

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

Номер и название раздела программы 3. Создание и развитие корпусных ресурсов по языкам народов России	
Название проекта Создание корпусов на диалектах языков Поволжья	
Научный руководитель проекта (ФИО полностью, уч. ст.) Наталья Львовна Шибасова, к.ф.н., и.о. м.н.с. Института языкознания РАН	
Е-mail, телефон, факс (с кодом города) почтовый адрес руководителя проекта	nshibasova@yandex.ru
Полное и краткое название организации – адресата финансирования ФГБУН Институт языкознания (ИЯз РАН)	ФИО (полностью) руководителя организации – адресата финансирования Алпатов Владимир Михайлович
	ФИО (полностью) главного бухгалтера организации — адресата финансирования Клезович Ирина Ивановна
	Телефон, факс (с кодом города), Е-mail организации – адресата финансирования тел. (495)690-35-85, факс (495)690-05-28 iling@iling-ran.ru
Год начала – год окончания проекта	2012—2014
Объем финансирования, полученного в 2012 г.	Объем финансирования, запрашиваемый на 2013 г.
Исполнители (ФИО, уч. ст.)	Мудрак О. А. , д.ф.н., Институт языкознания РАН
	Девяткина Е. М., к.ф.н., Институт языкознания РАН
	Минлос Ф. Р., к.ф.н., Институт славяноведения РАН
	Якунин М. В., Институт языкознания РАН
	Беляев О. И., Институт языкознания РАН
	Гращенков Павел Валерьевич, к.ф.н., ИВРАН
	Рамазанова Д.Б., д.ф.н., ИЯЛИ АН РТ
	Иванова Н.Г., доц., ЧРИО
	Каримова Р., н.с. ИИЯЛ УНЦ РАН
	Сердобольская Н. В., к.ф.н., РГГУ
	Толдова С. Ю., к.ф.н., РГГУ
	Усачева М. Н., МГУ
	Шматова М. С., МГУ
Дата сдачи отчета 20 ноября 2012 г.	Подпись руководителя проекта:

УТВЕРЖДАЕМ

Координатор Программы

акад. Вяч. Вс. Иванов

Координатор Программы

чл-корр. РАН В.А. Плунгян

«__» _____ 2012 г.

1. Название направления

Создание и развитие корпусных ресурсов по языкам народов России

2. Название проекта

Создание корпусов на диалектах языков Поволжья

3. Руководитель проекта (ФИО полностью, ученая степень, должность, место работы)

Наталья Львовна Шибасова, к.ф.н., и.о м.н.с. Института языкознания РАН

4. Основные участники проекта (ФИО полностью, ученая степень, должность, место работы)

Мудрак Олег Алексеевич, д.ф.н., в.н.с. Института языкознания РАН

Девяткина Екатерина Михайловна, н.с. Института языкознания РАН

Минлос Филипп Робертович, к.ф.н., н.с. Института славяноведения РАН

Рамазанова Дария Байрамовна, д.ф.н., Институт языка, литературы и истории АН РТ

Якунин Михаил Викторович, ст. лаб., Институт языкознания РАН

Беляев Олег Игоревич., м.н.с. Институт языкознания РАН

Гращенко Павел Валерьевич, к.ф.н., н.с. Институт востоковедения РАН

Иванова Надежда Григорьевна, доц., ЧРИО

Каримова Рамиля, н.с. ИИЯЛ УНЦ РАН

Сердобольская Наталья Вадимовна, Российский государственный гуманитарный университет, доцент

Толдова Светлана Юрьевна, к.ф.н., Российский государственный гуманитарный университет, доцент

Усачева Мария Николаевна, Московский государственный университет им. М.В. Ломоносова, аспирант

Шматова Мария Сергеевна, Московский государственный университет им. М.В. Ломоносова, аспирант

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

1. В 2012 году была запущена полнофункциональная версия корпусного ресурса, который находится на сайте Института языкознания РАН: corpora.iling-ran.ru. Для этого был куплен и установлен сервер, на которые было перенесено написанное ранее программное обеспечение. На этот сайт были выложены все корпуса, сделанные в ходе предыдущих корпусных проектов — как участниками данного проекта, так и другими исследователями. Сейчас на нем уже находится 516 текстов на 18 языках народов России и мира (в том числе на языках семей майя и западноатлантической), сайт продолжает пополняться.

2. В рамках подготовки корпусов на диалектах языков Поволжья созданы следующие корпуса:

- эрзя-мордовский (шокшинский диалект) — 2000 предложений (ок. 10000 слов);
- коми-зырянский (печорский диалект) — 1149 предложений (ок. 6000 слов);
- удмуртский (бесермянский диалект) — 2500 предложений (ок. 12500 слов);
- чувашский (северо-западный диалект) — расшифровка 5 часов текста (около 12000 слов).

