, T.I. (2021). Algorithmization of potentially insulting speech acts analysis: messages posted on social network site. *RUDN Journal of Language Studies, Semiotics and Semantics*, 12(3), 619–631. DOI: 10.22363/2313-2299-2021-12-3-619-631
- Sobkin, V.S. (2021). Osobennosti ispolzovaniija netsenzurnoj rechi rossiskimi podrostkami. *Voprosy psichologii*, 2, 28–42.
- Sherstinova, T.Ju., Petrova, I.A. (2023). Modelirovanie povsednevного rechevogo povedenija: korpus ustnoj rechi molodezhi, ili ORD V. 2.0. *Sotsio- i psiholingvisticheskie issledovaniya*, 11, 7–13.
- Fägersten, K.B. (2012). *Who's swearing now?: The social aspects of conversational swearing*. Södertörns högskola: Engelska Newcastle.
- Kennison, S.M., Messer, R.H. (2017). Cursing as a form of risk-taking. *Current Psychology*, 36(1), 119–126.

- Kovalev, M., Platonov, R., Hutchings, S. (2014). The Function of Russian obscene language in late Soviet and Post-Soviet prose. Manchester: University of Manchester.
- Sliusarenko, M. (2023). Obscene response as an individually responsible act. *Scientific studios on social and political psychology*, 29(1), 38–42.
-

Сведения об авторах

Чеповецкая София Вадимовна – стажер-исследователь Лаборатории языковой конвергенции, Национальный исследовательский университет «Высшая школа экономики» – Санкт-Петербург, лаборант-исследователь Научно-исследовательского института проблем государственного языка, Санкт-Петербургский государственный университет, Россия, Санкт-Петербург, avchepovetskaya@edu.hse.ru

Мелкозёрова Валерия Игоревна – стажер-исследователь Лаборатории языковой конвергенции, Национальный исследовательский университет «Высшая школа экономики», Россия, Санкт-Петербург, vimelkozrova@edu.hse.ru

Шерстинова Татьяна Юрьевна – кандидат филологических наук, доцент департамента филологии и академический руководитель ОП «Филология», Национальный исследовательский университет «Высшая школа экономики» – Санкт-Петербург, Россия, Санкт-Петербург, tsherstinova@hse.ru

About the authors

Chepovetskaya Sofia Vadimovna – Research Assistant at Linguistic Convergence Laboratory, National Research University “Higher School of Economics” – Saint-Petersburg, Laboratory Researcher at Research Institute for Official Language Problems, Saint-Petersburg State University, Russia, Saint-Petersburg, avchepovetskaya@edu.hse.ru

Melkozrova Valeriia Igorevna – Research Assistant at Linguistic Convergence Laboratory, National Research University “Higher School of Economics” – Saint-Petersburg, Russia, Saint-Petersburg, vimelkozrova@edu.hse.ru

Sherstinova Tatiana Yurievna – PhD in Philology, Associate Professor of the Department of Philology and Academic Director of the Philology educational program, National Research University “Higher School of Economics” – Saint-Petersburg, Russia, Saint-Petersburg, tsherstinova@hse.ru