



РОССИЙСКАЯ АКАДЕМИЯ НАУК

ОТДЕЛЕНИЕ

ИСТОРИКО-ФИЛОЛОГИЧЕСКИХ НАУК

ТРУДЫ
ОТДЕЛЕНИЯ
ИСТОРИКО-
ФИЛОЛОГИЧЕСКИХ
НАУК

2023

Том 13

Ответственный редактор
академик РАН В.А. ТИШКОВ

МОСКВА, 2024

УДК 001.32
ББК 72.3
Т 78

Издание основано в 2003 году

Редакционная коллегия:

В.Л. КЛЯУС, Н.А. МАКАРОВ, А.М. МОЛДОВАН, Е.И. ПИВОВАР,
Н.В. ТАРАСОВА (составитель), В.А. ТИШКОВ (ответственный редактор)

Рецензенты:

член-корреспондент РАН, доктор исторических наук Д.М. БОНДАРЕНКО
член-корреспондент РАН, доктор филологических наук А.В. ДЫБО

Научно-организационная подготовка А.А. САРИБЕГЯН

Т 78 **Труды Отделения историко-филологических наук РАН.** / Отд-ние
ист.-филол. наук РАН. – М.: РАН, 2024 – 428 с.

2023. Том 13 / отв. ред. В.А. Тишков; [сост. Н.В. Тарасова]. – 2024. - 428 с.

ISBN 978-5-907645-23-3

Тринадцатый том ежегодника «Труды Отделения историко-филологических наук РАН» представляет материалы научных сессий общих собраний и заседаний бюро ОИФН РАН. Авторами первой части (научная сессия «Культурное наследие России: академическая наука и практика», 22 мая 2023 г.) обсуждаются проблемы изучения и сохранения научного, археологического, музейного, архитектурного, литературного наследия нашей страны. Во второй части опубликованы материалы научной сессии «Трёхсотлетие российской науки» (11 декабря 2023 г.), в которых история Российской академии наук рассматривается через этапы и события, определившие тенденции развития отечественной науки в XVIII-XX вв. Авторы докладов стратегической сессии бюро ОИФН РАН «Информационные технологии и мониторинг этнокультурных и языковых трансформаций в России» (18 октября 2023 г.) дают представление о работе по созданию информационных ресурсов изучения, сохранения, популяризации знаний об этнической, фольклорной и языковой культуре народов Российской Федерации. Также в книге публикуются статьи членов РАН, профессоров РАН, представителей вузов по вопросам истории, филологии, истории науки, культурного и научного наследия.

Для историков, филологов, культурологов, широкого круга читателей.

ISBN 978-5-907645-23-3

© Отделение историко-филологических наук РАН, 2024
© Тарасова Н.В., составление, 2024

СОДЕРЖАНИЕ

<i>Тишков В.А.</i> Предисловие	6
--------------------------------------	---

Часть 1

Научная сессия Общего собрания Отделения историко-филологических наук РАН «Культурное наследие России: академическая наука и практика» (22 мая 2023 г.)

<i>Пиотровский М.Б.</i> Мосты культуры и войны памяти	14
<i>Макаров Н.А.</i> Московский Кремль как археологический памятник	31
<i>Швидковский Д.О.</i> Восстановление исторического облика России: архитектурные проблемы	50
<i>Дмитриева Е.Е.</i> «Тарас Бульба» Гоголя: прагматика vs герменевтика	59

Часть 2

Научная сессия Общего собрания Отделения историко-филологических наук РАН «Трёхсотлетие российской науки» (11 декабря 2023 г.)

<i>Басаргина Е.Ю.</i> Столетний юбилей Академии наук в 1826 г.	80
<i>Фандо Р.А.</i> 300 лет служения человечеству. Открытия Российской академии наук, изменившие мир	95
<i>Батурин Ю.М.</i> Неархимедова мера истории	109
<i>Бондарь Л.Д., Тункина И.В.</i> Академия наук СССР и Белорусская академия наук в начале 1930-х гг.: наставничество и взаимодействие	126
<i>Побережников И.В., Артёмов Е.Т.</i> Наука и технологии в контексте советской позднеиндустриальной модернизации	136
<i>Рынков В.М.</i> Новосибирский Академгородок: технология «броска» советской науки на восток	143

Часть 3
Стратегическая сессия Бюро
Отделения историко-филологических наук РАН
«Информационные технологии и мониторинг
этнокультурных и языковых трансформаций в России»
(18 октября 2023 г.)

