

Титульный лист отчета о работе в 2012 г.
по **Программе фундаментальных исследований Президиума РАН**
«Традиции и инновации в истории и культуре»

Номер и название направления Программы <u>Корпусная лингвистика</u> Создание и развитие корпусных ресурсов по языкам народов России.	
Название проекта Корпуса языков Дальнего Востока	
Научный руководитель проекта (ФИО полностью, уч. ст.) Перехвальская Елена Всеволодовна, доктор филологических наук	
E-mail, телефон, факс (с кодом города) почтовый адрес руководителя проекта	elenap96@gmail.com
Полное и краткое название организации – адресата финансирования Институт лингвистических исследований РАН ИЛИ РАН	ФИО (полностью) руководителя организации – адресата финансирования Казанский Николай Николаевич
	ФИО (полностью) главного бухгалтера организации – адресата финансирования 199053 Санкт-Петербург, Тучков пер., 9 тел. (812)3281611 факс (812)3284611 dir@iling.nw.ru; http://iling.spb.ru
	Телефон, факс (с кодом города), E-mail организации – адресата финансирования Белозерова Надежда Павловна тел. (812)3282155; факс (812)3284611 buch@iling.nw.ru
Год начала – год окончания проекта	2012 - 2014
Объем финансирования, полученного в 2012 г.	Объем финансирования, запрашиваемый на 2013 г.
Исполнители (ФИО, уч. ст.)	Оскольская Софья Алексеевна
	Пупынина Мария Юрьевна
Дата сдачи отчета 20.11.2012	Подпись руководителя проекта:

УТВЕРЖДАЮ

**Координатор Программы
фундаментальных исследований Президиума РАН
«Традиции и инновации в истории и культуре»
«__» _____ 2012 г.**

1. Название направления **Создание и развитие корпусных ресурсов по языкам народов России**
2. Название проекта **Корпуса языков Дальнего Востока**
3. Руководитель проекта (ФИО *полностью*, ученая степень, должность, место работы) **Перехвальская Елена Всеволодовна, доктор филологических наук, ведущий научный сотрудник Отдела языков России ИЛИ РАН**
4. Основные участники проекта (ФИО *полностью*, ученая степень, должность, место работы)
Оскольская Софья Алексеевна, младший научный сотрудник Отдела языков России ИЛИ РАН;
Пупынина Мария Юрьевна, младший научный сотрудник Отдела языков России ИЛИ РАН.
5. Содержание фактически проделанной работы, полученные результаты, их новизна, научная и практическая значимость, актуальность, соответствие мировому научному уровню (ок. 5 стр.)

Удэгейский язык.

Продолжалась работа по пополнению корпуса текстов на удэгейском языке, который состоит из двух типов текстов:

- ранее опубликованные тексты, работа над которыми состоит в том, что производится их компьютерный набор с дальнейшим глоссированием с использованием программы Toolbox;
- тексты из рукописного архива Е.Н.Баскаковой 1926 г., ранее не публиковавшиеся;
- тексты собранные в ходе лингвистических экспедиций 2006-2011 гг., которые не были опубликованы.

В 2012 г. работа велась со всеми тремя типами текстов.

Были полностью сканированы, распознаны и сличены с печатным текстом фольклорные произведения, опубликованные в 1998 г. в Новосибирске (Фольклор удэгейцев, под ред. М.Хасановой и М. Симонова). В настоящее время создана полная электронная версия этих текстов, из которых 15 глоссированы.

Большая работа проводилась по анализу и переводу рукописного фонда Е.Н. Баскаковой. К настоящему моменту произведен полный компьютерный набор текстов из V и VI тетрадей, а также 9 текстов из первой тетра-

ди, что составляет примерно 70% текстов архива. 21 текст оггоссирован с помощью программы Toolbox.

Проводилось транскрибирование, компьютерный набор, перевод на русский язык текстов, записанных в последние годы в ходе лингвистических экспедиций. Это тексты разного содержания, записанные у представителей самого старшего поколения удэгейцев (более 80 лет). Проводилось также транскрибирование текстов, записанных у более молодых носителей – это тексты, демонстрирующие значительное количество ситуаций переключения кодов.

