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

<table>
<tr><td colspan="2">Номер и название раздела программы Направление 1. Создание и развитие корпусных ресурсов по современному русскому языку</td></tr>
<tr><td colspan="2">Название проекта
Пополнение базы текстов XVIII и XIX веков в Национальном корпусе русского языка</td></tr>
<tr><td colspan="2">Научный руководитель проекта (ФИО полностью, уч. ст.)
Дмитренко Сергей Юрьевич, кандидат филологических наук</td></tr>
<tr><td>E-mail*, телефон, факс почтовый адрес руководителя проекта</td><td>dmitrserg@mail.ru</td></tr>
<tr><td rowspan="4">Полное и краткое название организации – адресата финансирования

Учреждение Российской академии наук Институт лингвистических исследований РАН</td><td>ФИО руководителя организации – адресата финансирования
Казанский Николай Николаевич, академик РАН</td></tr>
<tr><td>Почтовый адрес, телефон, факс (с кодом города), E-mail организации – адресата финансирования
Тучков переулок д. 9, 199053, Санкт-Петербург; тел.: (812)3281611; факс: (812)3284611; dir@iling.spb.ru</td></tr>
<tr><td>ФИО (полностью) главного бухгалтера организации – адресата финансирования
Белозерова Надежда Павловна</td></tr>
<tr><td>Телефон, факс (с кодом города), E-mail главного бухгалтера организации – адресата финансирования
тел.: (812)3282155; факс: (812)3284611; dir@iling.spb.ru</td></tr>
<tr><td>Год начала – год окончания проекта</td><td>2012—2014</td></tr>
<tr><td>Объем финансирования, полученного в 2012 г.</td><td>Объем финансирования, запрашиваемый на 2013 г.</td></tr>
<tr><td rowspan="3">Исполнители (ФИО, уч. ст., Институт)</td><td>Круглов В. М., дфн, ИЛИ РАН</td></tr>
<tr><td>Кузнецова И. Е., ИЛИ РАН</td></tr>
<tr><td>Оскольская С. А., ИЛИ РАН</td></tr>
<tr><td>Дата сдачи отчета
20.11.2012</td><td>Подпись руководителя проекта</td></tr>
</table>

Форма 1

УТВЕРЖДАЕМ

Координатор Программы акад. Вяч. Вс. Иванов

Координатор Программы чл-корр. РАН В.А. Плунгян

«___»_____2012 г.

1. Название направления

Направление 1. Создание и развитие корпусных ресурсов по современному русскому языку

2. Название проекта

Пополнение базы текстов XVIII и XIX веков в Национальном корпусе русского языка

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

Дмитренко Сергей Юрьевич, к.ф.н., зам. директора ИЛИ РАН;

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

Круглов Василий Михайлович, дфн, зав. Лабораторией информационных лингвистических исследований;

Кузнецова Ирина Евгеньевна, нс ИЛИ РАН;

Оскольская София Алексеевна, лаборант ИЛИ РАН.

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

Диахроническая часть НКРЯ *объединяет несколько подкорпусов, среди которых в качестве отдельного выделены тексты XVIII и XIX вв. Структура корпуса и разная степень разработанности его отдельных фрагментов обусловила постановку задач, которые были решены в рамках проекта в 2012 г. В этом состоит <u>актуальность</u> проведенного исследования.*

<u>Теоретическая значимость</u> проекта связана с тем, что основное внимание в 2012 г. было уделено теоретической подготовке и выработке принципов размещения в Корпусе текстов с сохранением оригинальной орфографии источника. В этом участники проекта видят свою основную задачу и считают, что именно такой подход позволит обогатить диахроническую часть НКРЯ и поднять ее на качественно новый уровень.

В результате сплошного просмотра изданий гражданской печати XVIII века, а также переводных и оригинальных рукописных памятников этого периода, выполненных скорописью, были сформулированы требова-

ния к шрифтовому оформлению текстов, включающему буквы «ять», «юс малый», «от», «ук», «омега», «фита», прописное и строчное «і» с одной, двумя точками и без точек, а также различные типы надстрочных знаков. Были выработаны принципы подготовки текстов гражданской печати XVIII века, в которых наблюдается варьирование в слитном и раздельном написании некоторых служебных слов (например, что бы и чтобы, по елико и поелико, по елику и поелику, до толе и дотоле, да бы и дабы, зде же и здеже и т. п.). Стали предметом обсуждения и были обобщены основные принципы подготовки рукописных текстов, выполненных скорописью. Дело в том, что воспроизведение рукописей названного периода всегда связано с необходимостью упрощенной передачи и некоторой стандартизации их графики. Это касается как буквенных, так и надстрочных знаков, слитного и раздельного написания слов, границ предложения, употребления заглавных и строчных букв. Были обобщены следующие принципы набора рукописных текстов и составлены следующие рекомендации.

