

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

Номер и название направления Программы Корпусная лингвистика	
Название проекта Корпус среднерусских текстов XV—XVII вв.	
Научный руководитель проекта (ФИО полностью, уч. ст.) Молдован Александр Михайлович, академик	
Е-mail, телефон, факс (с кодом города) почтовый адрес руководителя проекта	8-495-695 26 60 alexandr.moldovan@gmail.com
Полное и краткое название организации – адресата финансирования Федеральное государственное бюджетное учреждение науки Институт русского языка им. В. В. Виноградова РАН ИРЯ РАН	ФИО (полностью) руководителя организации – адресата финансирования Молдован Александр Михайлович
	ФИО (полностью) главного бухгалтера организации – адресата финансирования Глебова Татьяна Николаевна
	Телефон, факс (с кодом города), Е-mail организации – адресата финансирования Телефон: (+7 495) 695-26-60 Факс: (+7 495) 695-26-03
Год начала – год окончания проекта	2012—2014
Объем финансирования, полученного в 2012 г.	Объем финансирования, запрашиваемый на 2013 г.
Исполнители (ФИО, уч. ст.)	Кузнецова Анастасия Валерьевна
	Израелян Анри Гарегинович
	Смирнова Екатерина Андреевна
Дата сдачи отчета 15 ноября 2012 г.	Подпись руководителя проекта:

Координатор Программы

акад. Вяч. Вс. Иванов

Координатор Программы

чл-корр. РАН В.А. Плунгян

«___»_____2012 г.

1. Название направления **2. Создание и развитие корпусных ресурсов по истории русского языка**
2. Название проекта **«Корпус среднерусских текстов XV—XVII вв.»**
3. Руководитель проекта (ФИО *полностью*, ученая степень, должность, место работы) **Молдован Александр Михайлович, академик**
4. Основные участники проекта (ФИО *полностью*, ученая степень, должность, место работы)
Кузнецова Анастасия Валерьевна, ИРЯ РАН, научный сотрудник;
Израелян Анри Гарегинович, ИРЯ РАН, аспирант;
к.ф.н. Смирнова Екатерина Андреевна, ИРЯ РАН, младший научный сотрудник.
5. Содержание фактически проделанной работы, полученные результаты, их новизна, научная и практическая значимость, актуальность, соответствие мировому научному уровню (до 5 стр.)
Обработаны и подготовлены для включения в Национальный корпус русского языка созданные до этого электронные копии текстов древнерусских памятников XIV—XVII вв.: Четыре летописца середины XVI – первой половины XVII в. (ПСРЛ, т. 34) (29 п. л.);
А. Лызлов. Скифская история (18,2 п. л.);
Путешествие стольника П. А. Толстого по Европе (1697–1699) (18,2 п. л.);
Посольские книги по связям России с Ногайской Ордой. 1551—1561 гг. (17 п.л.);
Рязанская писцовая приправочная книга конца XVI в. (14,6 п. л.);
Путешествия русских послов XVI–XVII вв. Статейные списки (15 п. л.);
Акты писцового дела (1644–1661 гг.) (13,8 п. л.);
Кунгурские акты XVII в. (12,5 п. л.);
Акты феодального землевладения. Т. 1 (18,75 п.л.);
Акты феодального землевладения. Т. 2 (29,5 п.л.);
Акты феодального землевладения. Т. 3 (23 п.л.);

Кормленая книга Костромской чети 1613–1627 гг. (РИБ, № 15) (11,5 п.л.);
Новгородская кабальная книга 1597 г. (РИБ, № 15) (3 п.л.);
Владенная выпись с чувашского судного дела 1672–1674 гг. (РИБ, № 15)
(1 п.л.);
Строенная книга города Воронежа 1670 г. 0,7 п.л. Приходная книга Нов-
городской четверти 1619/20 г. и Другие прходно-расходные книги
(13 п.л.);
Писцовая книга дворцовой ловецкой Рыбной слободы 1674, 1675 и 1676
годов (3,65 п. л.);
Наказ Афанасию Филипповичу Пашкову на воеводство в Даурской зем-
ле. 1655 г. (РИБ, № 15) (2 п.л.);
Наказ Пазухиным, посланным в Бухару, Балх и Юргенч. 1669 г. (РИБ,
№ 15) (4,6 п.л.);
Грамотки XVII – начала XVIII вв. (вторая часть, с № 343) 8,5 п. л.

В ходе проведенных в 2012 г. работ подготовлены тексты на рус-
ском языке разных жанров XIV-XVII вв. для исторического корпуса об-
щим объемом почти 2 млн. словоупотреблений.

Некоторые тексты сканировались, распознавались (программой Fine
Reader 8.0), вычитывались и обрабатывались для размещения в корпусе
(разбивались на файлы, снабжались необходимыми метаданными: ав-
тор, годы написания текста, годы копирования текста, жанр, языковая
разновидность, издание, номер в издании, сведения о архиве).

В некоторых случаях были использованы отчасти готовые резуль-
таты работы (например, уже отсканированные источники).

На данный момент этот корпус может быть использован для поис-
ка конкретных словоформ (или их сочетаний). В будущем этот корпус
позволит разработать компьютерную морфологию древнерусского язы-
ка и, соответственно, поиск лексем и поиск по заданным морфологиче-
ским характеристикам. Использование обработанных и централизован-
ных корпусных данных поднимает работу над исторической лексиколо-
гией и грамматикой на качественно новый уровень.

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 — 0

7. Список опубликованных монографий и сборников статей, с полным указа-
нием 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.) — 0

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.) — **не предусмотрено**
11. Конференции, организованные в рамках проекта (название, место и сроки проведения, обсуждаемые проблемы, результаты) — **не предусмотрено**
12. Важнейшие научные результаты работы по проекту (ок. 0,5 стр. для публикации на сайте Программы)
В ходе проведенных в 2012 г. работ подготовлены тексты на русском языке разных жанров XIV-XVII вв. для исторического корпуса общим объемом почти 2 млн. словоупотреблений. На данный момент этот корпус может быть использован для поиска конкретных словоформ (или их сочетаний). В будущем этот корпус позволит разработать компьютерную морфологию старорусского языка и, соответственно, поиск лексем и поиск по заданным морфологическим характеристикам. Использование обработанных и централизованных корпусных данных поднимает работу над исторической лексикологией и грамматикой на качественно новый уровень
13. Наиболее значимый научный результат проекта (5-6 строк для сводного отчета в Президиум РАН)
В ходе проведенных в 2012 г. работ подготовлены тексты на русском языке разных жанров XIV-XVII вв. для исторического корпуса общим объемом почти 2 млн. словоупотреблений. На данный момент этот корпус может быть использован для поиска конкретных словоформ (или их сочетаний).
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). **Сканирование, распознавание и вычитка памятников XV—XVII вв. Подготовка текстов к вывеске на сайте НКРЯ. Общий объем около 2 млн. словоупотреблений. Запрашиваемая сумма пойдет на выплату заработной платы исполнителям проекта.**

Подпись руководителя проекта

А.М. Молдован

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
	Корпус среднерусских текстов XV—XVII вв.	ИРЯ РАН	Молдован А.М. +3		Сканирование, распознавание и вычитка памятников XV—XVII вв. Подготовка текстов к вывеске на сайте НКРЯ. Общий объем около 2 млн. словоупотреблений.