В ходе глоссирования был значительно увеличен вспомогательный удэгейский словарь, который в настоящее время насчитывает 5077 вхождений.

Нанайский язык

В 2012 году продолжалась работа над пополнением корпуса нанайского языка. Работа велась в нескольких направлениях: 1) сбор новых текстов, 2) расшифровка и глоссирование имеющихся текстов, 3) создание грамматического анализатора.

Сбор новых текстов.

Большая часть новых текстов была записана в ходе экспедиции в Хабаровский край в августе 2012 г. Сбором новых текстов занимались С.А. Оскольская и мл. науч. сотрудник ИРЯ им. В.В. Виноградова Н.М. Стойнова. Во время экспедиции было записано 4 ч 59 мин 47 с устных текстов на нанайском языке, из которых около 5 мин – тексты на горинском говоре нижнеамурского диалекта (записаны в с. Кондон и с. Харпичан). Остальные тексты представляют найхинский говор среднеамурского диалекта (записи сделаны в с. Даерга, с. Найхин, пос. Троицкое, с. Джари, г. Хабаровск). Кроме того, около 20 мин устных текстов были записаны С. А. Оскольской в октябре в Петербурге при работе с носителем из с. Найхин. Все тексты представляют собой рассказы этнографического характера (рецепты нанайской кухни, рассказы о костюмах, о лекарствах, о традициях), нанайские легенды, автобиографические рассказы.

Расшифровка и глоссирование имеющихся текстов.

В ходе экспедиции в Хабаровский край в августе 2012 г. было расшифровано 49 мин 31 с устных текстов на среднеамурском диалекте и 1 ч 9 мин 55 с на нижнеамурском диалекте нанайского языка. Все тексты расшифровывались вместе с носителями соответствующих диалектов в программе ELAN 4.4.1. Работа представляла собой дословную запись устного текста на нанайском языке, синхронизация его со звуковой дорожкой и перевод на русский язык. Глоссирование текстов проходило уже после экспедиции. Общая сумма расшифрованных текстов в 2012 г. — 8400 словоупотреблений.

Большая часть расшифрованных текстов была записана в августе 2011 г. во время экспедиции в Хабаровский край. Часть разобранных текстов была записана в экспедиции 2012 г. Разобранные тексты представляют собой нанайские сказки и предания, рассказы этнографического характера, автобиографические рассказы.

Создание грамматического анализатора.

Начата работа над созданием грамматического анализатора для текстов на нанайском языке. Данный грамматический анализатор позволит делать морфологическую разметку как относительно многочисленных текстов, опубликованных на нанайском языке в советский период, так и текстов, записанных в ходе экспедиций последних лет. В основе анализатора лежит «Нанайско-русский словарь» С. Н. Оненко (1980), а также грамматика. Работа над грамматической частью анализатора будет вестись в 2013 году С. А. Оскольской и Н. М. Стойновой на основе «Грамматики нанайского языка» В. А. Аврорина (1959, 1961) и на основе собственных исследований. На данный момент ведется работа над распознаванием и форматированием словаря.

Чукотский язык.

Корпус чукотского языка включает:

- литературу на чукотском языке (как аутентичную, так и переводную);
- собранные во время экспедиций тексты;
- расшифрованные аудиозаписи радиопередач на чукотском языке.

Раздел корпуса, включающий уже изданные тексты, был существенно пополнен: к настоящему времени отсканировано двенадцать печатных изданий объемом от 30 до 318 страниц, в том числе шесть произведений художественной литературы, четыре книги сказок и два переводных издания. Тексты сохранялись в формате *pdf. Тексты четырёх книг (объемом 318, 80, 109 и 100 страниц) после автоматического распознавания были скопированы MS Word с сохранением постраничной разбивки и сверены с оригиналом.