<i>Павленко О.В., Борисов Н.А., Баскакова И.А.</i> «Интерактивный атлас коренных малочисленных народов Севера, Сибири и Дальнего Востока: языки и культуры»: исследовательский и просветительский ресурс по цифровизации и популяризации природного и культурного наследия коренных народов России	155
<i>Черных А.В., Чернышева Ю.С.</i> Реестры объектов нематериального этнокультурного достояния Российской Федерации: опыт и новые задачи	170
<i>Муллонен И.И., Новак И.П., Родионова А.П., Крижановская Н.Б.</i> Открытый корпус вепского и карельского языков: структура, функциональность, перспективы развития	187
<i>Силантьев И.В., Кошкарёва Н.Б., Федина Н.Н., Шапошникова И.В., Кузьмина Е.Н., Гриневич А.А., Шахов П.С., Тирон Е.Л.</i> Электронные базы данных языков и фольклора народов Сибири и Дальнего Востока	197

Часть 4
Актуальные вопросы гуманитарных наук

<i>Алпатов В.М.</i> Факторы, способствующие исчезновению языков, и факторы, способствующие их сохранению	206
<i>Бурлак С.А.</i> Языковое родство и языковые контакты	213
<i>Пивовар Е.И., Катагощина М.В.</i> Российское интеллектуальное зарубежье XX – начала XXI века: глобальные миграции идей и технологий	237
<i>Зайцев И.В.</i> Коран Узбека: ильханидская рукопись для золотоордынского хана	254
<i>Сиренов А.В.</i> Дискуссия о типах письма древнерусских рукописей	265
<i>Черкасов П.П.</i> Попытка реставрации монархии во Франции в 1873 году (по свидетельству российского посла князя Н.А. Орлова)	275
<i>Репина Л.П.</i> «Спасительные гавани» историка, или о пользе изучения интеллектуальных традиций	289

Часть 5

Из истории науки

<i>Алиева А.И.</i> Фольклор, этнография, религиозные верования ингушей в российских научных изданиях второй половины XIX – начала XX вв.	306
<i>Вахромеева О.Б.</i> К истории списка трудов Ольги Антоновны Добиаш-Рождественской	339
<i>Ананьев В.Г., Бухарин М.Д.</i> «...поставить русскую науку в благоприятные условия для плодотворной научной работы». М.И. Ростовцев о политике археологических исследований	361
<i>Батурин Ю.М.</i> Академия наук и успехи космонавтики (малоизвестные страницы истории)	391
СПИСОК СОКРАЩЕНИЙ	420
СВЕДЕНИЯ ОБ АВТОРАХ	424

Открытый корпус вепского и карельского языков: структура, функциональность, перспективы развития¹

И.И. Муллонен, И.П. Новак, А.П. Родионова, Н.Б. Крижановская

Аннотация. ВепКар – Открытый корпус вепского и карельского языков, ведущий свою историю с 2009 г. Тогда была начата работа над «Корпусом вепского языка», а в 2016 г. было принято решение расширить его за счёт языка титульного этноса Карелии – карельского. За годы работы над ресурсом языковедами ИЯЛИ КарНЦ РАН и программистами ИПМИ КарНЦ РАН была проделана большая работа по его наполнению не только текстами, но и словарями, а также инструментами, обеспечивающими удобную работу и возможность проведения научных исследований. В настоящий момент в корпусе размещено около 6 тыс. текстов (2 млн. словоупотреблений) на диалектах и нормированных вариантах карельского и вепского языков, более 67 тыс. словарных статей. ВепКар стал базой для исследования прибалтийско-финских языков Северо-Запада России. Материалы корпуса активно используются в процессах разработки новых словарей и грамматик, редактирования правил и норм новописьменных вариантов языков, в обучающих процессах. Они послужили основой для таких ресурсов, как Аудиокарта прибалтийско-финских языков Карелии, Приложение для проверки орфографии и Мультимедийный словарь карельского языка LiPaS. При условии дальнейшего расширения объёма и структурного потенциала платформы в будущем эти данные планируется направить на решение таких сложных задач как, например, распознавание и синтез речи, машинный перевод, построение компьютерных моделей карельского и вепского языков.

