

Титульный лист отчета о работе в 2012 г.
по Программе фундаментальных исследований Президиума РАН
«Корпусная лингвистика»

Номер и название направления: Программа фундаментальных исследований Президиума РАН «Корпусная лингвистика» 2012—2014	
Направление 3 «Создание и развитие корпусных ресурсов по языкам народов России»	
Название проекта: « Корпус вепского языка: пополнение и развитие электронного ресурса »	
Научный руководитель проекта (ФИО полностью, уч. ст.): доктор филологических наук Зайцева Нина Григорьевна	
Е-mail, телефон, факс (с кодом города) почтовый адрес руководителя проекта	zaitseva@karelia.ru
Полное и краткое название организации – адресата финансирования: Учреждение Российской академии наук Институт языка, литературы и истории Карельского научного центра РАН ИЯЛИ КарНЦ РАН	ФИО (полностью) руководителя организации – адресата финансирования: Муллонен Ирма Ивановна
	ФИО (полностью) главного бухгалтера организации – адресата финансирования: Соболева Ирина Евгеньевна
	Телефон, факс (с кодом города), Е-mail организации – адресата финансирования: 185910 г. Петрозаводск, ул. Пушкинская, д. 11; тел./факс 8 (814 2) 781886, illh@krc.karelia.ru
Год начала – год окончания проекта	2012—2014
Объем финансирования, полученного в 2012 г.:	Объем финансирования, запрашиваемый на 2013 г.:
Исполнители (ФИО, уч. ст.)	Научный консультант проекта: Муллонен Ирма Ивановна, директор ИЯЛИ КарНЦ РАН, доктор филол. наук, профессор.
	Исполнитель проекта: Шибанова Нина Леонидовна, главный специалист по информационным технологиям ИЯЛИ КарНЦ РАН.
	Исполнитель проекта: Луговая Наталья Борисовна, ведущий инженер-программист лаборатории информационных компьютерных технологий Института прикладных математических исследований КарНЦ РАН.
	Исполнитель проекта: Жукова Ольга Юрьевна, ст. преподаватель Факультета прибалтийско-финской филологии и культуры Петрозаводского университета, канд. филол. наук.
	Исполнитель проекта: Бовин Владимир Борисович, ведущий инженер Фонограммархива ИЯЛИ КарНЦ РАН.
Дата сдачи отчета: 15.11.2012.	Подпись руководителя проекта: Зайцева Н.Г.

Координатор Программы

акад. Вяч. Вс. Иванов

Координатор Программы

чл-корр. РАН В.А. Плунгян

«__»_____2012 г.

1. Название программы и направления: Программа фундаментальных исследований Президиума РАН «**Корпусная лингвистика**» 2012—2014; Направление 3 «**Создание и развитие корпусных ресурсов по языкам народов России**».
2. Название проекта «**Корпус вепсского языка: пополнение и развитие электронного ресурса**».
3. Руководитель проекта (ФИО *полностью*, ученая степень, должность, место работы): **Зайцева Нина Григорьевна**, доктор филологических наук, зав. сектором языкознания ИЯЛИ КарНЦ РАН.
4. Основные участники проекта (ФИО *полностью*, ученая степень, должность, место работы):
 - 1) **Муллонен Ирма Ивановна**, директор ИЯЛИ КарНЦ РАН, доктор филол. наук, профессор;
 - 2) **Шибанова Нина Леонидовна**, главный специалист по информационным технологиям ИЯЛИ КарНЦ РАН;
 - 3) **Бовин Владимир Борисович**, ведущий инженер Фонограммархива ИЯЛИ КарНЦ РАН (услуги инженера в 2012 г. не были востребованы);
 - 4) **Жукова Ольга Юрьевна**, старший преподаватель Факультета прибалтийско-финской филологии и культуры Петрозаводского государственного университета, кандидат филол. наук (в 2012 г. находилась в отпуске по уходу за ребенком и в работе и участвовала в работе в течение 1 месяца);
 - 5) **Луговая Наталья Борисовна**, ведущий инженер-программист лаборатории информационных компьютерных технологий Института прикладных математических исследований КарНЦ РАН. (В начале работы над проектом в 2012 г. Луговая Н.Б. была заменена кандидатурой **Крижановского Андрея Анатольевича**, кандидата технических наук, старшего научного сотрудника лаборатории информационных компьютерных технологий Института прикладных математических исследований КарНЦ РАН, который переехал на работу в Карельский научный центр в начале 2012 г. Его научные интересы

сосредоточены именно в области компьютерной лингвистики, автоматического извлечения данных из Интернета).

