**УДК 811.511.112’366**
**DOI: 10.30624/2220-4156-2020-10-4-679-691**

# Разработка правил генерации именных словоформ для новописьменных вариантов карельского языка

**И. П. Новак**
*Институт языка, литературы и истории*
*Карельского научного центра Российской академии наук,*
*г. Петрозаводск, Российская Федерация,*
*novak@krc.karelia.ru*

**Н. Б. Крижановская**
*Институт прикладных математических исследований*
*Карельского научного центра Российской академии наук,*
*г. Петрозаводск, Российская Федерация,*
*nataly@krc.karelia.ru*

**Т. П. Бойко**
*Институт языка, литературы и истории*
*Карельского научного центра Российской академии наук,*
*г. Петрозаводск, Российская Федерация,*
*boiko@krc.karelia.ru*

**Н. А. Пеллинен**
*Институт языка, литературы и истории*
*Карельского научного центра Российской академии наук,*
*г. Петрозаводск, Российская Федерация,*
*nataliapellinen@gmail.com*

**АННОТАЦИЯ**

**Введение**. Связывание слов текста (токенов) со значениями лемм в словаре корпуса ВепКар существенно облегчает дальнейшую работу по семантической разметке текстов. Для вепсского подкорпуса ВепКар в 2019 г. были разработаны флективные правила, на их основе в корпус была добавлена функция генерации полной парадигмы по базовым словоформам.

При пополнении словарными статьями трёх подкорпусов карельского языка редакторам необходимо вводить большое число словоформ (около 30 для имён и 150 для глаголов), поэтому разработка алгоритма и компьютерной программы генерации словоформ карельского языка оказалась своевременной.

**Цель**: проиллюстрировать, как с помощью списка основ именных частей речи двух новописьменных наречий карельского языка можно составить правила для автоматической генерации словоформ.

**Материалы исследования**: леммы и словоформы из Открытого корпуса вепсского и карельского языков, Корпуса Приграничной Карелии, электронной версии Словаря карельского языка.

**Результаты и научная новизна**. На основе изученных по теоретическим источникам и выявленных в ходе многолетних наблюдений грамматических закономерностей, а также проведённых в исследовании экспериментов сформирован список основ и псевдооснов именного словоизменения, разработана система правил генерации словоформ, написана и проверена соответствующая программа. Научная новизна исследования заключается во впервые предпринимаемой разработке системы единых правил автоматической генерации словоформ для двух наречий карельского языка.

***Ключевые слова***: карельский язык, новописьменный язык, корпусная лингвистика, морфология, именное словоизменение, генерация словоформ.

*Благодарности*: Исследование проведено в рамках выполнения государственного задания КарНЦ РАН. Раздел «Разработка программы генерации» подготовлен Н.Б. Крижановской в рамках проекта РФФИ 18-012-00117.

*Для цитирования*: 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.

# Development of rules of generation of nominal word forms for new-written variants of the Karelian language

**I. P. Novak**
*Institute of Linguistics, Literature and History,*
*Karelian Research Centre of the Russian Academy of Sciences,*
*Petrozavodsk, Russian Federation,*
*novak@krc.karelia.ru*

**N. B. Krizhanovskaya**
*Institute of Applied Mathematical Research,*
*Karelian Research Centre of the Russian Academy of Sciences,*
*Petrozavodsk, Russian Federation,*
*nataly@krc.karelia.ru*

**T. P. Boyko**
*Institute of Linguistics, Literature and History,*
*Karelian Research Centre of the Russian Academy of Sciences,*
*Petrozavodsk, Russian Federation,*
*boiko@krc.karelia.ru*

**N. A. Pellinen**
*Institute of Linguistics, Literature and History,*
*Karelian Research Centre of the Russian Academy of Sciences,*
*Petrozavodsk, Russian Federation,*
*nataliapellinen@gmail.com*

**ABSTRACT**

**Introduction:** linking of words of texts (tokens) with meanings of lemmas in the dictionary of VepKar corpus significantly facilitates further work on semantic markup of texts. In 2019, inflectional rules were developed for the Vepsian subcorpora VepKar. To the corpus on the base of these rules a function for generation of a complete paradigm on basic word forms was added.

VepKar editors need to enter a large number of word forms when they create dictionary entries in three Karelian subcorpora (about 30 for names and 150 for verbs). Therefore, the development of an algorithm and a computer program for generation of word forms of the Karelian language turned out to be timely.

**Objective:** to illustrate how you can use the list of the stems of the nominal parts of speech of two new-written dialects of the Karelian language to create rules for automatic generation of word forms.

**Research materials:** lemmas and word forms from the Open corpus of the Vepsian and Karelian languages, the Corpus of Border Karelia, and the electronic version of the Dictionary of the Karelian language.

**Results and novelty of the research:** grammatical patterns were studied over many years from theoretical sources, and they were also discovered through experiments. Thanks to this, the list of stems and pseudo-stems of word forms was formed for the nominal parts of speech, the system of rules for generation of word forms was developed, and the corresponding computer program is written and tested. The scientific novelty of the study lies in the first attempt to develop uniform rules for the automatic generation of word forms for two dialects of the Karelian language.

***Key words*:** Karelian language, new-written language, corpus linguistics, morphology, nominal inflection, generation of word forms.

*Acknowledgements*: the study is carried out under the state order of the Karelian Research Centre of the Russian Academy of Sciences. The section «Development of the program of generation» was written by N. B. Krizhanovskaya in the framework of the project of the Russian Foundation for Basic Research No. 18-012-00117.

*For citation*: Novak I. P., Krizhanovskaya N. B., Boiko T. P., Pellinen N. A. Development of rules of generation of nominal word forms for new-written variants of the Karelian language // Vestnik ugrovedenia = Bulletin of Ugric Studies. 2020; 10 (4): 679–691.

## Введение

Открытый корпус вепсского и карельского языков (ВепКар[1]) является продолжением Корпуса вепсского языка (рук. Н. Г. Зайцева), созданного в 2009 г. [16, 391]. Карельский язык был включён в него в 2016 г. Корпус представляет собой электронную информационно-справочную систему, содержащую публицистические, художественные и диалектные тексты, разбитые на предложения и слова-токены. Часть текстов снабжена переводом на русский язык, выравненным с оригиналом по предложениям.

Важнейшей составляющей системы является словарь. Связывание слов текста со значениями

лемм в словаре существенно облегчает дальнейшую работу по их семантической разметке. Однако процесс по заполнению словаря продвигается очень медленно, что объясняется нехваткой специалистов и отсутствием необходимых для этого программ. Например, заполнение вручную именной словоизменительной парадигмы одной леммы (31 словоформа) занимает в среднем около 4–5 минут, а глагольной (149 словоформ) – 15–20 минут.

По карельскому языку вся работа в корпусе автоматически увеличивается в три раза, поскольку в нём представлено три карельских подкорпуса (рис. 1). Такое деление осуществлено в соответствии с тремя наречиями: собственно карельским, ливвиковским и людиковским, обнаруживающими между собой существенные различия на всех языковых уровнях [2, 21–27; 11, 57–58].

