

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

Номер и название направления Программы Направление 3. Создание и развитие корпусных ресурсов по языкам народов России	
Название проекта Национальный корпус осетинского языка: расширение и развитие	
Научный руководитель проекта (ФИО полностью, уч. ст.) Выдрин Арсений Павлович, к.ф.н.	
E-mail, телефон, факс (с кодом города) почтовый адрес руководителя проекта	senjacom@gmail.com
Полное и краткое название организации – адресата финансирования Учреждение Российской Академии Наук Институт Лингвистических Исследований (ИЛИ РАН)	ФИО (полностью) руководителя организации – адресата финансирования Казанский Николай Николаевич
	ФИО (полностью) главного бухгалтера организации – адресата финансирования Белозерова Надежда Павловна
	Телефон, факс (с кодом города), E-mail организации – адресата финансирования 199053, Санкт-Петербург, Тучков пер., 9. Тел.: (812) 328-16-11, факс: (812) 328-46-11. Email: iliran@mail.ru
Год начала – год окончания проекта	2012—2014
Объем финансирования, полученного в 2012 г.	Объем финансирования, запрашиваемый на 2013 г.
Исполнители (ФИО, уч. ст.)	Мазурова Юлия Викторовна, к.ф.н., с.н.с.
	Беляев Олег Игоревич, м.н.с
	Дарчиева Мадина Владимировна, к.ф.н., н.с.
Дата сдачи отчета	Подпись руководителя проекта:

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

Национальный корпус осетинского языка: расширение и развитие

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

Выдрин Арсений Павлович, кандидат филологических наук, научный сотрудник Института лингвистических исследований РАН

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

Беляев Олег Игоревич, младший научный сотрудник Института языкознания РАН

Дарчиева Мадина Владимировна, кандидат филологических наук, научный сотрудник Северо-Осетинского института гуманитарных и социальных исследований им. В.И.Абаева Владикавказского научного центра и Правительства РСО–Алания

Мазурова Юлия Викторовна, кандидат филологических наук, старший научный сотрудник Института языкознания РАН

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

За время работы над проектом в 2012 г. на основе доступной литературы (Исаев М.И. Осетинские имена // Справочник личных имен народов РСФСР. М., 1965; Цогоев В.Г. Словарь осетинских личных имен. Орджоникидзе, 1990; Fritz S. Iranisches Personennamenbuch, Band III: Neuiranische Personennamen, Faszikel 3: Die Ossetischen Personennamen. Vienna, 2006) в систему автоматического разбора был добавлен словарь осетинских имен собственных. Участники проекта существенно расширили список словоизменяемых и словообразовательных морфем. В частности были добавлены следующие суффиксы: *-он* (ADJ), *-дзинад* (деривационный суффикс, использующийся для образования абстрактных имен), *-ад* (деривационный суффикс, использующийся для образования абстрактных имен), *-джын/-чын* (ADJ), *-хъом* (POT), *-аг* (отыменный суффикс), *-даер* (ADD) и многие другие. Осетинский язык обладает немалым количеством сложных слов, которые сильно затрудняют автоматический разбор. На данном этапе работы над усовершенствованием автоматического разбора осетинских текстов было принято решение включить в список морфем наиболее употребительные составляющие части сложных слов. А именно, были включены следующие языковые единицы, способные образовывать композиты: *гæн-* и ее деривационные формы (основа настоящего времени глагола *кæнын* 'делать', ср. *ахуыр-гæнæг* (учение-делать.PRS-PART) 'учитель'), *-афон* ('время', ср. *аходæн-афон* (завтрак-время) 'время завтрака'), *-хуыз* ('имеющий вид', ср. *айк-хуыз* (яйцо-вид) 'яйцевидный, яйцеобразный') и многие другие.

В результате проделанной работы процент автоматического разбора в корпусе осетинского языка повысился до 80 %.

Существенному улучшению подвергся поиск, куда было добавлено большое количество новых показателей. Начата работа по составлению подробной инструкции пользования поиском и корпусом.

Доработан вид автоматического разбора, который теперь доступен на двух языках, английском и русском.