5. Содержание фактически проделанной работы, полученные результаты, их новизна, научная и практическая значимость, актуальность, соответствие мировому научному уровню (до 5 стр.):

Создание лингвистических корпусов – сбалансированных коллекций текстов со встроенным лингвистическим аппаратом или разметкой, является на сегодняшний день одним из актуальнейших направлений современной лингвистики, которое способно вывести верификацию лингвистических гипотез, составление словарей и грамматик, исследований фольклорного текста на иной уровень. До сих пор исследователь редкого языка, каковым является вепсский, был лишен возможности проверить те или иные данные, приводимые в научных работах. В то же время лингвистам, лексикографам, фольклористам, этнографам и заинтересованным читателям для полноценного изучения языка и культуры народов необходим свободный доступ к текстам на названных языках. Развитие компьютерных технологий позволяет удовлетворить такую потребность и обеспечить удобный доступ к корпусу текстов и словарям языков, представленных в сети Интернет.

Как известно, корпус должен характеризоваться представительностью и сбалансированным составом текстов. Корпус вепсского языка, таким образом, должен содержать пропорционально по возможности все типы текстов; диалектные, фольклорные, младописьменные. Ясно, что представительность корпуса языка малочисленного народа, занесенного в Красную книгу языков народов России (Энциклопедический словарь-справочник. М., 1994), не может достигать десятков и сотен миллионов словоупотреблений, как это имеет место в Национальном русском корпусе, тем не менее, создатели Корпуса вепсского языка понимают свои задачи и стремятся включить в корпус разнообразные материалы достаточно пропорционально, чтобы представить вепсский язык на данном этапе развития максимально более полно.

Работа по реструктуризации Корпуса вепсского языка, пополнению его новыми материалами и идеями, которая начата в 2012 году, позволяет говорить о возможности разработки модели документации языков малочисленных народов, которые уже в ближайшие десятилетия могут исчезнуть с лингвистической карты России. Можно утверждать, что при выполнении данного проекта в настоящее время создается своеобразный музей вепсского языка с широким и доступным кругом экспонатов различного характера: бытового и этнографического, фольклорного и прозаического, диалектного и младописьменного, на которые можно ориентироваться при описании других языков малочисленных народов.

Материалы корпуса могут стимулировать появление новейших работ с использованием вепсского языка в российском и зарубежном языкознании, что косвенным образом может повлиять и на популяризацию вепсского языка и культуры в целом.

На 2012 г. по проекту: «Корпус вепсского языка: пополнение и развитие электронного ресурса» было запланировано:

- 1). Пополнение корпуса новыми текстами: 100 текстов различного характера; расшифровка текстов нового для корпуса жанра внеобрядовых причитаний с маг. записей и снабжение текстов причитаний параллельными переводами на русский язык (10 текстов);
- 2). Лемматизация: 800 лемм и 1000 словоформ; переводы лемм на русский и английский языки;
- 3). Решение проблем грамматической разметки; составление подробного перечня всех грамматических форм и их сокращений.

2012 год был посвящен решению проблем реструктуризации сайта и базы данных. Ранее он представлял собой единый массив лишь с выделенным параллельным подкорпусом вепсских причитаний (текст вепсского причитания и его перевод на русский язык), для которого наряду с общей поисковой системой была разработана и свойственная только ему система поиска по диалектной принадлежности текста причитания и его типу. В текущем году разрабатывались возможности по созданию 5 подкорпусов, которые в настоящее время находятся по адресу <http://vepsian.krc.karelia.ru/text>:

- 1) подкорпус диалектной вепсской речи;
- 2) подкорпус младописьменных вепсских текстов;
- 3) подкорпус переводных текстов на младописьменном вепсском языке (тексты перевода Нового Завета);

Фольклорные вепские материалы представлены в двух подкорпусах:

- 4) параллельный подкорпус вепсских причитаний;
- 5) подкорпус вепсских народных сказок с переводами на русский язык.