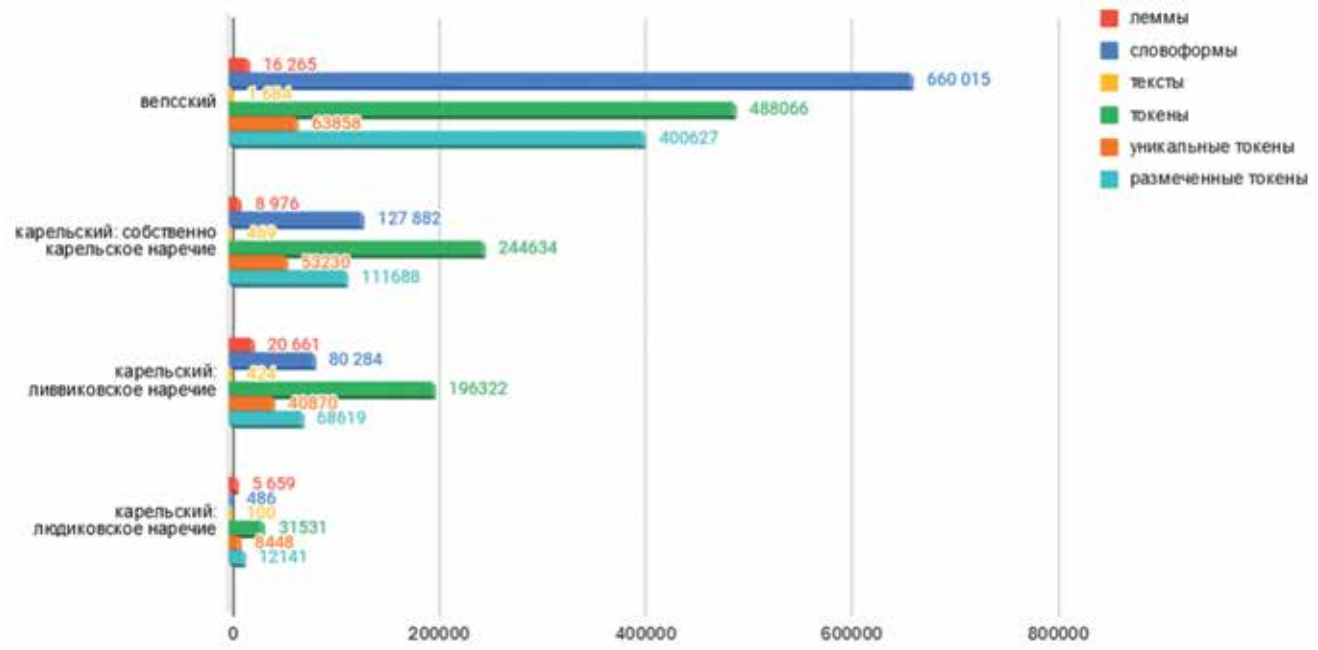


**Рис. 1.** Соотношение объёмов словарей и подкорпусов ВепКар (данные на 16.05.2020).

Для вепсского языка в 2019 г. были разработаны флективные правила и в корпус добавлена функция генерации полной парадигмы по базовым словоформам, отсюда большая доля не только вепсских словоформ, но и размеченных токенов (85%). Ускорить работу по разметке текстов и пополнению словарей карельских подкорпусов также можно при помощи программы автоматической генерации словоформ. В настоящем исследовании предполагается проанализировать возможности создания такой программы на примере именной словоизменительной системы новописьменных вариантов карельского языка Республики Карелия, разработанных на базе ливвиковского и собственно карельского наречий [4, 75–77].

Для новописьменных вариантов карельского языка определены орфографические нормы и грамматические правила, закреплённые в грамматиках П. М. Зайкова «Karielan kielioppi» [17], «Vienankarjalan kielioppi» [18] (собственно карельское наречие), Л. Ф. Маркиановой «Karjalan kielioppi» [12] и Р. Пюёли «Livvin karjalan kielioppi» [14] (ливвиковское наречие). В них приведены основные словоизменительные типы и парадигмы. Однако в процессе работы был выявлен ряд моментов, оставленных в грамматиках без внимания, что существенно осложнило разработку системы правил генерации и потребовало проведения дополнительных исследований.

В качестве теоретических источников были привлечены также справочник «Карельский язык в грамматиках» [2] и Академическая грамматика близкородственного карельскому финского языка [8].

В ходе исследования были использованы материалы Языкового банка Финляндии [7], в том числе Корпуса Приграничной Карелии (ок. 850 тыс. токенов) [13], основанного на записанных в 1960-х гг. образцах речи карелов Финляндии [10]. В качестве языковой базы для анализа карельской именной словоизменительной систем

[1] См. http://dictorpus.krc.karelia.ru

привлечены также данные электронной версии словаря карельского языка (88 тыс. словарных статей на диалектах собственно карельского и ливвиковского наречий) [9].

В процессе работы над программой был использован опыт создания электронных параллельных словарей уральских языков (например, людиковско-русско-финский словарь) [15], а также морфологического анализатора для родственного финскому квенского языка [5]. Кроме того, планируется использовать разработки программ морфологического анализа и морфологической генерации, в том числе для малоресурсных языков, разработанных и обученных на основе тщательно размеченных текстов Библии на английском языке и её переводов на сотни других языков [6]. Наличие переводов Библии на вепсский и карельский языки позволяет надеяться на возможность создания для них морфологических инструментов.

У коллектива редакторов корпуса ВепКар также имеется опыт работы над программой генерации словоформ. В 2019 г. такая программа была разработана для нормированного варианта карельского языка, развивающегося в Тверской области) [1]. Её применение позволило пополнить словарь более чем на 200 тыс. словоформ и увеличить разметку корпуса на 45%. В ходе работы предполагается переработать и распространить программу на собственно карельский и ливвиковский новописьменные варианты.

## Материалы и методы

Исследование проведено на базе материалов словаря корпуса ВепКар на собственно карельском (около 9 тыс. лемм и 129 тыс. словоформ) и ливвиковском (около 21 тыс. лемм и 557 тыс. словоформ) новописьменных языках. В 2018 году в ВепКар вручную было добавлено 18 тыс. лемм с основными словоформами (ном., ген., парт. ед. и парт. мн. для имен) из «Большого карельско-русского словаря (ливвиковское наречие)» [1]. Наличие этих материалов позволило провести эксперименты по установлению закономерностей образования основы множественного числа[1].

В качестве базовых, наряду с экспериментами, выступили метод словообразовательного анализа, а также сравнительный и статистический методы, применение которых в совокупности позволило определить алгоритм разработки системы правил автоматической генерации словоформ: 1) формирование списка грамматических форм; 2) выделение словоизменительных типов; 3) определение количества необходимых для генерации основ; 4) разработка системы правил генерации; 5) написание программы генерации и её проверка. В рамках статьи предлагается рассмотреть все этапы этого процесса на примере именной словоизменительной системы.

## Результаты

### 1. Формирование парадигмы именного словоизменения

Карельский язык, как все прибалтийско-финские языки, является агглютинативным. Словоизменение в нём происходит путём присоединения аффиксов, содержащих грамматическое значение, к лексической основе слова [2, 163; 8, 53]. Последовательность аффиксов строго регламентирована: лексическая основа – показатель множественного числа – падежное окончание [14, 22; 18, 59]. Напр., *pien-i-h* от *pieni* ‘маленький’, где *piene-* – основа, *-i-* – суффикс множественного числа, *-h* – падежное окончание иллатива.

Склонение имён карельского языка происходит по падежам и числам. Категории притяжательности и сравнения в прибалтийско-финском языкознании рассматриваются как пограничные случаи словоизменения и словообразования [8, 62, 95, 296], в связи с чем они были изъяты из анализа.