В экспедиции в Чукотский автономный округ, проведённой в июле-августе 2012 года, были записаны, расшифрованы и глоссированы новые тексты на чукотском языке. Всего было записано двенадцать текстов продолжительностью около 2-х часов. В частности, были записаны тексты на малоизученном хатырском (южном) говоре чукотского языка. Пополнение корпуса материалами на различных говорах чукотского языка представляется особенно актуальным ввиду того, что чукотский язык находится под угрозой исчезновения, причём южный говор имеет наименьшую степень сохранности.

В с. Усть-Белая были сделаны записи речи женщины-монолингва, владеющей только чукотским языком.

Кроме того, велась работа над тремя текстами, записанными и расшифрованными ранее: уточнялись значения некоторых слов и морфем, упорядочивалась система глоссирования.

В ходе экспедиции были получены в дар аудиозаписи интервью с носителями чукотского языка, сделанные по заказу чукотского радио, в том числе интервью с оленеводом хатырской оленеводческой бригады на хатырском (южном) говоре.

Корякский язык

Корпус корякских текстов состоит только из уже опубликованных материалов. Для корякского корпуса были отсканированы и распознаны пять книг

на корякском языке, все они относятся к переводной литературе. Одно произведение объемом 86 страниц было после распознавания сверено с оригиналом.

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

6.1. количество монографий

6.2. количество сборников статей

6.3. количество статей – 3.

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.)

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

1. Волков О.С., Даниэль М.А., Пупынина М.Ю., Рыжова Д.А. К характеристике аспектуальной системы чукотско-камчатских языков / Acta linguistica Petropolitana: Труды Института лингвистических исследований РАН, том VIII, часть 2. Исследования по теории грамматики. Выпуск 6: Типология аспектуальных систем и категорий. Санкт-Петербург: Наука, 2012. С. 396-457.

2. Перехвальская Е.В. Выражение аспектуальных значений в удэгейском языке / Acta linguistica Petropolitana: Труды Института лингвистических исследований РАН, том VIII, часть 2. Исследования по теории грамматики. Выпуск 6: Типология аспектуальных систем и категорий. Санкт-Петербург: Наука, 2012. С. 361-395.

3. Перехвальская Е.В. Рукописный архив Е.Н. Баскаковой 1936 г. (тексты на самаргинском диалекте удэгейского языка) / Вопросы языкознания 2012, № 5. С.

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.)

1. Чукотский автономный округ (г. Анадырь, с. Усть-Белая, с. Снежное, с. Хатырка), июль-август 2012 года. Запись текстов на чукотском языке, их обработка.

2. Хабаровский и Приморский края (Хабаровский край: гор. Хабаровск, пос. Найхин, Дада, Троицкое, с. Кондон, Харпичан, Даерга; Приморский край: пос. Верхний Перевал, Красный Яр), август 2012 г. Сбор материалов по нанайскому языку, запись текстов.

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)

На собранном корпусном материале было проведено исследование аспектуальных систем удэгейского языка, а также сравнительное исследование аспектуальных систем чукотско-камчатских языков. В отношении удэгейского языка было показано, что аспектуальные значения в этом языке выражаются, как грамматически — кумулятивно с выражением времени теми или иными формами глагола («глагольными временами»), так и морфологически — словообразовательными аффиксами, несущими аспектуальное значение. , что основными средствами передачи аспектуальных значения являются показатели глагольной деривации. Последние в разной степени грамматикализованы. Некоторые из них не могут встречаться внутри одной словоформы. Два показателя (имперфектив и сингулятив), присоединяясь к исходным основам, образуют семантические пары, противопоставленные как «процесс» — «законченное действие». Одновременно в языке присутствует грамматическая оппозиция претерит — перфект. Формы перфекта выражают достаточно типичные для данной формы аспектуальные значения.

