

**РЕЧЕВОЙ КОРПУС ВЕПКАР КАК ИНСТРУМЕНТ
СОХРАНЕНИЯ ДИАЛЕКТНОЙ РЕЧИ
ПРИБАЛТИЙСКО-ФИНСКИХ НАРОДОВ КАРЕЛИИ¹**

Статья посвящена описанию Речевого корпуса прибалтийско-финской речи, созданного на платформе «Открытого корпуса вепского и карельского языков» (ВепКар), его архитектуре и возможностям. Разработанный сотрудниками Института языка литературы и истории (ИЯЛИ) и Института прикладных математических исследований (ИМПИ) Речевой корпус представляет собой собрание звучащих текстов на разных диалектах карельского и вепского языков, снабженных транскрипцией, разметкой и переводом на русский язык, а также необходимые для работы поисковые фильтры (поиск по языку/диалекту, месту и году записи, информанту и собирателю, источнику). Актуальность исследования обусловлена необходимостью дальнейшего развития корпуса ВепКар, широко востребованного как в научных исследованиях, так и в процессе развития литературных форм карельского и вепского языков. Применение современных технологий и методик к накопленному на протяжении многих десятилетий полевому материалу в совокупности с новейшими данными позволит восполнить целый ряд лакун, выявленных лингвистами в данной системе ранее. Для наполнения корпуса аудиозаписями карельской и вепской речи исследователи используют три основных источника: аудиокolleкции Фонограммархива ИЯЛИ КарНЦ РАН, аудиозаписи передач на ливвиковском наречии карельского языка, а также полевые материалы авторов, записанные в ходе экспедиций. Научная новизна обоснована недостатком речевых корпусов прибалтийско-финских языков. Цифровизация архивных и полевых аудио-образцов карельской и вепской речи в формате Речевого корпуса в дальнейшем сможет упростить обработку и хранение материалов, позволит ввести в научный оборот и представить в открытый доступ уникальные аудиоматериалы, отражающие состояние карельских и вепских диалектов начиная с середины прошлого столетия.

Ключевые слова: Речевой корпус, вепский язык, карельский язык, корпусная лингвистика, аудио-образцы, разметка текста.

DOI: 10.35634/2224-9443-2023-17-3-343-351

Введение

Начавшееся десятилетие (2022–2032 гг.) названо в мире Десятилетием языков коренных народов, которое, в первую очередь, сосредоточено на правах носителей языков коренных народов. Для сохранения языкового богатства и последующего изучения языков коренных народов создаются лингвистические корпуса.

В мире существует большое количество лингвистических корпусов. К наиболее известным в мировом масштабе относятся: Национальный корпус русского языка [НКРЯ], Британский национальный корпус [БНК], Чешский национальный корпус [ЧНК]. Известны три наиболее крупных корпуса финно-угорских языков: Языковой банк Финляндии [ЯБФ], Сводный корпус эстонского языка [СКЭЯ] и Венгерский национальный корпус [ВНК]. Среди корпусов финно-угорских республик России можно выделить Национальный корпус удмуртского языка [НКУЯ], Корпус лугового марийского языка [КЛМЯ], Корпус коми-зырянского [ККЗЯ] и Коми-пермяцкого языков [ККПЯ] и т. д. В условиях цифровизации научного знания языковеды ИЯЛИ КарНЦ РАН совместно с исследователями ИПМИ в 2016 году занялись разработкой нового направления: созданием интернет-ресурса «Открытый корпус вепского и карельского языков» [ВепКар]. Корпус ВепКар является многофункциональным, т. к. содержит большое количество инструментов, позволяющих языковедам успешно использовать этот ресурс в своих исследованиях. В настоящее время размещено более 4,1 тыс. текстов на 46 диалектах карельского и вепского языков, словари и компьютерные программы для обработки, поиска и представления данных. Основу корпуса составляют письменные тексты различных жанров и

¹ Исследование выполнено при финансовой поддержке Российского научного фонда совместно с органами власти Республики Карелия с финансированием из Фонда венчурных инвестиций Республики Карелия (ФВИ РК) проект № 22-28-20215 «Создание речевого корпуса прибалтийско-финских языков Карелии».

