

Людиковские материалы в Открытом корпусе вепсского и карельского языков

В настоящее время активно развивается направление лингвистической науки – корпусной лингвистики, которое основано на использовании языковых электронных корпусов и корпусных технологий. В мире уже существуют огромные электронные ресурсы, которые воплощены, например, в Национальный корпус русского языка¹, Британский национальный корпус², Чешский национальный корпус³. Известен Венгерский национальный корпус⁴, Языковой банк Финляндии⁵ и т.д. В настоящей статье мы остановимся на описании «Открытого корпуса вепсского и карельского языков».

Проект «Открытый корпус вепсского и карельского языков» (ВепКар)⁶ ведет свою историю с 2009 г., когда в Институте языка, литературы и истории Карельского научного центра РАН (ИЯЛИ КарНЦ РАН)⁷ под руководством д. ф. н. Н. Г. Зайцевой началась работа по созданию «Корпуса вепсского языка»⁸. Корпус вепсского языка, включающий в себя электронный словарь и 5 текстовых корпусов: диалектные тексты; фольклорные тексты с двумя подкорпусами (причитания и сказки); два младописьменных корпуса (переводы Нового Завета; тексты на младописьменном вепском языке), содержит собственные системы поиска по различного рода характеристикам: по отдельным словам, диалектам, жанрам фольклора и жанрам младописьменных текстов и т.д. [1, С. 16]. «Корпус вепсского языка» оказал большую помощь в подготовке «Орфографического словаря вепсского языка» [2] его создателям (Н.Г. Зайцевой, О.Ю. Жуковой и Е.Е. Харитоновой), поскольку смог обеспечить возможность поиска в диалектном материале при отборе форм для создания правил орфографии вепсского языка [3, С. 20].

Представителями вепсского и карельского народов решаются идентичные проблемы по возрождению языков и культур, и всё это требует значительных языковых ресурсов, которые позволят исследователям создавать и развивать, например, правила карельской орфографии, введут в научный и общественный оборот значительное количество материалов, востребованных авторами учебников по языку и истории родного края,

¹ <http://www.ruscorpora.ru/new/>

² <http://www.natcorp.ox.ac.uk/>

³ <https://www.korpus.cz/>

⁴ http://mnsz.nytud.hu/index_hun.html

⁵ <https://sanat.csc.fi/>

⁶ dictorpus.krc.karelia.ru

⁷ <http://illhportal.krc.karelia.ru/>

⁸ <http://vepsian.krc.karelia.ru/about/>

учителями, писателями. В связи с этим, в целях сохранения, развития и популяризации не только вепского, но и карельского языка в 2016 году сотрудники ИЯЛИ и Института прикладных математических исследований КарНЦ РАН приступили к созданию многоязычного корпуса. И, таким образом, в 2016 г., корпус стал многоязычным: на базе вепского корпуса был создан карельский корпус, включающий в себя деление на три подкорпуса в соответствии с тремя основными наречиями (собственно карельским (с-к), ливвиковским (ливв.) и людиковским (люд.)) [4, С. 82]. Иными словами, объединённый корпус стал продолжением «Вепского корпуса» и получил название «Открытый корпус вепского и карельского языков» (ВепКар).

Открытый корпус состоит из двух частей: текстов и электронного словаря. На данный момент в корпусе насчитывается около 60 тысяч словарных статей и около 3 тысяч текстов на 6 языках и 46 диалектах (прим. данные на октябрь 2020). На сей день в корпусе преобладают тексты и леммы в вепском подкорпусе (вепс.). При этом количество текстов и слов в карельском подкорпусе ежедневно увеличивается.

На настоящий момент карельский подкорпус представлен 25 собственно карельскими, 8 ливвиковскими и 4 людиковскими диалектами из традиционной классификации П.М. Зайкова [5, С. 27]. Кроме этого в корпус включены четыре младописьменные формы: севернокарельская, тверская, ливвиковская и людиковская. Следует также подчеркнуть, что в корпусе ВепКар три карельских наречия представлены отдельно, а не единым языком, связано это с тем, что развившиеся в наречиях существенные различия диктуют необходимость использования в нём трех разных словарей карельского языка [6, С. 289].

Карелы-людики – одно из подразделений карельского этноса, традиционно проживающих в ряде деревень и поселков юго-восточной части Республики Карелия в Олонецком, Пряжинском и Кондопожском районах. Территория, где проживают карелы-людики, тянется примерно на 200 км., что сказывается на различиях между людиковскими диалектами, поскольку на западе она граничит с ливвиковской, где говорят на ливвиковском наречии карельского языка, на севере с собственно карельской, на востоке и юге – с русской.