Для карельского языка характерна бинарная структура категории числа. Форма единственного числа не маркирована (*hyvä-h tyttö-h* илл. ед. от *hyvä tyttö* ‘хорошая девочка’). Множественное число номинатива образуется при помощи окончания ***-t***. Для образования форм множественного числа косвенных падежей употребляются показатели ***-i-***, ***-loi-*** / ***-löi-***, ***-zi-*** (*hyv**i**-h tyttö-**löi**-h* илл. мн.) [12, 34–36; 14, 26–29; 17, 48–49; 18, 71–72].

В диалектах карельского языка выделяют от 9 до 16 продуктивных падежей, поскольку система находится в процессе непрерывного становления. Для собственно карельского новописьменного варианта определено 14 продуктивных падежей [17, 55; 18, 75], а для ливвиковского – 16 [12, 37–38] / 17 [14, 30]. Большая часть падежей имеет единое окончание как в единственном, так и во множественном числе. Исключением является форма номинатива, особое внимание следует уделить также образованию форм партитива и генитива.

[1] См. http://dictorpus.krc.karelia.ru/ru/experiments/vowel_gradation

*Таблица 1*

**Парадигма именного словоизменения новописьменных вариантов карельского языка[1]**

| Падеж | Одноосновное имя | | Двуосновное имя | |
|---|---|---|---|---|
| | Ед. ч. | Мн. ч. | Ед. ч. | Мн. ч. |
| Ном. | **tyttö** ʻдевочкаʼ | tytö-t | *lammaš/s* ʻовцаʼ | **lamp/baha**-t |
| Ген. | tytö-n | **tyttö**-j/löi-en/n | **lamp/baha**-n | **lamp/bahi**-en |
| Парт. | **tytty**-ö | **tyttö**-jä/löi | *lammaš/s*-ta/tu | **lamp/bahi**-e/i |
| Эсс. | **tyttö**-nä / tytö-nny | **tyttö**-löi-nä/nny | **lamp/baha**-na/nnu | **lamp/bahi**-na/nnu |
| Транс. | tytö-ksi/kse | **tyttö**-löi-ksi/kse | **lamp/baha**-kši/kse | **lamp/bahi**-ksi/kse |
| Инесс. | tytö-ššä/s | **tyttö**-löi-ssä/s | **lamp/baha**-šša/s | **lamp/bahi**-ssa/s |
| Элат. | tytö-štä/s(späi) | **tyttö**-löi-stä/s(späi) | **lamp/baha**-šta/s(späi) | **lamp/bahi**-sta/s(späi) |
| Илл. | **tyttö**-h | **tyttö**-löi-h | **lamp/baha**-h | **lamp/bahi**-h |
| Адесс. | tytö-llä/l | **tyttö**-löi-llä/l | **lamp/baha**-lla/l | **lamp/bahi**-lla/l |
| Абл. | tytö-ltä/l(päi) | **tyttö**-löi-ltä/l(päi) | **lamp/baha**-lta/l(päi) | **lamp/bahi**-lta/l(päi) |
| Алл. | - / tytö-le | - / **tyttö**-löi-le | **lamp/baha**-lla/le | **lamp/bahi**-lla/le |
| Абесс. | tytö-ttä/ttäh | **tyttö**-löi-ttä/ttäh | **lamp/baha**-tta/ttah | **lamp/bahi**-tta/ttah |
| Ком. | - / tytö-nke | **tyttö**-löi-neh/nneh(nke) | - / **lambaha**-nke | **lamp/bahi**-eh/nneh(nke) |
| Прол. | - / tytö-či | **tyttö**-löi-neh | - / **lambaha**-či | - / **lambahi**-či |
| Инстр. | - | **tyttö**-löi-n | - | **lamp/bahi**-n |

Представленные в таблице 1 словоизменительные парадигмы демонстрируют, что карельский язык далёк от агглютинативного идеала, поскольку в нём развились такие флективные способы словоизменения, как чередования конечных гласных основы (*tyttö-h* илл. ед. – *tytty-ö* парт. ед., *lamp*/*baha-h* илл. ед. – *lamp*/*bahi-h* илл. мн.) и консонантная альтернация, на основании которой производится разделение основ на слабые и сильные (*tytö-t* ном. мн. – *tyttö-h* илл. ед.). Для языка характерно наличие как одноосновных, обладающих гласной основой, и двуосновных, обладающих гласной и согласной основами, имён (*tytö-n* ном. ед. – *tytty-ö* парт. ед., *lamp*/*baha-n* ген. ед. – *lammaš*/*s-ta*/*tu* парт. ед.). Следовательно, для выявления словоизменительных типов имён необходимо наличие информации о следующих словоформах: **словарная**, формы **генитива** (сл. гл. осн. одноосн. имён и сильн. гл. осн. двуосн.) и **партитива** (сильн. гл. осн. с чередованием гласных одноосн. имён, согл. осн. двуосн.) **единственного числа**, форма **генитива множественного числа** (осн. мн. ч.).

## 2. Определение количества типов словоизменения имён

Словоизменительные типы имён карельского языка восходят к прибалтийско-финскому языку-основе, поэтому практически не имеют отличий на уровне наречий. Основной критерий, позволяющий отнести имя к тому или иному словоизменительному типу, – наличие или отсутствие у него согласной основы. Количество именных словоизменительных типов в нормативных грамматиках варьируется от 6 до 18 [12, 31–33; 14, 22–25; 17, 51–55; 18, 60–69]. В таблице 2 приведены примеры основных возможных типов имён, выявленных в материалах словарей корпуса ВепКар.

*Таблица 2*

**Словоизменительные типы имён новописьменных вариантов карельского языка[2]**

| **С.ф.** на | **Ном. ед.** | **Ген. ед.** | **Парт. ед.** | **Ген. мн.** |
|---|---|---|---|---|
| **Одноосновные имена** | | | | |
| *-a*, *-ä* / + *-u*, *-y* | ranta / rand**u** ʻберегʼ | ra**nn**a-n | rant/*d***u**-a | rant**o**-j-en / ra**nno**-in |
| *-a*, *-ä* / + *-u*, *-y* | leipä / leib**y** ʻхлебʼ | lei**v**ä-n | leip**y**/leib**i**-ä | lei*p*/*b***i**-en |
| *-u*, *-y* | tukku ʻкучаʼ | tu**k**u-n | tukku-o | tukku-j/loi-en/n |
| *-o*, *-ö* | pelt/*d*o ʻполеʼ | pe**ll**o-n | pelt/*d***u**-o | pelt/*d*o-j/loi-en/n |

[1] Принятые обозначения: / – отделение собственно карельского от ливвиковского варианта (фонема, окончание, словоформа), обычный шрифт – слабая гласная основа, полужирный – сильная гласная основа, курсив – согласная снова, подчёркивание – изменения гласного компонента основы.
[2] Принятые обозначения: / – отделение собственно карельского от ливвиковского варианта, курсив – отличающиеся в наречиях фонемы, + – наличие второго возможного варианта, наряду с первым, в ливвиковском наречии, полужирный шрифт – чередования гласных и согласных (слабоступенный вариант).