Фольклорная часть Корпуса вепсского языка в 2012 г. был пополнена подкорпусом «Вепские народные сказки». Интерес к вепским фольклорным материалам активен, но список специалистов со знанием вепсского языка исключительно скуден, поэтому с целью популяризации жанров вепсского фольклора, которые являются важнейшими компонентами самобытной духовной культуры вепсов, было решено и вепские народные сказки представить в электронном виде с переводами на русский язык. В течение всего 2012 г. велась компьютерная подготовка текстов. Сканирование, набор, редактирование и перевод текстов осуществляли Шибанова Н.Л., Зайцева Н.Г. Были расшифрованы с магнито-

фонных записей и переведены на русский язык 7 сказок из Фонограммархива ИЯЛИ РАН (Зайцева Н.Г.). Руководитель проекта Зайцева Н.Г. руководила практикой студентов Петрозаводского университета, обладающих знанием вепсского языка, которые в течение месяца занимались лемматизацией и вычиткой вепсскоязычных текстов. Работа студентов, а также иных помощников в работе была оплачена средствами из зарплаты по проекту Зайцевой Н.Г.

В настоящее время подкорпус вепсских народных сказок включает в себя 55 сказок с переводами на русский язык. Метатекстовая разметка подкорпуса сказок включает в себя:

- 1) фамилию исполнителя сказки, место и год его рождения;
- 2) наименование диалекта и говора, на котором исполнена сказка;
- 3) фамилию собирателя сказки;
- 4) соотношение со сравнительным указателем сюжетов восточнославянских сказок (Сравнительный указатель сюжетов: Восточнославянская сказка / Сост. Л.Г. Бараг, И.П. Березовский, К.П. Кабашников, Н.В. Новиков. Отв. редактор К.В. Чистов. Л., 1979 г.).

Подкорпус вепсских народных сказок находится пока в закрытом от посетителей режиме, поскольку готовится вступительная статья к подкорпусу, а также решаются проблемы поиска. Для подкорпуса уже характерна общая для всего Корпуса вепсского языка поисковая система, которая содержит поиск по текстам и поиск по электронному словарю. Для подкорпуса сказок будет введен дополнительно поиск по диалекту. В дальнейшем будет разработан поиск по исполнителю сказки и ее сюжету, поскольку такие данные содержатся в метатекстовой разметке.

В течение года велась работа в области электронного словаря. Было введено около 400 новых лемм и 1000 словоформ. В настоящее время все реже встречаются новые леммы, в большей степени идет наращивание материала словоформ, поскольку вепсский язык обладает исключительно развитой и богатой грамматической системой имени и глагола.

Материалы электронного словаря были активно использованы в течение 2012 г. при работе над «Орфографическим словарем вепсского языка», который находится в издательстве и выйдет в свет в 2012 г. впервые за всю историю функционирования младописьменного вепсского языка. При разработке орфограмм были необходимы сравнительные материалы, которые поддерживали бы выбор правил правописания в том или ином направлении. В качестве примера приведем сложности с выбором правил правописания для абстрактных существительных на **-uz, -ez**; čomuz' «красота», hūvūz' «доброта, добро»; käveluz «прогулка», armastuz «любовь», и т.д. Эти существительные обладают по диалектам разными вариантами основ: **-ude-** или **-use-**: čomuz' – čomude-n, armastuz – armastuse-n и т.д. С помощью электронного словаря, который пока функционирует, к со-

жалению, в недостаточно полном режиме, и нам пришлось вручную заниматься поисками каждого подобного существительного, удалось зафиксировать соотнесенность мягкого согласного звука на конце формы отадъективного абстрактного существительного -z' и основы на -ude-, и, в свою очередь, твердого согласного -z, которое образовано от глагола и основы на -use-. Это разработанное в том числе и с помощью электронного словаря и текстового материала Корпуса правило орфографии исключительно полезно, поскольку круг подобных существительных в языке немал и вызывал постоянные затруднения при изучении младописьменного вепского языка у учащихся, учителей и всех, кто связан с написанным словом.