1) Последовательная расстановка букв «й», «Ї», «ї», «І», «і» даже в том случае, если в оригинале они употребляются, но не всюду читаются ясно.

2) Расстановка заглавных букв в соответствии с правилами современной орфографии и границами предложений, выполненными при работе с текстом.

3) Помещение выносных букв в строку и выделение их курсивом.

4) Упрощенное воспроизведение надстрочных знаков: печатаются только титла, а ерики и паерки, расставленные в рукописи не всегда последовательно и существенно затруднившие бы чтение текста, не воспроизводятся. Не воспроизводятся и прочие надстрочные знаки.

5) Слитное написание служебных слов со знаменательными, характерное для рукописных памятников начала XVIII века, не воспроизводится. Словоделение выполняется в соответствии с современными орфографическими нормами; это касается и знака дефиса.

6) Деление текста на предложения выполняется издателем. Точно установить в рукописи границы предложений в большинстве случаев не представляется возможным, так как используемые переписчиком пунктуационные знаки, которые указывают, как правило, на конец предложения, расставлены не всегда последовательно и допускают определенные варианты.

7) Пунктуационные знаки внутри предложения также расставляются заново в соответствии с современными правилами пунктуации, в той мере, в какой это позволяет сделать имеющийся порядок слов.

8) Знаки «?», «!», «—», если они отсутствуют в рукописи, расставляются в соответствии со смысловым членением текста; то же относится и к кавычкам. Это поможет пользователям Корпуса правильно прочитать и интерпретировать полученные примеры.

9) Для удобства работы с текстами оригинальная фолиация памятников заменяется пагинацией, если последняя вообще необходима.

10) Слова, допускающие в рукописи слитное и раздельное написание, в подготовленных текстах пишутся по заранее оговоренным правилам. Так, слова иже, ниже, идеже, дондеже, доселе, дотоле, кийждо, нимало, неудобь, сиречь, отвнутрь, отвне, отинуду, никаковой, чтобы, поелику, дотоле, дабы пишутся слитно; а слова еще же, яко же, ничто же, никто же, ни мало, после жде, никогда же, однако ж и т. п.; формы типа крыл бы ся, боял бы ся и т. п., пишутся раздельно.

Представляется, что изложенные принципы помогут стандартизировать подготовку рукописных текстов и изданий гражданской печати для НКРЯ и обеспечат соответствующую теоретическую источниковедческую базу и, как следствие, высокий научный уровень проекта в целом.

В 2012 г. продолжалась работа по пополнению Корпуса новыми текстами. Были подготовлены и размещены в Корпусе тексты общим объемом 3 982 655 словоформ. Внимание уделялось как произведениям художественной литературы, так и мемуарам, публицистике, эпистолярной и исторической прозе. Речь идет, в частности, о произведениях Г. И. Успенского («Растеряевские типы и сцены», «Из путевых заметок», «Пришло на память», «Поездки к переселенцам», «Письма с дороги», «Очерки переходного времени» и т. д.), А. К. Шеллер-Михайлова («Над обрывом»), Е. М. Шавровой («Маркиза»), Н. П. Шаликовой («Семейные сцены»), Е. Н. Ахматовой («Кенелм Чиллингли, его приключения и взгляды на жизнь»), Е. А. Салиаса («Аракчеевский сынок», «Аракчеевский подкидыш»), И. В. Омулевского («Шаг за шагом»), В. Ф. Одоевского («Сказки дедушки Иринея»), Д. И. Иловайского («Регенство Бирона», «Черный ящик», «Начало Руси», «Краткие очерки русской истории», «История Рязанского княжества»), Н. Э. Гейнце («Аракчеев»), А. И. Эртеля («Записки степняка») и других авторов.

Таким образом, полученные в рамках проекта результаты характеризуются <u>новизной</u> и <u>практической значимостью</u>.