| | | | | |
|---|---|---|---|---|
| *-i* | pappi ʻсвященникʼ | pa**p**i-n | pappi-e/i | pappi-en / -loi-n |
| *-i* | leht**i** ʻлистʼ | le***h***/*ht*e-n | leht**i**-e | leht**i**-en / +-löi-n |
| *VV* | mua ʻземляʼ | mua-n | mua-ta/du | m**ai**-j-en / mua-loi-n |
| **Двуосновные имена** | | | | |
| *-hi*, *-li*, *-mi*, *-ni*, *-ri*, *-si*, *-ši* | tuoh**i** ʻбересташ’<br>nuor**i** ʻмолодойʼ | tuohe-n<br>nuore-n | tuoh-ta/tu<br>nuor-ta/du | tuoh**i**-en / +-loi-n<br>nuor**i**-en |
| *-si*, *-ši* / *-zi*, *-ži* | ve*s*/*z***i** ʻводаʼ | v***e****j*/*ie*-n | vet-tä/ty | ve*s*/*z***i**-en / +-löi-n |
| *-ni* / *-ne* | nai**n*i*/*e*** ʻженщинаʼ | nai**s*/*ze***-n | nais-ta/tu | nai*s*/*z***i**-en |
| *-Ce* | vua**te** ʻодеждаʼ | vuattie-n | vuatet-ta/tu | vuatt**ei**-j/loi-en/n |
| *-C* | pereh ʻсемьяʼ | perehe-n | pereh-tä/ty | pereh**i**-en |
| *-š* / *-s* | armaš/*s* ʻлюбимыйʼ | armaha-n | armaš/*s*-ta/tu | armah**i**-en |
| *-š*, *-s* | kirveš/*s* ʻтопорʼ | kirvehe-n | kirveš/*s*-ta/ty | kirveh**i**-en |
| *-š* / *-s* | vereš/*s* ʻсвежийʼ | verekš/*s*e-n | vereš/*s*-tä/ty | verekš/*s***i**-en |
| *-š* / *-s*, *-t* | vähyš/*s* ʻнедостатокʼ<br>lyhyt ʻкороткийʼ | vähyö-n<br>lyhyö-n | vähyt-tä/ty<br>lyhyt-tä/ty | väh**yi**-jen / vähy**zi**-en<br>lyh**yi**-jen / lyhy**zi**-en |

### 3. Определение основ для генерации словоформ

Важнейшим условием этого этапа (таблица 3) является стремление к использованию минимального набора исходных грамматических данных о лексеме. Поскольку словоформа далеко не всегда содержит в себе однозначное указание на внешний вид гласной (слабой гласной) основы (*kirveš*/*s* – *kirvehe-*, *vereš*/*s* – *verekš*/*se-*), а также на отсутствие или наличие у лексемы согласной основы (*lehti* – *leh*/*hte-*; *pieni* – *piene-*, *pien-*), автоматическое их заполнение невозможно.

*Таблица 3*

**Выделение основ для генерации словоформ**[1]

| Осн. | Правила формирования основы | Пример |
|---|---|---|
| **с.ф.** | – \|\| и \| из словарного слова.<br>*С.с., в котором изменяемая часть отделяется от неизменяемой символом \|, а между составными частями сложного слова ставится \|\|.* | **ran*ta*/*du*** ʻберегʼ<br>**pel*t*/*d*o** ʻполеʼ<br>**mua** ʻземляʼ<br>**nuori** ʻмолодойʼ<br>**lyhyt** ʻкороткийʼ<br>**ves/*z*i** ʻводаʼ |
| **о.1** (сл. / гл. осн.) | **А. 1)** часть с.с. до \| + первое в скобках;<br>**2)** если нет \|, то с.с. + первое в скобках.<br>**Б.** Если в скобках два варианта п.о., то формируется о.1.1 (с левым вариантом п.о.) и о.1.2 (с правым вариантом п.о.).<br>*П.о. вносится в скобки: mua [], ran\|ta [na] / ran\|du [na], pel\|to [lo] / pel\|do [lo], nuor\|i [e], lyhy\|t [ö], ve\|si [je/te] / v\|ezi [ie/ede].* | A. ran+na=**ranna**<br>pel+lo=**pello**<br>**mua**<br>nuor+e=**nuore**<br>lyhy+ö=**lyhyö**<br>**Б.** ve+je=**veje**, ve+te=**vete** /<br>v+ie=**vie**, v+ede=**vede** |
| **о.2** (сильн. / согл. осн.) | **А.** Если в скобках одно (нуль) п.о., то с.ф. – конечные V, VV → + конечные V, VV из о.1.<br>**Б.** Если в скобках два п.о., то:<br>**1)** часть с.с. до \| + второе в скобках;<br>**2)** если нет \|, то с.с. + второе в скобках.<br>*Второе п.о. для двуосн. имен вносится в скобки через запятую от первого п.о.: nuor\|i [e, -], lyhy\|t [ö, t], ve\|si [je/te, t] / v\|ezi [ie/ede, et].* | **А.** ran*t*/*d*+a=**ran*t*/*d*a**<br>pel*t*/*d*+o=**pel*t*/*d*o**<br>**mua**<br>**Б.** nuor+0=**nuor**<br>lyhy+t=**lyhyt**<br>ve+t/v+et=**vet** |

[1] Принятые для таблиц 3 и 4 обозначения: || – отделение составных частей сложного слова, | – отделение неизменяемой части слова от изменяемой, – – убрать, + – добавить, = – равно, > – переход, / – отделение с.к. варианта от ливв., ~ – отделение заднерядного варианта показателя от переднерядного, ^ – отделение вариантов показателей, содержащих шипящие или свистящие согласные компоненты, V – гласный, C – согласный, → – переход к следующему действию, курсив в примерах – отличающиеся в наречиях фонемы, курсив в правилах – алгоритм выведения шаблона леммы.

| | | |
|---|---|---|
| **o.3** (вспом. осн. мн., сильн.) | **А.** Если о.2 заканч. на<br>**1)** C*i*, C*u*, C*y*, а также C*o*, C*ö* (в ***ливв***. всегда, в ***с.к.*** если в о.2 2 слога), то о.2+**loi**~**löi**;<br>**2) *с.к.*:** C*o*, C*ö* и в о.2 3 слога, то о.2+**i** (*karpalo*+*i*);<br>**3)** C*e*, то *e*>**i**;<br>**4)** C*a*, то *a*>**i** или *a*>**oi**[1];<br>**5)** C*ä*, то *ä*>**i** или *ä*>**öi**[2];<br>**6)** V*i*, то о.2+**loi**~**löi** (*täi*+*löi*);<br>**7)** VV (не V*i*), то: ***с.к.:*** – первый V → +**i**; ***ливв.:*** о.2+**loi**~**löi**.<br>**Б.** Если о.2 заканч. на C, то о.2>о.1 → если о.1 заканч. на<br>**1)** *e* и с.ф. заканч. на *si*, *ši*, *zi*, *ži*, то о.3=с.ф.;<br>**2)** C*e* и с.ф. не заканч. на *si*, *ši*, *zi*, *ži*, то *e*>**i**;<br>**3)** C*a*, C*ä*, то *a*, *ä*>**i**;<br>**4)** VV, то ***с.к.:*** *uo*>**ui**, *yö*>**yi**, *ie*>**ei** (*vuattie* > *vuattei*); ***ливв.:*** о.2+**loi**~**löi** (*vuattie*+*loi*), если о.1 на *uo*, *yö*, то – *o*, *ö* → +***zi***. | A. pel***t***/***d***o+loi=**pel*t*/*d*oloi**<br>ran***t***/***d***a>**ran*t*/*d*oi**<br>mua–u=ma+i=**mai** /<br>mua+loi=**mualoi**<br>Б. nuore>**nuori**<br>lyhyö>**lyhyi** /<br>lyhy–ö= lyhy+zi=**lyhyzi**<br>veje/vie**>ves/*z*i** |
| **o.4** (вспом. осн. мн., сл.) | **А.** Если о.3 заканч. на C*oi*~Cöi, о.2 на C*a*, C*ä* и кол-во слогов в о.3=о.2, то о.3>о.1 → если о.1 заканч. на<br>**1)** C*a*, C*ä*, то *a*>**oi**, *ä*>**öi**;<br>**2)** *ua*, *iä*, то *ua*>**avoi**, *iä*>**ävöi** (*padoi* > *pua* > *pavoi*).<br>**Б.** Если о.3 заканч. на *loi*~*löi* и кол-во слогов в о.3 больше чем в о.2, то о.4=о.3.<br>**В.** Если о.3 заканч. на V*i* и о.1 на VV, то о.4=о.3.<br>**Г.** Если о.3 заканч. на C*i* и перед конечным *i* согл. *čč*, *šš*, *ss*, *k*, *t*, *p*, *g*, *d*, *b*<br>**1)** нет, то о.4=о.3;<br>**2)** есть, то о.3>о.1 → если о.1 заканч. на CV, то V>**i**; если о.1 заканч. на *ie*>**ei**, *ua*>**ai**, *iä*>**äi** (*regi*>*rie*>*rei*). | A. ran***t***/***d***oi>ranna<br>>**rannoi**<br>Б. **pel*t*/*d*oloi**<br>**mualoi** (ливв.)<br>В. **mai** (с.к.)<br>**lyhyi** (с.к.)<br>Г. **nuori**<br>**lyhyzi** (ливв.)<br>**ves/*z*i** |