И в целом, Электронный словарь и материалы текстов Корпуса вепского языка исключительно часто привлекались при строительстве орфограмм и решении проблем правописания отдельных вепских слов. «Орфографический словарь вепского языка», опубликованного в конце 2012 г. и содержащего помету о том, что он исполнялся, в том числе, и при поддержке проекта «Корпус вепского языка: пополнение и развитие электронного ресурса».

В течение года решались проблемы грамматической разметки. До настоящего времени грамматическая разметка корпуса достаточно скромна: идет указание на часть речи, а также дается перевод леммы на русский язык. С 2012 года начат перевод лемм и на английский язык, чтобы расширить круг пользователей корпуса. Осуществлен перевод ~ 1000 лемм. В 2012 г. в корпус введено 100 текстов (55 сказок, 24 причитания, 21 текст на младописьменном вепском языке). В настоящее время система грамматической разметки находится в поле пристального внимания разработчиков в плане ее совершенствования.

В связи с переносом сайта «Корпус вепского языка» на сайт Карельского научного центра РАН возникло много проблем технического характера. Технические проблемы содержания сайта и его разработки и реструктуризации разрабатываются в ином ключе. Решены проблемы вепского алфавита, которые оставались открытыми до конца 2011 года, когда шло неразличение буквенных символов, содержащих точки вверху (ä, ö, ü), что вводило путаницу в вепский алфавит и затрудняло нахождение слов.

С технической стороны в отчете о работе за 2012 год (полная версия технического отчета размещена в интернете по адресу: <http://scipeople.com/publication/111636/>) следует отметить, что шло описание и разработка ключевых особенностей компьютерной онлайн-системы, включающей словарь и корпус текстов вепского языка. База данных корпуса текстов и словаря реализована в единой реляционной базе данных. Все таблицы базы данных условно разделены на две группы или *таблицы*: таблицы, обслуживающие словарь, и таблицы, обслуживающие корпус текстов. Несмотря на это деление таблицы взаимосвязаны и представляют единое целое:

○ 1) *lemma* – таблица лемм. Особенность таблицы в том, что лемма не является уникальной в данной таблице, число лемм равно числу значений данного слова. Например, слово “*leta*” имеет два значения в вепском языке: 1) «летать» и 2) «поднимать, поднять». В таблице *lemma* будет две одинаковых записи “*leta*”, однако при редактировании словаря пользователь будет видеть две записи “*leta*” и “*leta* [2]”.

2) *wordform* – таблица словоформ. Словоформа привязана к лемме с помощью поля *lemma_id*. *translation_lemma* – перевод какого-либо значения с главного языка на один из иностранных. «Главным» языком в данном проекте является вепский, иностранный – это русский или английский. Таблица содержит следующие поля:

- *lemma* – текст перевода на один из иностранных языков;
- *lang_id* – идентификатор языка (отсылка к таблице *lang*), указывающий на каком иностранном языке записан перевод в поле “*lemma*”;

○ *translation* – таблица, связывающая лемму с переводом, т.е. связывает таблицы *lemma* и *translation_lemma*.

3) таблицы корпуса текстов:

○ *text* – ключевая таблица, описывающая и содержащая текст на каком-либо языке, содержит следующие поля:

- *text* – сам текст;
- *lang_id* – идентификатор языка, указывающий на каком языке записан текст в поле “*text*”, аутентичный (в данном случае – вепский) или русский (для текстов с параллельным переводом с вепского языка);
- *title* – название текста;
- *source_id* – идентификатор источника текста, описанного в таблице *source*;

informant – данные об информанте, со слов которого был записан данный текст;

○ *text_pair* – таблица, связывающая тексты с параллельным переводом, например, первый текст – источник на вепском и второй текст – перевод первого на русский язык.

○ *label* – описание характеристик текста. Что касается подкорпуса вепских причитаний, то в нем идет указание на следующее:

- вид причитаний (свадебные, похоронные или поминальные причитания);
- диалект (северновепский, средневепский или южновепский);
- говор (восточный или западный).