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.)

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)

Объем Национального корпуса русского языка (подкорпуса, включающего тексты XVIII-XIX вв.) увеличен на 3 982 655 словоформ. Подготовлены для Корпуса и размещены в Интернете произведения русской художественной литературы, а также мемуары, публицистика и эпистолярная проза, относящиеся к указанной эпохе. Это, в частности, произведения Г. И. Успенского, А. К. Шеллер-Михайлова, Е. М. Шавровой, Н. П. Шаликовой, Е. Н. Ахматовой, Е. А. Салиаса, И. В. Омулевского, В. Ф. Одоевского, Д. И. Иловайского, Н. Э. Гейнце, А. И. Эртеля и некоторых других авторов.

Проанализирован материал и обобщены трудности, связанные с подготовкой и размещением в Корпусе русского языка печатных и рукописных текстов XVIII века в «старой» орфографии. Рассмотрены как технические, так и собственно лингвистические вопросы. К первым относится использование особых символов и специальных шрифтов при подготовке текстов, ко вторым – исследование разных типов вариативности в русском языке XVIII века и, в частности, в орфографии: слитного и раздельного написания служебных слов в печатных текстах. Кроме того, особое внимание было уделено изучению существующих в современной науке принципов публикации рукописных текстов, предполагающей ту или иную степень «упрощения» орфографии оригинала.

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

Корпус русского языка (подкорпус, включающий тексты XVIII-XIX вв.) увеличен на 3 982 655 словоформ. Подготовлены для Корпуса и размещены в Интернете произведения русской художественной литературы, а также мемуары, публицистика и эпистолярная проза, относящиеся к указанной эпохе. Разработаны общие принципы публикации в Корпусе печатных изданий и рукописных памятников XVIII-XIX вв. в «старой» орфографии.

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).

Диахроническая часть НКРЯ объединяет несколько подкорпусов, среди которых тексты XVIII и XIX вв. Структура корпуса и разная степень разработанности его отдельных фрагментов (подкорпусов) обусловливают постановку задач, которые будут решаться в рамках проекта в 2013 г.

Основное внимание при работе над проектом в 2013 г. планируется уделить отбору и размещению в Корпусе текстов с сохранением «старой орфографии». Речь идет о неизданных рукописных памятниках XVIII в., изданиях гражданской печати XVIII века, текстах XVIII века, изданных в XIX веке. Кроме того, будет осуществлено пополнение Корпуса текстами, созданными в начале XVIII века и изданными в авторитетных научных изданиях XX века.

Планируется пополнить Корпус текстами XVIII в. общим объемом 650 000 словоупотреблений:

100 тыс. – неизданные рукописи начала XVIII в.,

100 тыс. – издания гражданской печати первой трети XVIII в.,

300 тыс. – памятники XVIII в., изданные в XIX в.,

150 тыс. – памятники XVIII в., опубликованные в XX в.

Подбор и подготовка печатных и рукописных текстов XVIII века отличаются большой сложностью и трудоемкостью. Предполагаются работы по копированию источников, осуществлению компьютерного набора текста в библиотеке, вычитки набранного текста квалифицированным филологом русистом.

В 2013 г. будет также продолжена работа по определению единых правил воспроизведения и размещения в Корпусе неизданных рукописных памятников, выполненных скорописью XVII-XVIII вв., а также – по стандартизации шрифтового оформления обрабатываемых текстов.

Подпись руководителя проекта С. Ю. Дмитренко

Форма 2
Планируемое содержание работ на 2013 г.

№	Название проекта	Организация-исполнитель и учреждения-соисполнтели	Руководитель проекта (+ кол-во испол-нителей)	Запрашиваемый объем финансирования на 2013 г. (тыс. руб.)	Ожидаемые в 2013 г. результаты
	Пополнение базы текстов XVIII и XIX веков в Национальном корпусе русского языка	ИЛИ РАН	Дмитренко С. Ю. + 3 исполнителя		Размещение в НКРЯ печатных и рукописных памятников XVIII века (Всего 650 тыс. словоупотреблений: 100 тыс. – рукописи начала XVIII в., 100 тыс. – издание первой трети XVIII в., 300 тыс. – памятники рубежа XVIII-XIX вв., изданные в XIX в., 150 тыс. – памятники нач. XVIII в., опубл. в XX в.).