типов, созданных начиная с XIX столетия [Бойко 2021, 103, Krizhanovskaya 2022, 48]. В корпусе в настоящее время также организована удобная система поиска, которая помогает отфильтровать тексты не только по языковой или стилистической, но и по диалектной принадлежности, или, например, по информанту, собирателю или автору, году записи или году публикации (рис. 1). Поиск лемм возможен по диалектам, частям речи, грамматическим признакам и даже по лексико-семантическим категориям [Крижановский 2019, 289; Воуко 2022, 37;]. Таким образом, ВепКар стал основной базой для исследования прибалтийско-финских языков Северо-Запада России.

Однако одних лишь текстовых данных недостаточно для проведения качественных фонетических исследований с применением современных программ обработки и анализа речи. В связи с этим в 2022 году исследователи ИЯЛИ и ИПМИ приступили к работе над созданием Речевого корпуса прибалтийско-финских языков Карелии.

В процессе исследования использовались теоретические **методы** исследования (ознакомление с материалом, оценка основательности предположения), методы корпусной лингвистики, лексикографии, лингвистической географии. Собранный материал анализируется с использованием преимущественно сравнительно-сопоставительного и сравнительно-исторического методов. Отсутствие хронологических рамок при отборе материала позволяет проводить диахронические исследования языков корпуса.

При проведении исследования использовались теоретические источники [Бойко 2021; Зайцева 2012; Крижановский 2019; Кузнецова 2019; Новак 2021; Родионова 2022; Воуко 2022; Krizhanovskaya 2022]. Для наполнения корпуса аудиозаписями карельской и вепсской речи исследователи привлекают материалы Фонограммархива ИЯЛИ КарНЦ РАН [ФА], в том числе опубликованные и изданные в образцах карельской речи [Баранцев 1978; Макаров 1969; Näytteitä karjalan kielestä 1994], аудиозаписи передач на ливвиковском наречии карельского языка, а также полевые материалы авторов, записанные в ходе экспедиций.

Речевой корпус прибалтийско-финских языков Карелии

Ситуация, в которой пребывают прибалтийско-финские языки Карелии, можно назвать тревожной. Численность носителей карельского языка, являющегося языком титульной нации республики, и вепсского – языка коренного малочисленного народа Российской Федерации – стремительно сокращается из года в год. Кроме этого, для вышеназванных языков характерно сужение языкового пространства, поскольку говоры постепенно уходят вместе с деревнями, а молодому поколению преподаются нормированные варианты языков. При этом именно говоры способствуют сохранению национально-культурной идентичности народа.

Институт языка, литературы и истории КарНЦ РАН является ведущим российским центром исследования прибалтийско-финских языков России, традиции которого заложены еще основателем современного российского финно-угроведения Д. В. Бубрихом. Начиная с 1940-х годов [Кузнецова 2019, 185] накоплен обширный полевой материал, проведены исследования главных параметров прибалтийско-финских языков Карелии на всех языковых уровнях и выявлены их основные «лакуны». В Фонограммархиве института хранится 450 часов вепсских и около 3000 часов карельских записей [ФА]. Работа по сбору диалектного материала продолжается и в настоящее время.

Цифровизация архивных аудио-образцов карельской и вепсской речи не только упрощает обработку и хранение материалов, но также позволяет ввести в научный оборот и представить в открытый доступ уникальные аудиоматериалы, отражающие состояние карельских и вепсских диалектов начиная с середины прошлого столетия. Именно аудиокolleкции ФА стали одним из основных источников для наполнения корпуса аудиозаписями карельской и вепсской речи. Исследователи определили список говоров карельского и вепсского языков для дальнейшего отбора аудиоматериалов с целью их переноса в Речевой корпус. Была поставлена задача максимально полно представить все многообразие живой и утраченной прибалтийско-финской диалектной речи не только в Республике Карелия, но также в Мурманской, Вологодской, Ленинградской, Новгородской и Тверской областях. Решение расширить границы проекта до Северо-Запада Российской Федерации было принято в связи с тем, что носители карельского и вепсского языков проживают за пределами республики, и их говоры представляют собой не меньшую ценность для решения проблем карельской и вепсской диалектологии. В настоящий момент уже проанализированы материалы тверских и людиковских аудиокolleкций ФА, а именно выявлены населенные пункты, в которых производились записи; все записи были распределены по годам; отмечено наличие оцифрованных копий и расшифровок для отдельных



кассет; произведено их сопоставление с опубликованными образцами речи; все выявленные записи распределены по диалектам [Новак 2021, Родионова 2022].