### 4. Разработка системы правил генерации словоформ

Прежде чем перейти к рассмотрению правил (таблица 4), важно обратить внимание на такую особенность фонетической системы карельского языка, как сингармонизм. Многие аффиксы представлены как в переднерядном, так и заднерядном вокалическом оформлении (*na*~*nä*, *nnu*~*nny*, *tta*~*ttä*, *ttah*~*ttäh*). В связи с этим для выбора варианта грамматического показателя вводится следующее дополнительное правило: 1) если в с.ф. есть гласные *a*, *o*, *u*, используется левый от ~ вариант показателя; 2) если в с.ф. нет гласных *a*, *o*, *u*, используется правый от ~ вариант показателя. Для сложных слов программа должна ориентироваться на правую от || часть.

Кроме того, в собственно карельском новописьменном варианте важную роль играет дистрибуция свистящих и шипящих согласных, от которой также зависит выбор одного из парных вариантов падежных окончаний (*kši*^*ksi*, *šša*~*ššä*^*ssa*~*ssä*). В таком случае действует правило: 1) если о.1 не заканчивается на *i*, используется левый от ^ вариант; 2) если о.1 заканчивается на *i*, используется правый от ^ вариант.

*Таблица 4*

**Система правил генерации словоформ именного словоизменения новописьменных вариантов карельского языка**

| Правило | Пример | Правило | Пример |
|---|---|---|---|
| **Единственное число** | | **Множественное число** | |
| **Ном.**: с.ф. | ran*ta*/*du* ‘берег’<br>pel*t*/*d*o ‘поле’<br>mua ‘земля’<br>nuori ‘молодой’<br>lyhyt ‘короткий’<br>ves/*z*i ‘вода’ | **Ном.**: о.1(о.1.1)+**t** | ranna+t<br>pello+t<br>mua+t<br>nuore+t<br>lyhyö+t<br>veje+t / vie+t |

[1] А. a > i если в о.2: 1) 2 слога и в первом слоге есть u, o (в том числе в составе VV); 2) более 2-х слогов и о.2 заканч. на mpa / mba, ma или является прил. с о.2. на va.
Б. a > oi если в о.2: 1) 2 слога и в первом слоге есть гласные a, e, i (в том числе в составе VV); 2) более 2-х слогов (кроме случаев А.2).
[2] А. ä > i если в о.2: 1) 2 слога; 2) более 2-х слогов и о.2 заканч. на mä, vä, žä, sä, jä, pä, bä.
Б. ä > öi если в о.2 более 2-х слогов (кроме случаев А.2).

| | | | |
|---|---|---|---|
| **Ген**.: о.1(о.1.1)+**n** | ranna+n<br>pello+n<br>mua+n<br>nuore+n<br>lyhyö+n<br>veje+n / vie+n | **Ген**.: если о.3 заканч. на<br>**1)** C*i*, то о.3+**en**;<br>***лівв.***: **2)** *oi~öi*, то о.3+**n** и о.4+**n**;<br>***с.к.***: **3)** *loi~löi* и кол-во слогов в о.3 больше чем в о.2, то *loi~löi*>**jen**;<br>**4)** V*i* и кол-во слогов в о.3=о.2, а о.2 заканч. на<br>– CV, то *i*>**jen**;<br>– VV, то о.3+**jen** | rantoi>rantojen / randoi+n и<br>rannoi+n<br>peltoloi>peltojen /<br>peldoloi+n<br>mai+jen / mualoi+n<br>nuori+en<br>lyhy*i*/*zi*+jen/en<br>ve*s*/*z*i+en |
| **Парт**.: если о.2 заканч. на<br>**1)** CV, то *a*>*u***a**, *ä*>*y***ä**/*i***ä**, *u*>*u***o**, *y*>*y***ö**, *o*>*u***o**, *ö*>*y***ö**, *e*>*i***e**, *i*>*i***e**/*i***i**;<br>**2)** VV, то о.2+**ta~tä/ du~dy;**<br>**3)** C, то о.2+**ta~tä**/**du~dy** (о.2 на *l*, *n*, *r*) или **ty~ty** (о.2 на *h*, *s*, *t*) | ran*t*/*d*ua<br>pel*t*/*d*uo<br>mua+ta/du<br>nuor+ta/du<br>lyhyt+tä/ty<br>vet+tä/ty | **Парт**.: если о.3 заканч. на<br>**1)** C*i*, то о.3+**e/i**;<br>***лівв.***: **2)** *oi~öi*, то=о.3;<br>***с.к.***: **3)** *loi~löi* и кол-во слогов в о.3 больше чем в о.2, то *loi~löi*>**ja~jä**;<br>**4)** V*i* и кол-во слогов в о.3=о.2 о, а о.2 заканч. на<br>– CV, то *i*>**ja~jä**;<br>– VV, то о.3+**ta~tä** | rantoi>rantoja / randoi<br>peltoloi>peltoja / peldoloi<br>mai+ta / mualoi<br>nuori+e/i<br>lyhy*i*/*zi*+tä/i<br>ve*s*/zi+e/i |
| **Эсс**.:<br>***с.к.*:** если о.2 заканч. на<br>**1)** V, то о.2+**na~nä**;<br>**2)** C, то о.1(о.1.2)+**na~nä**<br>***лівв.*:** о.1(о.1.1)+**nnu~nny** | ranta+na / ranna+nnu<br>pelto+na /<br>pello+nnu<br>mua+na/nnu<br>nuore+na/nnu<br>perehe+nä/nny<br>lyhyö+nä/nny<br>vete+nä / vie+nny | **Эсс**.: о.3+**na~nä** / о.4+**nnu~nny**<br>От о.3/о.4 образуются:<br>*ком*. (с.к.): о.3+**neh**<br>*ком.* (лівв.): о.4+**nke**(**nneh**) (о.4. на *oi*, *öi*) или **enke**(**nneh**) (о.4. на C*i*) | rantoi+na / rannoi+nnu<br>leipi+nä / leivi-nny<br>pel*t*/*d*oloi+na/nnu<br>mai+na / mualoi+nnu<br>nuori+na/nnu<br>lyhy*i*/*zi*+nä/nny<br>ve*s*/zi+nä/nny |
| **Транс**.: о.1(о.1.1)+**kši^ksi/ kse**<br>От о.1(о.1.1)+ образуются:<br>*инесс*.: **šša~ššä^ssa~ssä/s**<br>*элат*.: **šta~štä^sta~stä/späi**<br>*адесс*.: **lla~llä/l**<br>*абл*.: **lta~ltä/lpäi**<br>*абесс*.: **tta~ttä/ttah~ttäh**<br>*алл*. (лівв.): **le**<br>*прол*. (лівв.): **či**<br>*ком*. (лівв.): **nke** | ranna+t<br>pello+t<br>mua+t<br>nuore+t<br>lyhyö+t<br>veje+t / vie+t | **Транс**.: о.4+**ksi/kse**<br>От о.4+ образуются:<br>*инесс*.: **ssa~ssä/s**<br>*элат*.: **sta~stä/späi**<br>*адесс*.: **lla~llä/l**<br>*абл*.: **lta~ltä/lpäi**<br>*абесс*.: **tta~ttä/ttah~ttäh**<br>*инстр*.: **n**<br>*алл*. (лівв.): **le**<br>*прол*. (лівв.): **či** | rannoi+ksi/kse<br>pel*t*/*d*oloi+ksi/kse<br>mai+ksi / mualoi+kse<br>nuori+ksi/kse<br>lyhy*i*/*zi*+ksi/kse<br>ve*s*/zi+ksi/kse |
| **Илл**.: если о.2 заканч. на<br>**1)** V, то о.2+**h**;<br>**2)** C, то о.1(о.1.2)+**h** | ran*t*/*d*a+h<br>pel*t*/*d*o+h<br>mua+h<br>nuore+h<br>lyhyö+h<br>ve*t*/*d*e+h | **Илл**.: о.3+**h** | ran*t*/*d*oi+h<br>pel*t*/*d*oloi+h<br>mai/mualoi+h<br>nuori+h<br>lyhy*i*/*zi*+h<br>ve*s*/zi+h |