Ключевые слова: карельский язык; вепский язык; корпусная лингвистика; Открытый корпус вепского и карельского языков.

Развитие корпусной лингвистики в Институте языка, литературы и истории Карельского научного центра Российской академии наук стало возможным благодаря тесной и плодотворной работе с программистами и исследователями Института прикладных математических исследований Карельского научного центра Российской академии наук.

Открытый корпус вепского и карельского языков, созданный с целью сохранения языкового богатства и последующего изучения прибалтийско-финских языков Карелии, их развития и популяризации, ведёт свою историю с 2009 г., когда под руководством д.ф.н. Н.Г. Зайцевой была начата работа над «Корпусом вепского языка²», который включил в себя пять текстовых под-

¹ Статья подготовлена в рамках выполнения госзадания ФГБУ ФИЦ «Карельский научный центр Российской академии наук»

² См. <http://vepsian.krc.karelia.ru/about/>

корпусов (диалектные тексты, два подкорпуса фольклорных текстов и два младодисциплинарных корпуса), электронный словарь и системы поиска [Бойко и др., 2021: 102; Родионова и др., 2023: 343]. «Корпус вепского языка» оказал большую помощь в подготовке «Орфографического словаря вепского языка» в процессе отбора форм для создания правил орфографии [Зайцева и др., 2012].

В 2016 г. было принято решение расширить корпус за счёт включения в него языка титульного этноса Карелии – карельского, поскольку для карельского народа проблемы, связанные с ревитализацией языка, не менее актуальны, чем для вепского. Оба языка находятся под серьёзной угрозой исчезновения. Новый корпус получил название «Открытый корпус вепского и карельского языков», сокращенно – ВепКар (<http://dictorpus.krc.karelia.ru>). За эти годы языковедами и математиками КарНЦ РАН была проделана большая работа по наполнению ресурса не только текстами, но и словарями, а также инструментами, обеспечивающими удобную работу в корпусе и возможность проведения научных исследований.

В первые годы корпус развивался благодаря поддержке Программы Отделения историко-филологических наук РАН и грантам [Зайцева, Крижановская, 2018]. С 2021 г. направление корпусной лингвистики было включено в план научно-исследовательской работы сектора языкознания ИЯЛИ КарНЦ РАН.

Структура и функциональность корпуса ВепКар

ВепКар – это электронный открытый динамический многоязычный разножанровый полнотекстовый диахронический размеченный смешанный многофункциональный языковой корпус (рис. 1).



Рис. 1. Структура ВепКар. Данные по состоянию на 4 марта 2024 г.

Открытость корпуса подразумевает наличие свободного доступа и отсутствие ограничений на использование материалов. Динамичность – его постоянное пополнение. Сегодня ВепКар представлен 6 тыс. текстами (2 млн. словоупотреблений). Источниками для пополнения служат периодические издания, сборники диалектных и фольклорных текстов, учебные пособия, художественные произведения. Наполнение ресурса осуществляется по соглашению с издательствами и авторами текстов. Неисчерпаемым источником материалов остаются архивы. Все тексты размещаются полностью, а не фрагментарно (полнотекстовость).

Корпус ВепКар состоит из подкорпусов. Одним из параметров их выделения является языковая принадлежность текстов. В корпусе представлен один вепсский и три карельских подкорпуса [Крижановский и др., 2019: 289]. Причиной выделения последних послужила диалектная разобщённость карельского языка. В нём функционируют три наречия (собственно карельское, ливвиковское и людиковское), на каждом из которых уже на протяжении 30 лет развивается и внедряется через систему образования и СМИ своя новописьменная традиция [Бойко и др., 2021: 105]. Наличие нормированных вариантов позволяет, с одной стороны, отталкиваться от них в процессе обработки текстов, а с другой – перепроверять существующие нормы и, в случае выявления несоответствий, корректировать их. В связи с этим важно отметить, что практически все редакторы корпуса являются членами Республиканской термино-орфографической комиссии по карельскому и вепсскому языкам при Главе Республики Карелия. Кроме того, многоязычность корпуса подразумевает наличие текстов с параллельным выровненным по предложениям переводом на русский язык (рис. 2).