4) В особую группу выделены вспомогательные таблицы, содержащие «лексикографические константы», т.е. данные в этих таблицах изменяются очень редко, в результате согласованной работы инженера-программиста и лингвиста:

○ *lang* – таблица со списком языков, сейчас это вепский, русский и английский. В таблице указаны:

- уникальный идентификатор (поле “*lang_id*”);
 - название языка на русском;
 - название языка на английском;
 - буквенный код языка в соответствии со стандартом ISO 639.
- *part_of_speech* – список частей речи (сущ., прил., и т.д.), всего на данный момент в базе данных – 11 частей речи. Запись в таблице содержит уникальный идентификатор, название части речи на английском, на русском и текстовое сокращение для обозначений части речи.

Таким образом, на языке программирования PHP разработана объектно-ориентированная библиотека, как часть компьютерной онлайн-системы, включающей словарь и корпус вепского языка. В этой библиотеке для каждой из таблиц создан интерфейсный класс типа POPO (Plain Old PHP Object).

Руководитель проекта Зайцева Н.Г. организовала в Санкт-Петербурге в Институте лингвистических исследований *совещание* по возможности использования уже имеющихся Интернетресурсов: подкорпуса диалектных текстов, а также материалов электронного словаря - при работе по заполнению вопросников Лингвистического атласа вепского языка. В совещании участвовали специалисты в области вепского языка докт. филол. наук Зайцева Н.Г., докт. филол. наук, профессор Муллонен И.И., докт. филол. наук Мызников С.А., канд. филол. наук Бродский И.В.

В течение 2012 г. результаты работы по проекту, ее ключевые моменты и проблемы излагались на международных и региональных научных форумах:

Зайцева Н.Г. Вепские причитания в фокусе корпусной лингвистики и лингвофольклористики // XLI Международная филологическая конференция, г. Санкт-Петербург, 26-31.03, 2012 г. Секция «Уралистика».

Зайцева Н.Г. Вепский язык в контексте корпусной лингвистики // Конференция «Бубриховские чтения», Петрозаводск, 18.10.2012.

Зайцева Н.Г. Вепский язык в фокусе современных лингвистических исследований // Межрегиональная практическая конференция «Вепсы: история и современность: к 1150-летию зарождения российской государственности. С. Винницы Подпорожского р., Ленинградской обл., 10.06.2012.

Крижановский А. А., Луговая Н. Б., Круглов В. М. Извлечение и анализ дат произведений в корпусе цитат онлайн-словаря // Информационные технологии и письменное наследие: материалы VI междунар. науч. конф. E1'Manuscript-12 (Петрозаводск, 3-8 сентября 2012) / отв. ред. В.А.Баранов, А.Г.Варфоломеев. – Петрозаводск; Ижевск, 2012. – 328 с. – С.

137—142. ISBN 978-5-8021-1402- <http://textualheritage.org/content/view/398/188/lang,russian/>

6. Общее число опубликованных в 2012 г. по проекту работ: 3

- 6.1. количество монографий: 1 словарь, 1 сборник текстов, выполненные, в том числе, при поддержке программы
- 6.2. количество сборников статей
- 6.3. количество статей: 2 (1 опубликована, 1 находится в печати)

7. Список опубликованных монографий и сборников статей, с полным указанием выходных данных, объема /в п.л. и количество стр./, а также их краткие аннотации (до 0,5 стр.)

Зайцева Н.Г., Жукова О.Ю. Обернись-ка милой кукушечкой. Вепские причитания. Juminkeko (Финляндия)-Петрозаводск, 2012. (первое издание-март). 224 с. 13 п.л. Второе издание, исправленное и дополненное: октябрь-2012 г.. 228 с. 13, 5 п.л.

Книга является единственным на сегодняшний день изданием, открывающим научному миру и всем заинтересованным лицам в наиболее полном объеме вепские причитания различного характера (свадебные, похоронные, внеобрядовые). Публикуемые причитания вводятся в научный оборот впервые, представляя коллекцию записей Фонограммархива Института языка, литературы и истории КарНЦ РАН. В книге можно найти ответы на многие вопросы относительно вепских причитаний, являющихся важнейшим компонентом самобытной духовной культуры вепского народа, отражающих достаточно ярко и особенности его языка. Часть причитаний сборника (50 причитаний) размещена на сайте Корпуса вепского языка (<http://vepsian.krc.karelia.ru/text/>), представляя там специальный параллельный подкорпус причитаний, работа над которым осуществляется при поддержке программы «Корпусная лингвистика», 2012-2014. Сборник причитаний инициирован и подготовлен в русле работы над корпусом при поддержке программы «Корпусная лингвистика».