### 5. Разработка программы генерации

На основе полученных правил в комплексе программ Dictorpus реализованы функции автоматической генерации словоформ по минимизированному шаблону для двух наречий карельского языка. Редактор словаря ВепКар в поле «лемма» вводит шаблон (*mua* [], *ran*|*ta* [*na*] / *ran*|*du* [*na*], *nuor*|*i* [*e,* ]) и система автоматически создаёт 24 словоформы для собственно карельского или 31 – для ливвиковского наречия (рис. 2).

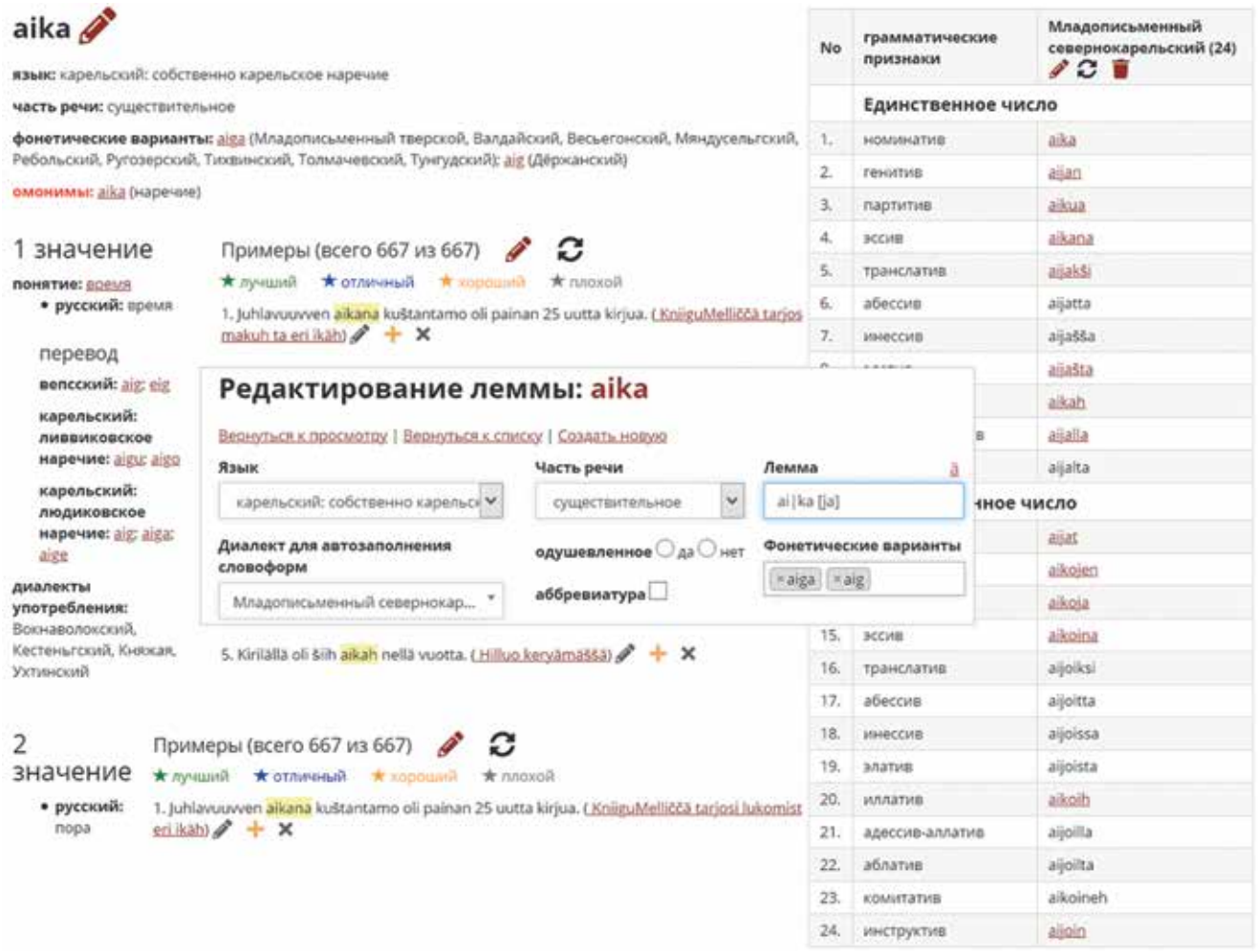


**Рис. 2.** Добавление словоформ через шаблон леммы в словаре ВепКар.

Благодаря новым правилам в словарь ВепКар добавлено 600 тыс. ливвиковских словоформ, что увеличило количество размеченных токенов в текстах подкорпуса на 35%. В ближайшее время предстоит заполнение собственно карельской части словаря.

Программный код написан на языке PHP с использованием платформы Laravel и размещён в открытом доступе на сайте GitHub[1].