Всего на настоящий момент в корпусе 156 людиковских текстов. Все тексты корпуса даны с подробной паспортизацией, где указаны диалект, данные информанта (имя, год рождения, место рождения), год и место записи, данные исследователя, записавшего текст, что очень нужно, прежде всего, для работы диалектологов.

Для людиковского подкорпуса, главным образом, используются тексты из опубликованных образцов речи [7; 8], подготовленных А. П. Баранцевым. Все тексты (а их около 200 в опубликованных образцах) представляют собой расшифровки магнитофонных

записей и передают естественную речь информанта со всеми ее индивидуальными и общезыковыми особенностями.

Кроме текстов в корпусе имеется и электронный словарь. В людиковском подкорпусе в настоящий момент около 6000 лемм. Изначально предполагалось пополнять Словарь корпуса исключительно теми леммами, которые используются в текстах корпуса. Впоследствии было принято решение включать в корпус материалы уже изданных словарей. Поэтому количество слов, например, на ливвиковском наречии значительно превосходит людиковский словник. Прежде всего, потому, что на ливвиковском, например, уже издано два больших академических словаря [9; 10], которые включает в себя, в том числе, и новейшую лексику. К большому сожалению, на людиковском наречии подобных словарей, русско-карельских и карельско-русских, в России до сих пор нет.

Многим исследователям знаком «Сопоставительно-ономасиологический словарь диалектов карельского вепсского, саамского языков», которые находятся между собой в генетическом родстве.

Людиковское наречие карельского языка представляют четыре говора следующих населенных пунктов: д. Койкары и с. Галлезеро Кондопожского района, с. Михайловское Олонецкого района и с. Святозеро Пряжинского района. Сбор материал для словаря осуществлялся в полевых условиях в 1979-1981 гг. знатоками людиковского наречия: А.П. Баранцевым (с. Галлезеро, с. Михайловское) и А.П. Созоновой (д. Койкары, с. Святозеро) [11]. Уникальность словаря заключается, прежде всего, в том, что он отражает состояние диалектов того времени, учитывая, что в настоящий момент из года в год записывать информантов становится сложнее.

Словарь составлялся по понятийно-тематическому принципу: каждый раздел начинается с темы, внутри каждой темы представлены слова-понятия.

Представим вашему вниманию слова-понятия в разделе «Вселенная», подразделе А.1.1. «Небо, небесные тела» [11, С. 14-15]:

Ключевое слово (по-русски)	Койкары	Галлезеро	Святозеро	Михайловское
земля, суша	mua	mua	mua	mua
небо	taivas	taivas	taivaz	taivaz
солнце	päiväne	päiväine, päiv, päivüöi	päiväine, päivyöin'e, päive	päivään'e
взойти (о небесных)	nosta, noussa	nosta	nosta	nousta

светилах)				
опуститься, закатиться (о небесных светилах)	laskeuduo	laskezeta	laskedakseh	lasktakse
луна	kuu, kuudamone, kuudain	kuldoi	kuu, kuudamaine, kuudain, kuudaine, kudei	kudmaan'e
звезда	tähti	tiäht'	tiähti	tiäht'

Уже первые семь понятий указывают на некоторые различия, прежде всего фонетические, между диалектами, связанные с влиянием других наречий (или языков) на них. Особенно эти различия преобладают в глаголах.

Обращаясь к отдельным понятиям словаря можно проследить разницу, прежде всего, в конечной огласовке слов. Согласно закону отпадения гласных в двусложных словах, где первый слог был исторически долгим, или закрытым, в номинативе в ряде людиковских говоров первичные конечные гласные *a*, *ä* перешли в *e*, или, как и в вепсском языке, отпали [12, С. 755-756]. С другой стороны, в диалекте Койкары конечная гласная *a* сохраняется, что говорит о сильном влиянии собственно карельского наречия на него.

Ключевое слово (по-русски)	Койкары	Галлезеро	Святозеро	Михайловское
амбар	aitta (ср. с-к. aitta)	aitt	aitte	ait (ср. вепс. ait)
ветвь	oksa (ср. с-к. oksa, okša) varba (ср. с- к. varpa)	oks, barb	okse, barbe	oks (ср. вепс. oks) barb (ср. вепс. barb)
погода сырая с дождем и мокрым снегом, слякоть	röččä (ср. с-к. röččä) rändä (ср. с-к. räntä)	röčč, ränd (ср. вепс. ränd) šl'ott	röčče, rände	röčč
лес	meččä (ср. с-к. meččä)	mečč	mečče	mecc (ср. вепс. mecc, mec)

Влияние собственно карельского и ливвиковского наречия, с одной стороны, и вепсского языка, с другой стороны, можно проследить и в оформлении показателей 1.инфинитива у людиковских глаголов.