№	Язык	Диалект	Подкорпус	Жанр	Заголовок	Перевод
1	вепсский	Средневепсский западный	диалектные тексты	бытовой рассказ	Maria Košeleva. <i>Kut leib tehtihe</i>	Кошелева Мария. Как хлеб заготовливали
2	карельский: ливвиковское наречие	Новописьменный ливвиковский	художественные тексты		Nikolai Karpin. <i>Vala vai, Vasle!</i>	

Рис. 1. Система поиска в Открытом корпусе вепсского и карельского языков

В настоящее время Речевой корпус пополнился аудиофрагментами, записанными в 1959–1990 годы на магнитофонные пленки: на кестеньгских, юшкозерских, подужемских, поросозерских, тихвинских, валдайских, весьегонских и толмачевских говорах собственно карельского наречия карельского языка, а также на северно- и средневепсских говорах. Кроме того, для наполнения Речевого корпуса были подобраны аудиозаписи, произведенные сотрудниками института уже в цифровом формате в 2003–2021 годы: керетьские, оулангские, кестеньгские, вокнаволоцкие, ребольские, паданские, толмачевские говоры собственно карельского наречия, средне-, южнолюдиковские и михайловские говоры людиковского наречия, а также южновепсские говоры.

Кроме материалов аудиоколлекций ФА Речевой корпус пополняется полевыми аудиозаписями, произведенными в местах компактного проживания карелов и вепсов, а также фрагментами аудиозаписей радиопередач на ливвиковском наречии карельского языка, подготовленными сотрудниками ГТРК «Карелия» на сяозерских, тулмозерских, ведлозерских, видлицких, коткозерских, рыпушкальских и некулльских говорах. Для расшифровки аудиофрагментов был разработан набор транскрипционных символов. В его основу положены как упрощенная финно-угорская транскрипция, с одной стороны не перегружающая текст дополнительными диакритическими знаками, в чем нет необходимости при наличии аудиозаписей, а с другой, отражающая на письме важнейшие фонетические диалектные особенности устной речи, так и современные алфавиты новописьменных карельских и вепсских языков, что позволяет значительно расширить пользовательскую аудиторию создаваемого ресурса. Все аудиофрагменты Речевого корпуса переведены на русский язык, а также произведена их полная разметка, грамматическая (определены грамматические формы каждого слова текста) и семантическая (отмечены значения каждого слова текста). Помимо этого, выявлены основные фонетические диалектодифференцирующие черты, к которым следует отнести, в первую очередь, особенности систем восходящих нисходящих дифтонгов и переднеязычных щелевых согласных, конечную огласовку начальных и словоизменительных форм слов, особенности альтернативной системы согласных и др. В ходе морфологической разметки расшифрованных текстов были определены основные диалектные особенности грамматических систем карельских и вепсских говоров, к которым стоит отнести особенности падеж-

ных систем, отличия в количестве временных форм условного наклонения, особенности образования возвратных глагольных форм и пр. Семантическая разметка позволила выявить лексические междиалектные соответствия, т. е. были определены лексемы, имеющие в говорах отличное значение, или выявлены понятия, для именованя которых в говорах используются разные слова.

Проделанная работа позволила создать модуль «Речевой корпус» в котором представлены тексты корпуса, сопровождаемые аудиозаписями, а также необходимые для работы поисковые фильтры (поиск по языку / диалекту, месту и году записи, информанту и собирателю, источнику) (рис. 2).

Язык	Диалект	Заголовок
<input type="text"/>	<input type="text"/>	<input type="text"/>
Область, республика записи	Район записи	Населенный пункт записи
<input type="text"/>	<input type="text"/>	<input type="text"/>
Область, республика рожд. информанта	Район рожд. информанта	Населенный пункт рожд. информанта
<input type="text"/>	<input type="text"/>	<input type="text"/>
Информант	Собиратель	Год (с) Год (по)
<input type="text"/>	<input type="text"/>	<input type="text"/> <input type="text"/>
Источник	по <input type="text"/> 10 записей ПОКАЗАТЬ	
<input type="text"/>		

Найдено 82 записи. Простой поиск ↑

No	Диалект	Говор	Заголовок	Перевод	Прослушать
1	Средневепский западный	<u>Немжа</u>	Maria Košeleva. <u>Kut leib tehtihe</u>	Кошелева Мария. Как хлеб заготавливали	0:00 / 1:30
2	Северновепский	<u>Рыбрека</u>	<u>Kil'bet'</u>	Баня	0:00 / 1:03

Рис. 2. Речевой корпус прибалтийско-финских языков Карелии

В настоящий момент Речевой корпус содержит более шестидесяти аудиофрагментов, длительностью от одной до трех минут, представляющих собой разнообразие карельской и веппской устной диалектной речи. Особую ценность представляет фрагмент записи валдайской речи, единственный обнаруженный к настоящему времени (рис. 3).