## Обсуждение и заключение

В процессе исследования именной словоизменительной системы новописьменных вариантов карельского языка с целью создания программы автоматической генерации словоформ удалось выделить и унифицировать правила образования всех грамматических форм для имён основных словоизменительных типов, совпадающих в ливвиковском и собственно карельском наречиях. Количество минимальной грамматической информации об имени, которую необходимо вручную вносить в шаблоны для дальнейшего автоматического заполнения полной парадигмы слова, удалось сократить до нуля или одного псевдоокончания для одноосновных имён (*mua []*, *koivu []*, *uk|ko [o]*) и до двух или трёх – для двуосновных (*arma|š [ha*, *š]* / *arma|s [ha*, *s]*, *ve|si [je/te*, *t]* / *v|ezi [ie/ede*, *et]*). Достигнуть таких результатов помогли эксперименты, на основе которых для одноосновных имён с основой на *a-*, *ä-* были разработаны правила образования форм множественного числа.

В процессе проверки программы было выявлено отсутствие унификации в образовании формы множественного числа генитива в ливвиковском новописьменном варианте: в литературных текстах обнаружено два варианта образования этой формы (от сильно- и от слабоступенной основы). Для корпуса ВепКар было принято решение о включении обоих вариантов с целью увеличения процента разметки текстов. При этом отмечена крайняя необходимость доработки правила для нормированного варианта языка.

Предложенные в статье правила применимы не ко всем именам, а только к существительным,

[1] См. https://github.com/componavt/dictorpus

прилагательным, числительным и второстепенным местоимениям (пришедшим из других частей речи). Первостепенные по происхождению местоимения (личные, указательные, вопросительные) и их словоформы в процессе развития языка претерпели значительные изменения и приобрели целый ряд особенностей, что делает невозможным применение к ним программы генерации. Поскольку количество таких местоимений невелико, ручное заполнение их словоизменительных парадигм не представляет труда.

Автоматизировать процесс генерации именной словоизменительной парадигмы полностью, т. е. обучить программу распознавать словоизменительный тип имени и учитывать особенности чередования ступеней согласных основы, не представляется возможным. Однако использование программы генерации именных словоформ в том виде, в котором она существует на настоящий момент, позволило сократить время заполнения одной именной парадигмы до 10–20 секунд или в среднем в 20 раз.

Проведённое исследование позволило существенно приблизиться к созданию приложения для проверки правописания карельского языка, а также подойти к разработке морфоанализатора и автоматического перевода. Правила образования форм множественного числа, которые удалось дополнить в ходе работы, найдут применение в практике преподавания карельского языка.

**Список сокращений**

абесс. – абессив
абл. – аблатив
адесс. – адессив
алл. – аллатив
вспом. осн. – вспомогательная основа
ед. – единственное число
ген. – генитив
гл. осн. – гласная основа
двуосн. – двуосновное имя
илл. – иллатив
инесс. – инессив
инстр. – инструктив
ком. – комитатив
ливв. – ливвиковское наречие
мн. – множественное
ном. – номинатив
одноосн. – одноосновное имя
осн. – основа
п.о. – псевдоокончание
парт. – партитив
прил. – прилагательное
прол. – пролатив
сильн. осн. – сильная основа
с.к. – собственно карельское наречие
с.с. – словарное слово
с.ф. – словарная форма
сл. осн. – слабая основа
согл. осн. – согласная основа
транс. – транслатив
ч. – число
элат. – элатив
эсс. – эссив

**Список источников и литературы**

1. Бойко Т. П. Большой карельско-русский словарь (ливвиковское наречие). Петрозаводск: Периодика. 2016. 352 с.

2. Карельский язык в грамматиках / Новак И., Пенттонен М., Руусканен А., Сиилин Л. Петрозаводск: КарНЦ РАН, 2019. 479 с.

3. Крижановская Н. Б., Новак И. П. New written Tver Karelian dialects word form generator. Свидетельство о гос. регистрации программы для ЭВМ Федеральной службы по интеллектуальной собственности № 2019665163 от 20 ноября 2019 г. URL: https://new.fips.ru/publication-web/publications/document?type=doc&tab=PrEVM&id=6E578F3E-989F-44C6-A973-54863D199455 (дата обращения: 16.05.2020).

4. Нагурная С. В. Карельская письменность // Народы Карелии: историко-этнографические очерки. Петрозаводск: Периодика, 2019. С. 65–77.

5. A morphological analyzer for Kven / Trosterud S. R., Trosterud T., Räisänen A. K., Niiranen L., Haavisto M., Maliniemi K. // Proceedings of the Third Workshop on Computational Linguistics for Uralic Languages, 2017. Pp. 76–88.

6. Fine-grained Morphosyntactic Analysis and Generation Tools for More Than One Thousand Languages / Garrett N., Dylan L., McCarthy A. D., Mueller A., Wu W., Yarowsky D. // Proceedings of The 12th Language Resources and Evaluation Conference. 2020. Pp. 3963–3972.

7. Institute for the Languages of Finland. The Helsinki Korp Version of Samples of Spoken Finnish [text corpus]. Kielipankki. URL: http://urn.fi/urn:nbn:fi:lb-2016042702 (дата обращения: 15.05.2020).

8. Iso suomen kielioppi / Hakulinen A., Vilkuna M., Korhonen R., Koivisto V., Heinonen T.R., Alho I. Helsinki: Suomalaisen Kirjallisuuden Seura. 2008. Verkkoversio. URL: http://scripta.kotus.fi/visk (дата обращения: 28.05.2020).

9. Karjalan kielen verkkosanakirja / Toim. M. Torikka. URL: http://kaino.kotus.fi/cgi-bin/kks/karjala.cgi (дата обращения: 29.05.2020).

10. Kielikorpuksia Suomen itärajalta / Palander M., Riionheimo H., Kemppanen H., Mäkisalo J. // Multi lingual Finnic. Helsinki: Suomalais-Ugrilainen Seura, 2019. Pp. 425–438.

11. Koivisto V. Border Karelian dialects – a diffuse variety of Karelian // On the border of language and dialect. Helsinki: Suomalaisen Kirjallisuuden Seura, 2018. Pp. 56–84.

12. Markianova L. Karjalan kielioppi. Petroskoi: Periodika, 2002. 296 s.

13. Palander M., Riionheimo H., Koivisto V. Kotimaisten kielten keskus, Institute for the Languages of Finland. Raja-Karjalan korpus. 2017. URL: http://urn.fi/urn:nbn:fi:lb-2014073033 (дата обращения: 28.05.2020).

14. Pyöli R. Livvinkarjalan kielioppi. Helsinki: Karjalan Kielen Seura, 2011. 200 p.

15. Rueter J., Hämäläinen M. Synchronized Mediawiki based analyzer dictionary development // Proceedings of the Third Workshop on Computational Linguistics for Uralic Languages. 2017. Pp. 1–7.

16. Zaiceva N. Veps language heritage in Karelia // Multi lingual Finnic. Helsinki: Suomalais-Ugrilainen Seura, 2019. Pp. 379–400.