Savina Renata Jurjevna. Ristitun uskond

подкорпус: художественные тексты

жанр: литературная (авторская) сказка

источник: Савина Рената Юрьевна, Kodima. № 11., 2022, с. 8

Savina Renata Jurjevna

Ristitun uskond

Вепсский
Младописьменный вепсский

Minun baboi eli vanhas karjalaizēs küläs kaiken elon. Hänen elos oli äi jügeduzid, no hän nikonz ei lanktend opalaha. Hän vägesti kaikid jügeduzid. Minä ihastun hänehe. Kerdan minä küzuin baboil, kuspäi hänes or "Minä en sanu Jumalale minun jügeduzile, miše minäi om Jum sanoid. Aig mäne, necešpäi virk minä omištan baboile.

ihastuda VERB (2)
восхищаться)

индикатив, презенс, 1 л.,
ед. ч., полож. ф.

Савина Рената Юрьевна

Сказка о человеческой вере

Русский

Моя бабушка всю жизнь прожила в старой карельской деревне. Ее жизнь была тяжелой, но она никогда не унывала. Она справлялась со всеми трудностями. Я восхищаюсь ей. Однажды я спросила как ей удается быть такой сильной. Бабушка ответила: "Я не говорю Богу о своих проблемах, а, наоборот, всем своим проблемам я говорю, что со мной Бог". Я запомнила ее слова. Спустя годы из этой фразы у меня родилась сказка, которую я посвящаю памяти моей бабушки.

Рис. 2. Пример текста из корпуса ВепКар с параллельным переводом на русский язык. Полный текст см. <http://dictorpus.krc.karelia.ru/ru/corpus/text/4611>

По стилистической принадлежности текстов в корпусе ВепКар сложились следующие подкорпусы: публицистический, художественный, диалектный, библейский, фольклорный (разножанровость) (рис. 3). Наличие жанровой системы позволяет, исходя из определённых задач, оформлять тематические коллекции текстов [Бойко и др., 2021: 107]. Сегодня на сайте корпуса представлены три оформленные фольклорные коллекции: вепские причитания, карельские руны и карельские предания³. Так, например, коллекция «Карельские руны» была подготовлена в рамках проведения в республике в 2021 г. Года карельских рун. К данной работе активно подключились фольклористы ИЯЛИ КарНЦ РАН. Для создания коллекции ими не только были отобраны, загружены в корпус и переведены 99 текстов, но и произведена их подробная классификация по сюжетам и темам. В коллекции «Вепские причитания» представлены свадебные и похоронно-поминальные причитания, их классификация дана исходя из локальных традиций.

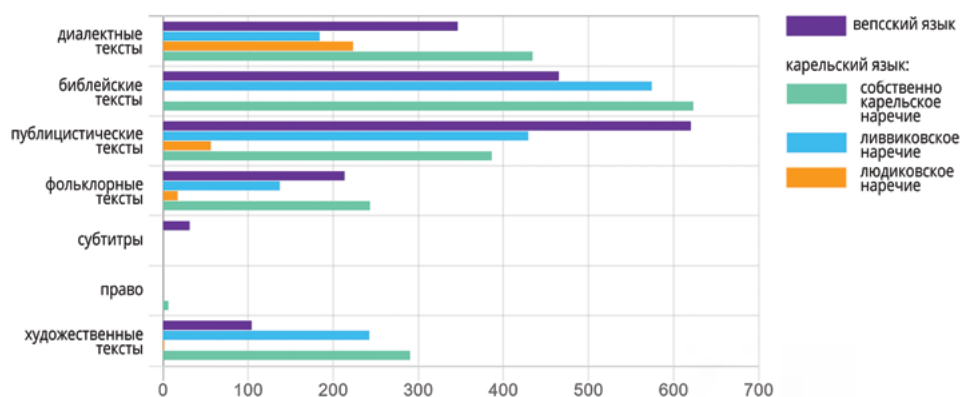


Рис. 3. Подкорпусы ВепКар. Соотношение текстов на вепском и карельском языках в подкорпусах. Данные по состоянию на 4 марта 2024 г. Актуальную информацию см. http://dictorpus.krc.karelia.ru/ru/stats/by_corpus