Tullah kozimmiehet

подкорпус: диалектные тексты

информант(ы): Павлова Мария Егоровна, 1900, Марково, Валдайский р-н, Новгородская обл.
место записи: Марково, Валдайский р-н, Новгородская обл., г. записи: 1990
записали: Пунжина Александра Васильевна

источник: Слушаю карельский говор. Образцы речи дёржанских и валдайских карел / сост. Пунжина А.В., 2001, с. 180-181



Tullah kozimmiehet

Карельский: собственно карельское наречие Валдайский

– Kozittih milma kozimmiehet. Rista tuli. Ka šanou: "Kozimmiehet tullah." No olla Kozittih, kozittih. A mie šanou tällä... "Seičas en lähe, vuuvveksi jätän." Jätin Ka vuotti, gul'aičči. I znakomči, gul'aičči hänenke kaikkii. A šiid' tuli kocižčoma Mie hänellä en otkažin. Ei pidän i... " srazu i nainun. A nyt, – šanou, – mid

tulla VERB (1)
возвращаться, заходить,
приходить, прибывать,
приезжать)
индикатив, имперфект, 3
л., ед. ч., полож. ф.

Пунжина Александра Васильевна

Сваты приходят

Русский

ня сваты. Крестный пришел. И
аты придут." Ну, пусть приходят
ти, сватали. А я и говорю... Я
час не пойду, на год оставлю."
год. Вот ждал, гулял.
я, гуляли с ним все. А потом
аться, сосватал. Я ему не

Рис. 3. Пример страницы текста Речевого корпуса

Для облегчения работы пользователей с Речевым корпусом и для обеспечения возможности наглядного представления звукового материала была разработана мультимедийная аудио-карта говоров прибалтийско-финской речи Карелии и сопредельных областей (рис. 4).



Рис. 4. Аудиокарта говоров прибалтийско-финской речи Карелии и сопредельных областей

На карте нашли отражение все подготовленные аудиофрагменты на ливвиковском наречии (ведлозерский, видлицкий, коткозерский, неккульский, рыпушкальский, сямозерский, тулмозерский диалекты), аудио-образцы на людиковском наречии (среднелюдиковский, южнолюдиковский и михайловский диалекты), на собственно карельском наречии (валдайский, весьегонский, вокनावолокский, дёржанский, керетьский, кестеньгский, оулангский, паданский, подужемский, поросозерский, ребольский, тихвинский, толмачевский, юшкозерский диалекты) и на вепском языке (северновепский, средневеппский восточный, средневеппский западный и южновеппский диалекты). Работа над пополнением Речевого корпуса продолжится и в будущем, с целью представить все многообразие живой и утраченной прибалтийско-финской диалектной речи Карелии.

Заключение

Значимость Речевого корпуса связана с тревожной ситуацией, в которой пребывают прибалтийско-финские языки Карелии. Численность носителей карельского языка, являющегося языком титульной нации республики, и вепского – языка коренного малочисленного народа Российской Федерации – стремительно сокращается из года в год. Практическая значимость Речевого корпуса продиктована также постоянно возрастающим объемом запросов от общественных организаций, культурных объединений и рядовых граждан на обнаружение «местных» – на уровне говоров и диалектов – материалов на карельском и вепском языках. При этом заинтересованных граждан, работников культуры и национальную общественность, наряду с фольклорными произведениями и материалами этнографического и топонимического характера, в последние годы все чаще интересуют собственно языковые особенности родных населенных пунктов. Отметим, что в местах компактного проживания карелов и вепсов преподавание национальных языков в школе, обучение населения на языковых курсах, использование фольклорных и языковых элементов в туризме, музейном деле и т. д., за редким исключением, осуществляются на литературных формах карельского и вепского языков. С этой точ-

ки зрения Речевой корпус удовлетворит запросам граждан, заинтересованных в изучении и сохранении именно местных форм языка и культуры. Актуальность запланированных нами исследований определяется не только их собственно научной составляющей, но и востребованностью практики языкового строительства, а также необходимостью сохранения диалектов карельского и вепского языков. Достижение поставленных целей будет способствовать как поддержанию жизнеспособности, популяризации и повышению грамматического потенциала новописьменных карельского и вепского языков, так и ревитализации и поднятию престижа диалектов этих прибалтийско-финских языков.