Корпус находится в свободном доступе в сети Интернет по адресу <http://corpus.ossetic-studies.org>.

Проведенная в 2012 г. работа по существенному улучшению корпуса письменных текстов осетинского языка представляет собой значительный вклад как в осетиноведение, так и в иранистику в целом. Среди других иранских языков до сих пор не существует аналогов созданному осетинскому корпусу, как по объему включенных текстов, так и по степени подробности разбора словоформ и разработанной системе поиска. Уже в настоящий момент корпус является удобным средством получения материала по отдельным аспектам грамматики осетинского языка, открытым не только для осетиноведов и иранистов, но и для лингвистов-типологов.

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

6.1. количество монографий – 0

6.2. количество сборников статей – 0

6.3. количество статей — 5

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.) — 0

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

Belyaev O. Towards an anaphoric approach to Ossetic correlatives // *Typology of Languages of Europe and Northern and Central Asia*. Amsterdam: John Benjamins. Принято к публикации, 1 а.л.

Мазурова Ю.В. Иранские и кавказские черты в системе пространственных показателей осетинского языка // Материалы Первой конференции-школы «Проблемы языка: взгляд молодых ученых» (в печати) 1 а.л.

Выдрин А. П. Две конструкции возможности в осетинском языке // ВЯ 2013 (в печати) 1,5 а.л.

Выдрин А. П. Редкий случай грамматикализации модального значения в осетинском языке // *Acta Linguistica Petropolitana*. Труды ИЛИ РАН / отв. ред. Н. Н. Казанский, 2013 (в печати).

Выдрин А. П. Полисемия модальных языковых единиц и контексты их семантической однозначности // *Studia Typologica*, 2013. 1 а.л. (принято к печати)

9. Список книг, сданных в печать или поданных на издательские гранты (указать ожидаемое время издания, объем в а.л.) — 0

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

Республика Северная Осетия – Алания, руководитель – О.И. Беляев, сроки – 15.03.2012 – 31.03.2012, подбор письменных текстов на иронском диалекте осетинского языка, организация сканирования 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)

Существенно расширена система автоматической разметки: добавлен словарь осетинских имен собственных, включен целый ряд словоизменительных и словообразовательных морфем, а также наиболее употребительные составляющие осетинских композитов. В результате проделанной работы процент автоматического разбора в 2012 г. достиг 80 % (по сравнению с 75% в 2011 г.). Значительному улучшению подвергся поиск, куда было добавлено большое количество новых показателей. Доработан вид автоматического разбора, который теперь доступен на двух языках, английском и русском.

Корпус осетинского языка находится в свободном доступе в сети Интернет по адресу <http://corpus.ossetic-studies.org>.

На основе расширенного и усовершенствованного корпуса участниками проекта был проведен ряд исследований по отдельным аспектам грамматики осетинского языка. Результаты исследований представлены в серии докладов и статей, принятых к печати в 2013 году.

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

За год проделанной работы была существенно улучшена система автоматического разбора, позволившая увеличить процент разбора с 75% в 2011 г. до 80 % в 2012 г. Доработке подвергся также вид автоматического разбора, который теперь доступен на двух языках, английском и русском. Также была усовершенствована система поиска. Корпус находится в свободном доступе в сети Интернет по адресу <http://corpus.ossetic-studies.org>.

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).

Расширение национального корпуса осетинского языка до 10 млн. словоупотреблений. Развитие системы автоматической разметки и доведение процента разобранного материала после автоматической разметки всего корпуса как минимум до 85%. Расширение системы поиска. Размещение результатов работы на ресурсе www.corpus.ossetic-studies.org

Подпись руководителя проекта

А. Выдрин

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
	Национальный корпус осетинского языка: расширение и развитие	ИЛИ РАН	Выдрин А. П. Мазурова Ю.В. Дарчиева М. В. Беяев О. И.		Расширение национального корпуса осетинского языка до 10 млн. словоупотреблений. Развитие системы автоматической разметки и доведение процента разобранного материала после автоматической разметки всего корпуса как минимум до 85%. Расширение системы поиска. Размещение результатов работы на ресурсе www.corpus.ossetic-studies.org