Рост подкорпуса художественных текстов открывает перспективы для исследования литературоведами прозы и поэзии на национальных языках. Можно, например, изучать произведения отдельно взятого автора (как вариант, на основе статистики использования им определённых частей речи, деминутивов, фразеологизмов и т.д.). А диалектные бытовые тексты, содержащие ценный этнографический материал, в свою очередь, могут быть полезны этнологам.

Основу диахронического корпуса составляют тексты, записанные в середине XX в. Самый ранний из представленных текстов (текст руны) был зафиксирован в 1888 г. ВепКар постоянно пополняется как материалами из свежих номеров газет на карельском и вепском языках, так и расшифровками полевых материалов ежегодных экспедиций.

³ См. <http://dictorpus.krc.karelia.ru/ru/corpus/collection>

В корпусе представлены три вида *разметки* (рис. 2):

- метатекстовая (название текста, дата создания, автор, жанр, место записи и др.);
- морфологическая (у слов в текстах указаны части речи и морфологические признаки);
- семантическая (слова в текстах связаны со значениями словарных статей).

Возможность морфологической и семантической разметки обеспечивает связь корпуса текстов со словарем (67 тыс. словарных статей) (рис. 1). Все загружаемые тексты автоматически размечаются: т.е. для каждого слова производится поиск соответствия в словаре. Контекст при морфологическом анализе не учитывается, анализатор выдает все возможные леммы для заданной словоформы. Обученная на массивах текстов программа в дальнейшем сможет прогнозировать нужные формы имён при глаголах (глагольное управление), определять

окончания именных и глагольных словоформ.

Сегодня автоматическая разметка достигла 78% благодаря двум вещам:

1) в словарь корпуса включены все изданные словари нормированных вариантов ливвиковского карельского и вепсского языков. С июля 2023 г. ведётся загрузка словарных статей рукописи Большого словаря собственно карельского языка П.М. Зайкова;

2) кроме слов в начальной форме в словаре представлены полные словоизменительные парадигмы (рис. 4). Разработанные алгоритмы генерации словоформ по всем подкорпусам позволили автоматически получить более 2,5 миллионов словоформ [Новак и др., 2020: 681]. При внесении нового слова редактором указываются одна-две основы, по которым генерируются полные парадигмы (более 40 форм для имён, 150 – для глаголов), параллельно производится поиск соответствий для всех новых введённых словоформ по текстам.

No	грамматические признаки	Новописьменный тверской (121)
Индикатив, презенс, положительные формы		
1.	1 л., ед. ч.	<u>elän</u>
2.	2 л., ед. ч.	<u>elät</u>
3.	3 л., ед. ч.	<u>eläy</u>
4.	1 л., мн. ч.	<u>elämmä</u>
5.	2 л., мн. ч.	<u>elättä</u>
6.	3 л., мн. ч.	<u>eletäh</u>
7.	ед. ч., коннегатив	<u>elä</u>
8.	мн. ч., коннегатив	<u>eletä</u>
Индикатив, презенс, отрицательные формы		
9.	1 л., ед. ч.	en elä
10.	2 л., ед. ч.	et elä
11.	3 л., ед. ч.	ei elä
12.	1 л., мн. ч.	emmä elä
13.	2 л., мн. ч.	että elä
14.	3 л., мн. ч.	ei eletä

Рис. 4. Фрагмент словоизменительной парадигмы глагола *eliä*. Полную информацию см. <http://dictorpus.krc.karelia.ru/ru/dict/lemma/23142>

Suaharuo himottau

Карельский: собственно карельское наречие
Толмачевский

Oi. Midä elettih? Ka täššä juablokka oli issutettu, pellot. Voinan jälgie ev midä šyvvä. Niin ka lumi šulau, Lidanke lähemmä,

jälgi NOUN (1) след,
отпечаток; метина) +
jälgi NOUN (2) остаток
чего-л., след) +
jälgi NOUN (3) колея) +