ЛИТЕРАТУРА

- Баранцев А. П.* Образцы людиковской речи (образцы корпуса людиковского идиолекта) / А. П. Баранцев. Петрозаводск: Карелия, 1978. 287 с.
- Бойко Т. П.* Лингвистический корпус ВепКар – «заповедник» прибалтийско-финских языков Карелии / Бойко Т. П., Зайцева Н. Г., Крижановская Н. Б., Крижановский А. А., Новак И. П., Пеллинен Н. А., Родионова А. П., Трубина Е. Д. // Труды КарНЦ РАН. – 2021. – № 7. – С. 100-115
- БНК – Британский национальный корпус: URL: <http://www.natcorp.ox.ac.uk/> (дата обращения: 18.11.2022).
- ВепКар – Открытый корпус вепского и карельского языков: URL: <http://www.dictorpus.krc.karelia.ru/> (дата обращения: 05.12.2022)
- ВНК – Венгерский национальный корпус: URL: http://mnsz.nytud.hu/index_eng.html/ (дата обращения: 16.11.2021).
- Зайцева Н. Г.* Вепские причитания в фокусе корпусной лингвистики и лингвофольклористики / Зайцева Н. Г. // Материалы ХLI Международной филологической конференции. 26-31 марта 2012 г. Секция «Уралистика». – Санкт-Петербург: Филологический факультет СПбГУ, 2012. – С. 16-26.
- Зайцева Н. Г.* Корпусная лингвистика в прибалтийско-финском исследовательском пространстве (на материале Корпуса вепского языка и Открытого корпуса вепского и карельского языков) / Зайцева Н. Г., Крижановская Н. Б. // Альманах североевропейских и балтийских исследований. – 2018. – Выпуск 3. – С. 264-273.
- ККЗЯ – Корпус коми-зырянского языка: URL: <http://komi-zyrian.webcorpora.net/> (дата обращения: 16.11.2022)
- ККПЯ – Корпус коми-пермяцкого языка: URL: <http://komi-permyak.web-corpora.net/> (дата обращения: 16.11.2022)
- КЛМЯ – Корпус лугового марийского языка: URL: <http://meadow-mari.web-corpora.net/> (дата обращения: 16.11.2022).
- Крижановский А. А.* Представление диалектов в Открытом корпусе вепского и карельского языков (ВепКар) / Крижановский А. А., Крижановская Н. Б., Новак И. П. // Труды международной конференции «Корпусная лингвистика – 2019». СПб, 2019. С. 288-295.
- Кузнецова В. П.* Формирование фондов Фонограммархива ИЯЛИ КарНЦ РАН, проблемы их сохранения и систематизации / Кузнецова В. П. // Кижский вестник. 2017. Вып. 17. С. 185-192.
- Макаров Г. Н.* Образцы карельской речи / Г. Н. Макаров, В. Д. Рягоев. Ленинград: Наука, 1969. 283 с.
- НКРЯ – Национальный корпус русского языка: URL: <http://www.ruscorpora.ru/> (дата обращения: 16.11.2022).
- НКУЯ – Национальный корпус удмуртского языка: URL: <http://udmcorpus.udman.ru/> (дата обращения: 16.11.2022).
- Новак И. П.* Коллекция тверских карельских диалектных материалов в Фонограммархиве ИЯЛИ КарНЦ РАН / Новак И. П. // Ученые записки Петрозаводского государственного университета. 2021. Т. 43, № 1. С. 41–51.
- Родионова А. П.* О коллекциях людиковских диалектных материалов Фонограммархива ИЯЛИ КарНЦ РАН / Родионова А. П. // Ученые записки Петрозаводского государственного университета. 2022 Т. 44, № 7 С. 64–70
- СКЭЯ – Сводный корпус эстонского языка: URL: <https://www.cl.ut.ee/korpused/segakorpus/index.php?lang=en/> (дата обращения: 16.11.2022).
- ФА – Фонограммархив ИЯЛИ КарНЦ РАН: URL: <http://phonogr.krc.karelia.ru/> (дата обращения: 16.11.2022).
- ЧНК – Чешский национальный корпус: URL: <https://www.korpus.cz/> (дата обращения: 17.11.2022).
- ЯБФ – Языковой банк Финляндии: URL: <https://www.kielipankki.fi/aineistot/ftc/> (дата обращения: 16.11.2022).