Зайцева Н.Г., Харитоновна Е.Е., Жукова О.Ю. Орфографический словарь вепского языка. 25.11 п.л. Издательство «Периодика». Петрозаводск, 2012. 25.11 п.л. 428 с.

«Орфографический словарь вепского языка» является на сегодняшний день самым крупным лексикографическим изданием: в нём представлено около 17 тысяч слов. Кроме сведений по правописанию, в словаре предусмотрен перевод всех лексем на русский язык, поскольку далеко не все пользователи обладают хорошими знаниями вепского языка. Словарные статьи, посвящённые именам (существительным, прилагательным, числительным), местоимениям и глаголам, обладают сведениями по всем типам основ. Глагольные статьи сопровождаются формами причастий (актива: презен-

са и имперфекта, пассива), а также некоторыми данными по глагольному управлению. В конце словаря приложен краткий грамматический очерк вепсского языка и таблицы склонения и спряжения. Словарные статьи нового издания впервые готовились при использовании текстовых материалов и материалов электронного словаря Корпуса вепсского языка, использование в работе которого оказало существенную помощь при формировании орфограмм младописьменного вепсского языка, что отмечено и в предисловии к словарю.

8. Список опубликованных по проекту статей (объемом не менее 1 п.л.):

Зайцева Н.Г. Вепсские причитания в фокусе корпусной лингвистики и лингвофольклористики // Материалы XLI международной филологической конференции. Уралистика. 26-31 марта 2012 г. Санкт-Петербург, 2012. С. 16-26 (1 п.л.).

Смирнов А.В., Круглов В.М., Крижановский А.А., Луговая Н.Б., Карпов А.А., Кипяткова И.С. Количественный анализ лексики русского WordNet и викисловарей // Труды СПИИ-РАН. План издания 2012 г. Статья находится в издательстве.

9. Список книг, сданных в печать или поданных на издательские гранты (указать ожидаемое время издания, объем в а.л.)

10. Экспедиции, проведенные в рамках проекта (регион, руководитель, сроки, тематика исследований, полученные результаты, их значимость – до 2 стр.)

11. Конференции, организованные в рамках проекта (название, место и сроки проведения, обсуждаемые проблемы, результаты)

12. Важнейшие научные результаты работы по проекту (ок. 0,5 стр. для публикации на сайте Программы):

Главная задача проекта «Корпус вепсского языка: пополнение и развитие электронного ресурса», решаемая в течение 2012 г.г., □ реструктуризация созданного впервые за всю историю исследования языка вепсов электронного ресурса вепсского языка «Корпус вепсского языка» (<http://vepsian.krc.karelia.ru/text>), который значительно пополнен и содержит на сегодняшний день 800 текстов на 270 тысяч словоупотреблений, а также электронный словарь на 3190 лемм и 8560 словоформ. В результате реструктуризации «Корпус вепсского языка» стал содержать 5 подкорпусов: диалектной вепсской речи; младописьменных

вепских текстов; переводных текстов на младописьменном вепском языке (тексты перевода Нового Завета); параллельные подкорпусы вепских причитаний и вепских народных сказок, которые благодаря русскоязычным переводам дают возможность многим пользователям более свободно ориентироваться в текстовом и словарном материале, способствуют наиболее точному его поиску.

Результат работы в проекте сказался в издании в 2012 году таких значимых, единственных на сегодняшний день в научном мире трудов как «Сборник вепских причитаний» (13 п.л. – в течение 2012 года состоялось два издания сборника) и «Орфографического словаря вепского языка» (25.11 п.л.), подготовке которых способствовали напрямую электронные материалы Корпуса вепского языка.

13. Наиболее значимый научный результат проекта (5-6 строк для сводного отчета в Президиум РАН)

В результате реструктуризации Корпус вепского языка включает в себя пять подкорпусов (диалектный, младописьменный, переводов, причитаний, сказок), одни из которых – подкорпус вепских народных сказок с их переводами на русский язык (55 сказок) – подготовлен в 2012 году. Материалы корпуса способствовали разработке и публикации в 2012 г. двух важнейших научных работ: «Вепские причитаний» - на сегодняшний день единственного издания данного жанра, открывающего самобытную духовную культуру вепсов, и «Орфографического словаря младописьменного вепского языка», изданного впервые за период функционирования вепской письменности.