новое значение / лемма

partitive, ed. ч. +



hmä. Da muamo midä šöimmä. Žentäh nyt kaikki i
verih, tabakalla torguimah.
innettu dengua. I toi suaharuo, ka
razie, šuurie. Lageh... Ka tuala
niun kodi ev edähänä. Matičalla
nie en šöis' suaharoloida. A
Vahnemmat jo kuolluot oldih.
atalla šain šene suaharot. Da,
diedo oli elošša, tuaton tuatto, diedo. Diedo tuli. Storožana
ruado, melliččä, tädä kladovoloida storožuičči. Diedo tulou, a
mie iten. Suaharuo šain, a riputtua-to miula ei šua. Ka i issun i
iten. Hiän šano... Tuli da i šanou: "Midä itet?" Lau: "Muamo
tulou, milma rubieu pergamah." "Mintän?" "Ka, – lau, – šain
suaharot. Otin palazen, kivi vaš. Nu, ka moni palas'ta oli. – Lau,

Рис. 5. Проверка экспертом автоматической разметки и выбор правильного значения и грамматических признаков слова в тексте

Добавить словоформу

Matičalla oli nuagla.

В матице был гоэодь. (Suaharuo himottau)

исправить слово в тексте

Словоформа

Matičalla

Лемма

Возможно, это один из следующих вариантов?

- ☐ Matičča, имя собственное, адессив-аллатив, ед. ч. (51.28%)
- ☐ matičča, существительное, адессив-аллатив, ед. ч. (46.15%)
- ☐ matičča, прилагательное, адессив-аллатив, ед. ч. (2.56%)

Значение для предсказанной леммы

Рис. 6. Разметка «нового» слова, т. е. не найденной в словаре словоформы. В данном окошке предлагается выбрать лемму (начальную форму) из списка, создать новую или воспользоваться предсказанием системы

Редакторами корпуса производится проверка автоматической разметки, снятие омонимии, а также ручная разметка нераспознанных программой слов. На рис. 5 приведён пример автоматической разметки художественного текста. Редактор, кликая на иконку "+", выбирает верное значение и подходящие грамматические признаки.

При клике на неразмеченное слово открывается окно (рис. 6), в котором можно вручную выбрать или создать новую лемму (с полной парадигмой). Редактору помогает модуль-предсказатель, который предлагает грамматические признаки слова.

ВепКар – смешанный корпус. Кроме письменных текстов с 2022 г. он постепенно наполняется аудиозаписями звучащей речи (в настоящий момент в нём представлено уже более 100 фрагментов⁴).

Многофункциональность корпуса заключается в том, что он предназначен как для обычных пользователей, так и для исследовательских целей. Любой желающий может пользоваться ВепКаром как электронной библиотекой и как полноценным электронным словарём. В этом помогает удобная система поиска. Языковеды используют программы обработки, поиска и представления данных для ре-

⁴ См. http://dictorpus.krc.karelia.ru/ru/corpus/speech_corpus

шения своих конкретных научных задач в области изучения лексики и грамматики карельского и вепского языков. Так, система лексико-грамматического поиска позволяет выбирать из текстов примеры заданной последовательности лемм и/или словоформ, обладающих определёнными грамматическими характеристиками, что позволяет отфильтровывать сложные грамматические конструкции (рис. 7). Частотные словари⁵ выступают базой для статистических исследований, а обратные помогают разобраться со словоизменительными типами имён и глаголов.

Расширенный поиск ↓

Язык x карельский: ливвиковское наречие (489)	Подкорпус 	по 10 записей	ИСКАТЬ
--	--	--	---------------

Слово 1 olla	Часть речи VERB	Грамматические признаки COND	Расстояние от 1 до 1
Слово 2 	Часть речи VERB	Грамматические признаки ACT,V.PTCP_PST	+

Найдено 32 текста, 44 вхождения.