Boyko Tatyana. The Open corpus of the Veps and Karelian languages: overview and applications / Boyko Tatyana, Zaitseva Nina, Krizhanovskaya Natalia, Krizhanovsky Andrew, Novak Irina, Pellinen Nataliya, Rodionova Aleksandra // KnE Social Sciences. –7(3). – 2022. – P. 29–40.

Krizhanovskaya, N. Morphological inflectional rules for Karelian Proper verbs / Krizhanovskaya, N., Novak, I., Krizhanovsky, A., Pellinen, N. Eesti Ja Soome-Ugri Keeleteaduse Ajakiri. Journal of Estonian and Finno-Ugric Linguistics, 13(2). 2022. P. 47–78.

Näytteitä karjalan kielestä I (Образцы карельской речи). Joensuu-Петрозаводск, 1994. 455 с.

Поступила в редакцию 29.05.2023

Родионова Александра Павловна

Кандидат филологических наук
научный сотрудник сектора языкознания
Института языка, литературы и истории
Карельского научного центра РАН
185910, Россия, г. Петрозаводск, ул. Пушкинская, 11
E-mail: santrar@krc.karelia.ru, sashenka22@yandex.ru

Крижановская Наталья Борисовна

ведущий инженер-исследователь
лаборатории информационных
компьютерных технологий
Институт прикладных математических исследований
Карельского научного центра РАН
185910, Россия, г. Петрозаводск, ул. Пушкинская, 11
E-mail: nataly@krc.karelia.ru

Пеллинен Наталия Александровна

Кандидат филологических наук,
младший научный сотрудник
сектора языкознания
Института языка, литературы и истории
Карельского научного центра РАН
185910, Россия, г. Петрозаводск, ул. Пушкинская, 11
E-mail: nataliapellinen@gmail.com

A. P. Rodionova, N. B. Krizhanovskaya, N. A. Pellinen

**VEPKAR SPEECH CORPUS AS A TOOL TO PRESERVE THE DIALECT SPEECH
OF THE BALTIC-FINNISH PEOPLE OF KARELIA**

DOI: 10.35634/2224-9443-2023-17-3-343-351

The article is devoted to the description of the Speech Corpus of the Baltic-Finnish Speech, created on the platform of the Open Corpus of the Vepsian and Karelian Languages (VepKar), its architecture and possibilities. The speech corpus was developed by the staff of the Institute for the Language, Literature and History and the Institute of Applied Mathematical Research. The corpus includes a collection of spoken texts in different dialects of the Karelian and Vepsian languages, provided with transcription, markup and translation into Russian. The corpus also contains search filters necessary for work (search by language/dialect, place and year of recording, informant and collector, source). The need to develop the VepKar corpus is very relevant, the corpus is in great demand both in scientific research and in the process of developing the literary forms of the Karelian and Vepsian languages. The use of modern technologies and methods, the field material accumulated over many decades and the latest data will make it possible to fill in a number of gaps that were previously identified by linguists in this system. Researchers use three main sources to fill the corpus with audio recordings of Karelian and Vepsian speech: audiocollections of the Phonogram Archive of the ILLH KRC RAS, audiorecordings of broadcasts in the Livvic dialect of the Karelian language, as well as field materials of the authors recorded during the expeditions. Scientific novelty is justified by the lack of speech corpora of the Baltic-Finnish languages. Digitization of archival and field audio samples of Karelian and Vepsian speech in the Speech Corpus format will further simplify the processing and storage of materials. It will also make it possible to introduce unique audio ma-

materials reflecting the state of the Karelian and Vepsian dialects since the middle of the last century into scientific circulation and make them available to the public.