Все тексты представляют собой полевые записи последних лет (после 2000 года). Жанр записей — бытовые рассказы, диалоги, небольшое число фольклорных текстов. Ни для одного из этих языков не существует других корпусов, по крайней мере, выложенных в открытый доступ.

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

- 6.1. количество монографий
- 6.2. количество сборников статей
- 6.3. количество статей

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.)

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.)

С 25 июля по 8 августа в д. Шамардан Юкаменского района республики Якутия находились студентки МГУ А. Смирнова и И. Зибер. Были записаны и разобраны тексты на бесермянском диалекте удмуртского языка; также были обработаны ранее записанные тексты.

Разобранные записи представляют собой монологические и диалогические тексты, порождённые как по просьбе исследователей (фольклорные канонические тексты,

а также тексты-импровизации на произвольные темы), так и спонтанно, в общении с другими носителями, что делает их максимально приближенными к речи живого повседневного общения. Среди них представлены бытовые диалоги, нарративы с элементами диалога и монологи. Тематически тексты довольно разнообразны. Это и автобиографический меморат, и рассуждения о современной жизни носителей, их обычаях, мировоззрении, понимании собственного национального своеобразия, самобытности языка, культуры и быта.

Общий объем разобранных текстов — около 1250 предложений. Они вошли в состав удмуртского корпуса и выложены на сайте.

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)

Запущена полнофункциональная версия корпусного ресурса corpora.iling-ran.ru. Для этого был куплен и установлен сервер, на который было перенесено написанное ранее программное обеспечение. На этот сайт были выложены все корпуса, сделанные в ходе предыдущих корпусных проектов — как участниками данного проекта, так и другими исследователями. Сейчас на нем уже находится 516 текстов на 18 языках народов России и мира (в том числе на языках семей майя и западноатлантической), сайт продолжает пополняться. Созданы новые корпуса на диалектах финно-угорских и тюркских языков Поволжья: эрзя-мордовский (шокшинский диалект) — 2000 предложений (ок. 10000 слов); коми-зырянский (печорский диалект) — 1149 предложений (ок. 6000 слов); удмуртский (бесермянский диалект) — 2500 предложений (ок. 12500 слов); чувашский (северо-западный диалект) — расшифровка 5 часов текста (около 12000 слов).

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

Создан общедоступный корпусной ресурс, размещенный на сайте Института языкознания РАН: corpora.iling-ran.ru; на ноябрь 2012 года на нем находилось более 500 текстов на 18 языках России и мира. Созданы новые корпуса на диалектах финно-угорских и тюркских языков Поволжья: эрзя-мордовский (шокшинский диалект) — 2000 предложений (ок. 10000 слов); коми-зырянский (печорский диалект) — 1149 предложений (ок. 6000 слов); удмуртский (бесермянский диалект) — 2500 предложений (ок. 12500 слов); чувашский (северо-западный диалект) — расшифровка 5 часов текста (около 12000 слов).

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

В 2013 году работа будет сосредоточена в основном на тех же языках, что в 2012 г., с добавлением марийского и татарского и, вероятно, с исключением бесермянского, поскольку все имеющиеся на сегодняшний день материалы уже введены в корпус. Предполагается сделать следующие объемы по языкам: коми-зырянский язык (печорский диалект) — увеличение корпуса на 5000 словоформ; луговой марийский язык (торьяльский диалект) — создание корпуса объемом 20000 словоформ; татарский язык — создание диалектного корпуса объемом около 15000 словоформ; чувашский язык (северо-западный диалект) — увеличение корпуса на 10000 словоформ.

Подпись руководителя проекта

Н.Л. Шибасова

Форма 2
Планируемое содержание работ на 2013 г.

№	Название проекта	Организация-исполнитель и учреждения-соисполнители	Руководитель проекта (+ кол-во исполнителей)	Запрашиваемый объем финансирования на 2013 г. (тыс. руб.)	Ожидаемые в 2013 г. результаты
	Создание корпусов на диалектах языков Поволжья	ИЯз РАН	Шибасова Н. Л. (+ 12 чел.)		<ul style="list-style-type: none"> — коми-зырянский язык (печорский диалект) — увеличение корпуса на 5000 словоформ — луговой марийский язык (торьяльский диалект) — создание корпуса объемом 20000 словоформ — татарский язык — создание диалектного корпуса объемом около 15000 словоформ — чувашский язык (северо-западный диалект) — увеличение корпуса на 10000 словоформ.