[Уточнить запрос](#)

- Igor Petrov. [Ristilahten tora](#) (Oma Mua. № 11; № 14; № 15; № 16; № 17, 2017, с. 10; 10; 10; 10; 10)

← Sobimuksen peittolizävykses oli sanottu, ku ruoččiloi soda-avus **olis** pidányh suaha Korelan ujezdu Korelan linnanke (nygöine Priozerskan linnu). →
- Juuli Kähäri. [Minun hiihtoloma](#) (Oma Mua. № 12, 2017, с. 10)

← Tama loma on ollut ylen vessel da toivozin, ku se ei vie **olis** loppunuh.
- Fodorova Anni (nygöi Ivanova). [Kargiet voinuaijat](#) (Oma mua. № 24, 2017, с. 8)

← Kai kyla **ollus** palanuh, ku vahnembat ei ehtittys. →
- Irina Kudeņnikova. [Bu'uu borkananke](#) (Oma mua. № 1, 2018, с. 11)

← Äijan rahvastu olluzin parandannuh, a työ kallehen syomizen kaimaitto!
- Ol'ga Ogneva. [Ongo Korzal kaimua?](#) (Oma mua. № 23, 2018, с. 6-7)

← Korzan kylan alustettih karjalazet, ga mis **olis** voinnuh lahtie kylan nimi? →

Рис. 7. С помощью лексико-грамматического поиска найдены примеры употребления ливвиковских глаголов в форме перфекта кондиционала

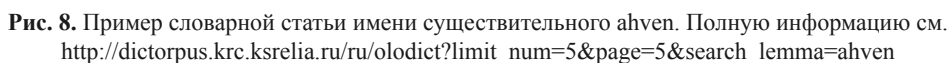
Кроме того, на основе данных корпуса разрабатываются приложения с упрощённым интерфейсом, более понятные широкой публике. К таким относятся Аудиокарта прибалтийско-финских языков Карелии⁶, Приложение для проверки орфографии⁷. Продолжается работа над Мультимедийным словарём карельского языка LiPaS⁸, содержащим материалы на ливвиковском новописьменном варианте. Целевой аудиторией этого проекта являются, в первую очередь, школьники, студенты, слушатели курсов карельского языка, преподаватели.

⁵ Частотный словарь лемм, размеченных в тексте (<http://dictorpus.krc.karelia.ru/ru/corpus/text/frequency/lemmas>); обратный словарь лемм (http://dictorpus.krc.karelia.ru/ru/dict/reverse_lemma).

⁶ См. <http://dictorpus.krc.karelia.ru/ru/corpus/audiotext/map>

⁷ См. <http://dictorpus.krc.karelia.ru/ru/corpus/spellchecking>

⁸ См. <http://dictorpus.krc.karelia.ru/ru/olodict>



На ближайшие годы работа по развитию корпуса ВепКар запланирована в нескольких основных направлениях.

194

для разработки и настройки различных автоматизированных систем (машинный перевод, распознавание речи и пр.). Карельский и вепский языки – языки малоресурсные, поэтому проблема дефицита языкового материала для решения таких масштабных задач является особо актуальной. Кроме того, предполагается также добавление новой метаразметки в целях привлечения к корпусу специалистов литературоведов, фольклористов, этнографов.

2. Наполнение словаря корпуса как за счёт разметки текстов, так и за счёт внесения данных из печатных словарей, в том числе диалектных. На основе словаря корпуса планируется продолжение работы по созданию удобных многофункциональных мультимедийных учебных и популярных словарей, в первую очередь, для новописьменных вариантов карельского и вепского языков, звучащих словников диалектной речи.

3. Завершение работ по созданию «золотого стандарта» или размеченной вручную части корпуса, которая в дальнейшем будет использована в различных экспериментах, направленных на разработку программы для автоматического снятия грамматической омонимии. Важно отметить, что работа по разметке корпуса продвигается в разы медленнее, чем загрузка текстов с их привязкой к метаданным, ввиду трудозатратности процесса и небольшого числа квалифицированных редакторов.