Keywords: Speech corpus, Vepsian language, Karelian language, corpus linguistics, audio samples, text markup.

Citation: Yearbook of Finno-Ugric Studies, 2023, vol.17, issue 3, pp. 343–351. In Russian.

REFERENCES

Barantsev A. P. *Obraztsy lyudikovskoy rechi (obraztsy korpusa lyudikovskogo idiolekta)* [Samples of Ludic's speech (samples of the corpus of Ludic's idiolect)]. Petrozavodsk: Karelia, 1978. 287 p. In Russian, In Karelian.

Boyko T. P., Zaitseva N. G., Krizhanovskaya N. B., Krizhanovsky A. A., Novak I. P., Pellinen N. A., Rodionova A. P., Trubina E. D. *Lingvisticheskiy korpus VepKar – «zapovednik» pribaltiysko-finskiy yazykov Karelii* [The linguistic corpus VepKar is a language refuge for the Baltic Finnish language of Karelia]. *Trudy Karelskogo nauchnogo tsentra Rossiyskoy akademii nauk.* [Transactions of the Karelian Research Centre of the Russian Academy of Sciences]. 2021, 7. P. 100-117. In Russian.

BNK – *Britanskiy national'nyy korpus* [British national corpus]. Available at: <http://www.natcorp.ox.ac.uk/> (accessed November 11, 2022). In English.

VepKar – *Otkrytyy korpus vepsskogo i karel'skogo yazykov* [Open corpus of Vepsian and Karelian languages]. Available at: <http://www.dictopus.krc.karelia.ru/> (accessed December 05, 2022). In Karelian, In Vepsian, In Russian, In English.

VNK – *Vengerskiy natsional'nyy korpus* [Hungarian national corpus]. Available at: http://mnsz.nyud.hu/index_eng.html/ (accessed November 16, 2021). In Hungarian.

Zaiceva N. G. *Vepsskie prichitaniya v fokuse korpusnoj lingvistiki i lingvofol'kloristiki* [Vepsian lamentations in the focus of corpus linguistics and linguistic folkloristics]. *Materialy XLI Mezhdunarodnoj filologicheskoy konferencii.* 26-31 marta 2012 g. Sekciya "Uralistika" [Materials of the XLI International Philological Conference. March 26-31, 2012 Section "Uralistics"]. SPb.: Filologicheskij fakul'tet SPbGU, 2012. P. 16-26. In Russian.

Zaiceva N. G., Krizhanovskaya N. B. *Korpusnaya lingvistika v pribaltiysko-finskom issledovatel'skom prostranstve (na materiale Korpusa vepsskogo yazyka i Otkrytogo korpusa vepsskogo i karel'skogo yazykov)* [Corpus linguistics in the Baltic-Finnish research space (based on the Vepsian Language Corpus and the Veps and Karelian Open Corpus)] *Al'manakh severoevropeskikh i baltiyskikh issledovaniy* [Almanac of Northern European and Baltic Studies]. Issue 3, 2018. P. 264-273. In Russian.

KKZJA – *Korpus komi-zyrianskogo yazyka* [Corpus of Komi-Zyrian language]. Available at: <http://komi-zyrian.webcorpora.net/> (accessed November 16, 2022). In Komi.

KKPJA – *Korpus komi-zyrianskogo yazyka* [Corpus of Komi-Permian language]. Available at: <http://komi-permyak.web-corpora.net/> (accessed November 16, 2022). In Komi.

KLMJA – *Korpus lugovogo marijskogo yazyka* [Corpus of Meadow Mari language]. Available at: <http://meadow-mari.web-corpora.net/> (accessed November 16, 2022). In Mari.

Krizhanovskij A. A., Krizhanovskaya N. B., Novak I. P. *Predstavlenie dialektov v Otkrytom korpusse vepsskogo i karel'skogo yazykov (VepKar)* [Representation of dialects in the Open Corpus of Veps and Karelian languages (VepKar)]. *Trudy mezhdunarodnoj konferencii «Korpusnaya lingvistika – 2019»* [Transactions of the international conference "Corpus linguistics – 2019"]. SPb, 2019. P. 288-295. In Russian.

Kuznetsova V. P. *Formirovanie fondov Fonogrammarhiva IYALI KarNC RAN, problemy ih sohraneniya i sistematizatsii* [Formation of the funds of Phonogram archive of the ILLH KRC RAS, problems of their preservation and sistematization]. *Kizhskij vestnik* [Kizhi's Bullenten]. 2017, 17. P. 185-192. In Russian.