14. Краткий финансовый отчет за 2012 г. (основные статьи расходов по проекту, сумма) –

15. Запрашиваемы на 2013 г. объем финансирования, с кратким обоснованием расходов:

16. Краткое обоснование научных работ на 2013 г., ожидаемые результаты – до 1 стр. и заполненная форма 2 (см. ниже).

Планируемое содержание работ на 2013 г.

1. Планируется дальнейшая реструктуризация электронного ресурса вепского языка.
2. Пополнение корпуса новыми текстами: 100 текстов различного характера;
3. Дополнение подкорпуса вепских причитаний звучащей составляющей (25 образцов- приложений);
4. Лемматизация и грамматическая разметка: 500 лемм и 1000 словоформ;
5. Переводы лемм на русский и английский языки;
6. Продолжение структурирования корпуса; урегулирование систем поиска;
7. В течение года предполагается изменение структуры базы данных и функциональности сайта. Работа будет вестись в двух направлениях:

Электронный словарь:

- 1). Добавление специальных таблиц и полей в базу данных *словаря* для разграничения омонимов и значений слов для реализации в дальнейшем семантического поиска по корпусу.
- 2). Разработка возможности поиска слов по обратному словарю.

Корпус текстов:

- 1). Преобразование структуры сайта – объединение поиска по текстам и поиска по подкорпусам с целью создания единой поисковой формы с возможностью выбора подкорпуса и указания дополнительных поисковых параметров (диалект, говор, тип причитаний и т.д.).
- 2). Реализация технической возможности для выравнивания текстов:
 - при указании номера стиха в подкорпусе переводных вепскоязычных текстов Нового Завета и номера стиха в Библии;
 - тексты сказок и параллельный русскоязычный перевод.
- 3). Проектирование HTML-страницы для одновременного представления аудиофайла и текста с параллельным переводом.

Подпись руководителя проекта:

Зайцева Н.Г.

Форма 2

№	<p>Название проекта: «Корпус вепского языка: пополнение и развитие электронного ресурса»</p>	<p>Организация-исполнитель и учреждения-соисполнители: Институт языка литературы и истории Карельского научного центра РАН</p>	<p>Руководитель проекта Зайцева Н.Г., д. филол. наук Исполнители проекта: 1) Научный консультант: Муллонен Ирма Ивановна, директор ИЯЛИ КарНЦ РАН, д. филол. наук, профессор. 2) Шибанова Нина Леонидовна, главный специалист по информационным технологиям ИЯЛИ КарНЦ РАН. 3) Жукова Ольга Юрьевна, мл. научн. сотр. сектора языкознания ИЯЛИ КарНЦ РАН, канд. филол. наук. 4). Герасимов Федор Сергеевич, техник-оператор Фонограммархива ИЯЛИ КарНЦ РАН. 5) Крижановской Андрей Анатольевич, канд. техн. наук, ст. научн. сотр. лаборатории информационных компьютерных технологий Института прикладных математических исследований КарНЦ РАН</p>	<p>Запрашиваемый объем финансирования на 2013 г. (тыс. руб.): 350 тыс. рублей</p>	<ol style="list-style-type: none"> 1. Пополнение корпуса новыми текстами: 100 текстов различного характера; 2. Дополнение подкорпуса вепских причитаний звучными причитаниями (25 приложений); 3. Лемматизация: 500 лемм и 1000 словоформ; переводы лемм на русский и английский языки; 4. Продолжение структурирования корпуса; создание единой поисковой формы с возможностью выбора подкорпуса и указания дополнительных поисковых параметров (диалект, говор, тип причитаний, перевод и т.д.). 5. Продолжение работы в параллельных подкорпусах причитаний и сказок: решение проблем «выравнивания» текстов и переводов. 6. Создание HTML-страницы для одновременного представления аудиофайла и текста с параллельным переводом.
---	---	--	---	--	---