Для чукотско-камчатских языков были рассмотрены формы, составляющие ядро видовременной системы этой группы языков. Эти формы противопоставлены, во-первых, по морфологическому признаку личности ~ адъективности. Первые используют специальные глагольные согласовательные показатели, вторые — структурно (но необязательно формально) тождественны адъективам. И личные, и адъективные формы противопоставлены по перфективности ~ имперфективности, однако в личных формах перфектив немаркирован, а имперфектив выражается особым суффиксом, в то время как у адъективных форм аспектуальные граммемы выражаются самим «адъективным» циркумфиксом.

Проведенные исследования отражены в двух публикациях.

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

На собранном корпусном материале было проведено исследование аспектуальных систем удэгейского языка, а также сравнительное исследование аспектуальных систем чукотско-камчатских языков. Прослежены возможные пути возникновения данных систем.

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

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

16. Краткое обоснование итогового этапа научных работ 2013 г., ожидаемые результаты – 1 стр. и заполненная форма 2 (см. отдельный файл).

В 2013 г. планируется:

Удэгейский язык:

- 1) Будет продолжена работа по глоссированию опубликованных текстов, будут глоссированы тексты, представляющие собой переводы с русского языка.
- 2) Планируется завершить компьютерный набор, перевод на русский язык и глоссирование текстов из архива Е.Н. Баскаковой.
- 3) Будет проводиться синхронизация аудиозаписей и глоссированных текстов с использованием программы Elan 4.4.1. Предполагается начать такую обработку для текстов, записанных в ходе экспедиций 2006-2011 гг, которые ранее не публиковались.
- 4) На основе корпусных материалов будет проведено исследование функций пространственных падежей удэгейского языка.

Нанайский язык:

- 1) продолжать работу над грамматическим анализатором для нанайского языка: планируется завершить работу над распознаванием и форматированием словаря, а также над грамматической частью анализатора.
- 2) Планируется расшифровать и отглоссировать около 120 мин текстов на разных диалектах нанайского языка.
- 3) Будет продолжено сканирование, распознавание и глоссирование текстов, опубликованных в советский период, как оригинальных текстов, так и переводов (всего около 10 тыс. словоупотреблений).

Чукотский язык

- 1) Будет проведена работа по транскрибированию, переводу и глоссированию аудиозаписей, сделанных по заказу чукотского радио на хатырском говоре.
- 2) Планируется расшифровать и отглоссировать тексты, записанные в 2012 году в ходе лингвистической экспедиции в Чукотский автономный округ.
- 3) Будет продолжена работа по сканированию, распознаванию и глоссированию текстов, опубликованных в послевоенный период, как оригинальных текстов, так и переводов с русского языка.

Корякский язык

- 1) Будет продолжена работа по сканированию и распознаванию опубликованных текстов на корякском языке.

Подпись руководителя проекта

Форма 2
Планируемое содержание работ на 2013 г.

№	Название проекта	Организация-исполнитель и учреждения-соисполнители	Руководитель проекта (+ кол-во исполнителей)	объем финансирования на 2013 г. (тыс. руб.)	Ожидаемые в 2013 г. результаты
	Корпуса языков Дальнего Востока	Институт лингвистических исследований РАН	Перехвальская Елена Всеволодовна + 2 исполнителя		<p>Удэгейский язык: гlossирование опубликованных текстов (переводы с русского языка); завершение обработки текстов из архива Е.Н. Баскаковой; синхронизация аудиозаписей и текстов с использованием программы Elan; исследование функций пространственных падежей удэгейского языка.</p> <p>Нанайский язык: работа над грамматическим анализатором для нанайского языка: планируется завершить работу над распознаванием и форматированием словаря, а также над грамматической частью анализатора; расшифровка и glossирование около 120 мин текстов на диалектах нанайского языка; распознавание и glossирование текстов, опубликованных в советский период.</p> <p>Чукотский язык транскрибирование, перевод и glossирование аудиозаписей на хатырском говоре; расшифровка и glossирование текстов, записанных в 2012 году в ходе экспедиции; сканирование, распознавание и glossирование текстов, опубликованных в послевоенный период.</p> <p>Корякский язык сканирование и распознавание опубликованных текстов на корякском языке.</p>