Ключевое слово (по-русски)	Койкары	Галлезеро	Святозеро	Михайловское
лаять	haukkuo (ср. с-к., ливв. haukkuo)	haukkuda	haukkuda	haukta (ср. вепс. houkta)
сметь	ruohtie (ср. с-к., ливв. ruohtie)	ruoht'ida	ruoht'ida	ruohtta (ср. вепс. rohtta)
хвалить	kiittää (ср. с-к. kiittää) kiittiä (ср. ливв. kiittiä)	kiittada	kiittädä	kitta (ср. вепс. kitta)
смотреть	kaččuo (ср. с-к., ливв. kaččuo)	kaččoda	kaččoda	kacta (ср. вепс. kacta)

К настоящему моменту весь материал Словаря включен в ВепКар. В процессе кодирования Словаря была выбрана система, позволяющая отразить не только особенности лексического состава исследуемого языка, но и фонетические особенности каждого диалекта.

База данных позволяет произвести подборку лексики для определенного диалекта по определенной теме, тем самым формируя, небольшие тематические словари.

Таким образом, в дальнейшем к понятию на русском языке будут привязаны лексические соответствия с учетом фонетических особенностей.

У людиковского наречия письменная норма до конца не сформирована: в данном случае выбор диалектной базы для создания единой нормы затруднен в связи с наличием грамматических и лексических различий в говорах. Поэтому первые учебные пособия, созданные на основе михайловского диалекта, не всегда принимаются носителями, например, севернолюдиковских говоров, и наоборот. Парадоксально, т. н. нормированный литературный язык стал главным «врагом» диалектов из-за стремления унифицировать, собрать воедино многочисленные языковые особенности. После того, как в ВепКаре будут представлены не только святозерский (южнолюдиковский) и михайловский диалекты, но и средне- и севернолюдиковские (диалекты д. Галлезеро и д. Койкары), на основе собранного и проанализированного материала появится реальная возможность создать единую общелюдиковскую норму.

Литература

1. Зайцева Н.Г. Вепские причитания в фокусе корпусной лингвистики и лингвофольклористики // Материалы ХLI Международной филологической конференции. 26-31 марта 2012 г. Секция "Уралистика". СПб.: Филологический факультет СПбГУ. 2012. С. 16-26.
2. Зайцева Н.Г., Харитоновна Е.Е., Жукова О.Ю. Орфографический словарь вепского языка. Петрозаводск: Карельский научный центр, 2012. 432 с.
3. Крижановский А. А., Крижановская Н. Б., Родионова А. П. Архитектура корпусного менеджера и разметка текстов корпуса ВепКар // Электронная письменность народов Российской Федерации : опыт, проблемы и перспективы: Материалы II Международной научной конференции (Уфа, 11–12 декабря 2019 г.). – Уфа: Башк. энцикл. 2019. С. 19-23.
4. Родионова А.П. Людиковский глагол и его диалектные источники // Ученые записки Петрозаводского государственного университета. № 1 (178). 2019. С. 80–84.
5. Зайков П.М. Глагол в карельском языке / П.М. Зайков. Петрозаводск: Изд-во ПетрГУ, 2000. 293 с.
6. Крижановский А.А., Крижановская Н.Б., Новак И.П. Представление диалектов в Открытом корпусе вепского и карельского языков (ВепКар) //Труды международной конференции «Корпусная лингвистика – 2019». СПб, 2019. С. 288-295.
7. Баранцев А.П. Образцы людиковской речи (образцы корпуса людиковского идиолекта) / А.П. Баранцев. Петрозаводск: «Карелия», 1978. 287 с.
8. Näytteitä karjalan kielestä I. Joensuu-Петрозаводск, 1994. 455 с.
9. Бойко Т. П., Маркианова Л. Ф. Большой русско-карельский словарь: (ливвиковское наречие). Петрозаводск: Periodika, 2018. 399 с
10. Бойко Т. П. Большой карельско-русский словарь(ливвиковское наречие). Петрозаводск: Periodika, 2016. - 351 с.
11. Сопоставительно-ономасиологический словарь диалектов карельского, вепского, саамского языков / Под общей редакцией Ю.С. Елисеева и Н.Г. Зайцевой. Петрозаводск, 2007.346 с.
12. Turunen A. Lyydiläismurteiden äännehistoria. I. Konsonantit // Suomalais-ugrilaisen Seuran Toimituksia. Helsinki, 1946. № 89. 338 с.