4. Если до настоящего момента упор был сделан на наполнение корпуса материалами и развитие его инструментальной базы, то теперь, когда набран минимально необходимый объём данных, перед коллективом поставлена задача активизации исследовательской деятельности по самым разным направлениям: диалектология, языковое нормирование, лексикография и пр. Представляется, что платформа ВепКар в обозримом будущем могла бы стать источником для междисциплинарных исследований в рамках филологических и исторических наук.

5. Развитие корпуса производится и за счёт расширения направлений взаимодействия с пользователями. ВепКар пока ещё остаётся востребованным ресурсом лишь для ограниченного круга лиц – финно-угроведов, студентов и преподавателей. Между тем, как показал проведённый в 2023 г. Марафон записи образцов карельской и вепской диалектной речи «Слушаю родной говор», формат обратной связи с потенциальными пользователями платформы положительно воспринимается населением и специалистами. Коллектив ВепКара планирует проводить мастер-классы по корпусной грамотности для заинтересованных лиц, разрабатывать игры и квесты на материалах корпуса, устраивать акции по сбору звучащей речи (запись приветов с малой родины, сбор топонимических преданий и пр.). Языковеды и фольклористы ИЯЛИ КарНЦ РАН крайне заинтересованы в проведении подобных мероприятий, поскольку возможности посетить все населённые пункты, в которых ещё можно записать стремительно угасающую вепскую и карельскую речь, нет.

Успешное достижение поставленных целей опирается в одну основную проблему – проблему развития кадрового потенциала. Костяк сотрудников, регулярно задействованных в работе над корпусом, сегодня составляет восемь

человек, из них пять языковедов, занимающихся непосредственно работой в корпусе (разметка, перевод текстов, наполнение словарей, проведение экспериментов, крайне необходимых для отлаживания работы лингвистических инструментов), два математика-программиста, обеспечивающих работу системы, и один инженер, пополняющий базу текстами.

Кадровую проблему удаётся частично решать благодаря:

- грантовой поддержке. Например, разработка и наполнение Речевого подкорпуса были осуществлены за счёт средств регионального гранта РНФ;

- привлечению коллег неязыковедческих направлений. Так, разработкой подкорпуса карельских рун занимался фольклорист ИЯЛИ КарНЦ РАН;

- привлечению студентов Петрозаводского государственного университета.

В ходе ежегодной диалектологической практики студенты-филологи занимаются подготовкой к загрузке в корпус образцов диалектной речи и словарного материала. А студентами-математиками на базе ВепКара разрабатываются удобные мобильные приложения.

Литература

Бойко Т.П., Зайцева Н.Г., Крижановская Н.Б., Крижановский А.А., Новак И.П., Пеллинен Н.А., Родионова А.П., Трубина Е.Д. Лингвистический корпус ВепКар – «заповедник» прибалтийско-финских языков Карелии // Труды Карельского научного центра Российской академии наук. 2021. № 7. С. 100-115.

Зайцева Н.Г., Крижановская Н.Б. Корпусная лингвистика в прибалтийско-финском исследовательском пространстве (на материале Корпуса вепсского языка и Открытого корпуса вепсского и карельского языков) // Альманах североευропейских и балтийских исследований. Вып. 3. 2018. С. 264-273. [Электронный ресурс]. URL: <https://nbsr.petsu.ru/journal/article.php?id=1062> (дата обращения: 04.03.2024)

Зайцева Н.Г., Харитонова Е.Е., Жукова О.Ю. Орфографический словарь вепсского языка. Петрозаводск: КарНЦ РАН, 2012. 432 с.

Крижановский А.А., Крижановская Н.Б., Новак И.П. Представление диалектов в Открытом корпусе вепсского и карельского языков (ВепКар) // Труды международной конференции «Корпусная лингвистика – 2019». СПб, 2019. С. 288-295.

Новак И.П., Крижановская Н.Б., Бойко Т.П., Пеллинен Н.А. Разработка правил генерации именных словоформ для новописьменных вариантов карельского языка // Вестник угроведения. 2020. Т. 10. № 4. С. 679-691.

Родионова А.П., Крижановская Н.Б., Пеллинен Н.А. Речевой корпус ВепКар как инструмент сохранения диалектной речи прибалтийско-финских народов Карелии // Ежегодник финно-угорских исследований. 2023. № 3. С. 343-351.