17. Zaikov P. Karjalan kielioppi. Petroskoi: Periodika, 2002. 208 p.

18. Zaikov P. Vienankarjalan kielioppi. Helsinki: Karjalan Sivistysseura, 2013. 284 p.

**References**

1. Boyko T. P. *Bol'shoj karel'sko-russkij slovar'* [Large Russian-Karelian dictionary]. Petrozavodsk: Periodika Publ., 2016. 352 p. (In Russian)

2. Novak I., Penttonen M., Ruuskanen A., Siilin L. *Karel'skiy yazyk v grammatikakh* [Karelian language in grammars]. Petrozavodsk: KarRC RAS Publ., 2019. 479 p. (In Russian)

3. Krizhanovskaya N. B., Novak I. P. New written Tver Karelian dialects word form generator. *Svidetel'stvo o gos. registratsii programmy dlya EVM Federal'noy sluzhby po intellektual'noy sobstvennosti № 2019665163 ot 20 noyabrya 2019 g.* [Certificate of state registration of the program, № 2019665163, November 20, 2019]. Available at: https://new.fips.ru/publication-web/publications/document?type=doc&tab=PrEVM&id=6E578F3E-989F-44C6-A973-54863D199455 (accessed May 16, 2020). (In Russian)

4. Nagurnaya S. V. *Karel'skaya pis'mennost'* [Karelian writing]. *Narody Karelii: istoriko-etnograficheskie ocherki* [Peoples of Karelia: historical-ethnographic essays]. Petrozavodsk: Periodika Publ., 2019. pp. 65–77. (In Russian)

5. A morphological analyzer for Kven. Trosterud S. R., Trosterud T., Räisänen A. K., Niiranen L., Haavisto M., Maliniemi K. *Proceedings of the Third Workshop on Computational Linguistics for Uralic Languages*, 2017, pp. 76–88. (In English)

6. Fine-grained Morphosyntactic Analysis and Generation Tools for More Than One Thousand Languages. Garrett N., Dylan L., McCarthy A. D., Mueller A., Wu W., Yarowsky D. *Proceedings of The 12th Language Resources and Evaluation Conference*, 2020, pp. 3963–3972. (In English)

7. Institute for the Languages of Finland. The Helsinki Korp Version of Samples of Spoken Finnish [text corpus]. Kielipankki. Available at: http://urn.fi/urn:nbn:fi:lb-2016042702 (accessed May 15, 2020). (In English)

8. Hakulinen A., Vilkuna M., Korhonen R., Koivisto V., Heinonen T.R., Alho I. *Iso suomen kielioppi.* Helsinki: Suomalaisen Kirjallisuuden Seura. 2008. Available at: http://scripta.kotus.fi/visk (accessed May 28, 2020). (In Finnish)

9. *Karjalan kielen verkkosanakirja.* Ed. by M. Torikka. Available at: http://kaino.kotus.fi/cgi-bin/kks/karjala.cgi (accessed May 29, 2020). (In Finnish)

10. *Kielikorpuksia Suomen itärajalta*. Palander M., Riionheimo H., Kemppanen H., Mäkisalo J. Multi lingual Finnic. Helsinki: Suomalais-Ugrilainen Seura, 2019. pp. 425–438. (In Finnish)

11. Palander M., Riionheimo H., Kemppanen H., Mäkisalo J. Multi lingual Finnic *Kielikorpuksia Suomen itärajalta*. Helsinki: Suomalais-Ugrilainen Seura, 2019. pp. 425–438. (In Finnish)

12. Markianova L. Karjalan kielioppi. Petrozavodsk: Periodika, 2002. 296 p. (In Karelian)

13. Palander M., Riionheimo H., Koivisto V. *Kotimaisten kielten keskus, Institute for the Languages of Finland. Raja-Karjalan korpus.* 2017. Available at: http://urn.fi/urn:nbn:fi:lb-2014073033 (accessed May 28, 2020). (In Finnish)

14. Pyöli R. *Livvinkarjalan kielioppi*. Helsinki: Karjalan Kielen Seura, 2011. 200 p. (In Finnish)

15. Rueter J., Hämäläinen M. Synchronized Mediawiki based analyzer dictionary development. *Proceedings of the Third Workshop on Computational Linguistics for Uralic Languages,* 2017. pp. 1–7. (In English)

16. Zaiceva N. *Veps language heritage in Karelia*. Multi lingual Finnic. Helsinki: Suomalais-Ugrilainen Seura, 2019. pp. 379–400. (In English)

17. Zaikov P. *Karjalan kielioppi*. Petrozavodsk: Periodika, 2002. 208 p. (In Karelian)

18. Zaikov P. *Vienankarjalan kielioppi*. Helsinki: Karjalan Sivistysseura, 2013. 284 p. (In Finnish)

**ИНФОРМАЦИЯ ОБ АВТОРАХ**

**Новак Ирина Петровна**, научный сотрудник сектора языкознания Института языка, литературы и истории КарНЦ РАН, ФИЦ «Карельский научный центр» (185910, Российская Федерация, Республика Карелия, г. Петрозаводск, ул. Пушкинская, д. 11), кандидат филологических наук.

novak@krc.karelia.ru

ORCID.ID: 0000-0002-9436-9460

**Крижановская Наталья Борисовна**, ведущий инженер-исследователь лаборатории информационных компьютерных технологий Института прикладных математических исследований КарНЦ РАН, ФИЦ «Карельский научный центр» (185910, Российская Федерация, Республика Карелия, г. Петрозаводск, ул. Пушкинская, д. 11).

nataly@krc.karelia.ru

ORCID ID: 0000-0002-9948-1910

**Бойко Татьяна Петровна**, научный сотрудник сектора языкознания Института языка, литературы и истории КарНЦ РАН, ФИЦ «Карельский научный центр» (185910, Российская Федерация, Республика Карелия, г. Петрозаводск, ул. Пушкинская, д. 11).

boiko@krc.karelia.ru

ORCID ID: 0000-0001-5095-2921

**Пеллинен Наталия Александровна**, младший научный сотрудник сектора языкознания Института языка, литературы и истории КарНЦ РАН, ФИЦ «Карельский научный центр» (185910, Российская Федерация, Республика Карелия, г. Петрозаводск, ул. Пушкинская, д. 11), кандидат филологических наук.

nataliapellinen@gmail.com

ORCID ID: 0000-0002-5648-6877

**ABOUT THE AUTHORS**

**Novak Irina Petrovna**, Researcher, Department of Linguistics, Institute of Linguistics, Literature and History, Karelian Research Centre of the RAS (185910, Russian Federation, Republic of Karelia, Petrozavodsk, Pushkinskaya st., 11), Candidate of Philological Sciences.

novak@krc.karelia.ru

ORCID ID: 0000-0002-9436-9460

**Krizhanovskaya Natalya Borisovna**, Leading Engineer Research, Laboratory for Information Computer Technologies, Institute of Applied Mathematical Research, Karelian Research Center of the RAS (185910, Russian Federation, Republic of Karelia, Petrozavodsk, Pushkinskaya st., 11).

nataly@krc.karelia.ru

ORCID ID: 0000-0002-9948-1910

**Boyko Tatyana Petrovna**, Researcher, Department of Linguistics, Institute of Linguistics, Literature and History, Karelian Research Centre of the RAS (185910, Russian Federation, Republic of Karelia, Petrozavodsk, Pushkinskaya st., 11).

boiko@krc.karelia.ru

ORCID ID: 0000-0001-5095-2921

**Pellinen Nataliya Aleksandrovna**, Junior Researcher, Department of Linguistics, Institute of Linguistics, Literature and History, Karelian Research Centre of the RAS (185910, Russian Federation, Republic of Karelia, Petrozavodsk, Pushkinskaya st., 11), Candidate of Philological Sciences.

nataliapellinen@gmail.com

ORCID ID: 0000-0002-5648-6877