Makarov G. N., Ryagoyev V. D. *Obraztsy karel'skoy rechi* [Samples of Karelian speech]. Leningrad: Nauka, 1969. 283 p. In Karelian.

NKRJA – *Natsionalnyy korpus russkogo yazyka* [Russian national corpus]. Available at: <http://www.ruscorpora.ru/> (accessed November 16, 2022). In Russian.

NKUJA – *Natsionalnyy korpus udmurtskogo yazyka* [National corpus of the Udmurtian language]. Available at: <http://udmcorpus.udman.ru/> (accessed November 16, 2022). In Udmurtian.

Novak I. P. *Kollektsiya tverskih karelskikh dialektnykh materialov v fonogrammarhive IJALI KarNTS RAN* [Collection of Tver Karelian materials in the phonogram archive of the Institute of Linguistics, Literature and History of the Karelian Research Centre of the Russian Academy of Sciences]. *Uchenye zapiski Petrozavodskogo gosudarstvennogo universiteta* [Proceedings of Petrozavodsk State University]. 2021, 43(1). P. 41–51. In Russian.

Rodionova A. P. *O kollektsiyah lyudikovskikh dialektnykh materialov Fonogrammarhiva IJALI KarNTS RAN* [Collection of the Ludic materials in the phonogram archive of the Institute of Linguistics, Literature and History of the Karelian Research Centre of Academy on Sciences]. *Uchenye zapiski Petrozavodskogo gosudarstvennogo universiteta* [Proceedings of Petrozavodsk State University]. 2022, 44 (7). P. 64-70. In Russian.

SKEJA – *Svodnyy korpus estonskogo yazyka* [Consolidated corpus of the Estonian language]. Available at:



<https://www.cl.ut.ee/korpused/segakorpus/index.php?lang=en/> (accessed November 16, 2022). In Estonian.

FA – *Fonogrammarhiv IJALI KarNTS RAN* [Phonogram archive of the ILLH KRC RAS]. Available at: <http://phonogr.krc.karelia.ru/> (accessed November 16, 2022). In Karelian, In Vepsian, In Russian.

CHNK – *Cheshskij natsionalnyj korpus* [Czechian national corpus]. URL: <https://www.korpus.cz/> (accessed November 17, 2022). In Czechian.

YABF – *Yazykovoj bank Finlyandii* [Language bank of Finland]. URL: <https://www.kielipankki.fi/aineistot/ftc/> (accessed November 16, 2022). In Finnish.

Boyko T., Zaitseva N., Krizhanovskaya N., Krizhanovsky A., Novak I., Pellinen N., Rodionova A. *The Open corpus of the Veps and Karelian languages: overview and applications. KnE Social Sciences.* 2022, 7(3). P. 29–40. In English.

Krizhanovskaya, N., Novak, I., Krizhanovsky, A., Pellinen, N. *Morphological inflectional rules for Karelian Proper verbs. Eesti Ja Soome-Ugri Keeleteaduse Ajakiri. Journal of Estonian and Finno-Ugric Linguistics.* 2022, 13(2). P. 47–78. In English.

Näytteitä karjalan kielestä I. Joensuu-Petrozavodsk, 1994. 455 p. In Karelian.

Received 29.05.2023

Rodionova Alexandra Pavlovna

Candidate of Philology,
Research Associate in the Linguistic Section
Institute of Language, Literature and History
of Karelian Research Centre RAS.
11, Pushkinskaya str., Petrozavodsk, 185910, Russia
E-mail: santrar@krc.karelia.ru, sashenka22@yandex.ru

Krizhanovskaya Natalya Borisovna

Leading Research Engineer of the Laboratory for
Information Computer Technologies
Institute of Applied Mathematical Research
of Karelian Research Centre RAS
11, Pushkinskaya str., Petrozavodsk, 185910, Russian Federation
E-mail: nataly@krc.karelia.ru

Pellinen Natalia Alexandrovna

Candidate of Philology
Junior Research Associate in the Linguistic Section
Institute of Language, Literature and History
of Karelian Research Centre RAS
11, Pushkinskaya str., Petrozavodsk, 185910, Russian Federation
E-mail: nataliapellinen@gmail.com