

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

Номер и название направления Программы <b>Направление 4. Создание и развитие корпусных ресурсов по языкам мира.</b>	
Название проекта <b>Корпус языка идиш</b>	
Научный руководитель проекта (ФИО полностью, уч. ст.) <b>К.ф.н. Кисилиер Максим Львович</b>	
Е-mail, телефон, факс (с кодом города) почтовый адрес руководителя проекта	<b>kisimac@gmail.com,</b>
Полное и краткое название организации – адресата финансирования  <b>Институт лингвистических исследований Российской Академии Наук ИЛИ РАН</b>	ФИО (полностью) руководителя организации – адресата финансирования <b>Казанский Николай Николаевич</b>
	ФИО (полностью) главного бухгалтера организации – адресата финансирования <b>Белозёрова Надежда Павловна</b>
	Телефон, факс (с кодом города), Е-mail организации – адресата финансирования тел., факс: (812)3284611 <b>iliran@mail.ru</b>
Год начала – год окончания проекта	<b>2012—2014</b>
Объем финансирования, полученного в 2012 г.:	Объем финансирования, запрашиваемый на 2013 г.
	<b>Перкова Наталья Викторовна, магистр</b>
	<b>Воробьев Григорий Михайлович, магистр</b>
	<b>Кириянов Денис Павлович, студент</b>
	<b>Панова Татьяна Александровна, студент</b>
Дата сдачи отчета 21.11.12	Подпись руководителя проекта:

Координатор Программы

акад. Вяч. Вс. Иванов

Координатор Программы

чл-корр. РАН В.А. Плунгян

«\_\_\_»\_\_\_\_\_2012 г.

1. Название направления: Направление 4. Создание и развитие корпусных ресурсов по языкам мира.
2. Название проекта: Корпус языка идиш.
3. Руководитель проекта (ФИО *полностью*, ученая степень, должность, место работы): Максим Львович Кисилиер, к.ф.н., с.н.с., ИЛИ РАН
4. Основные участники проекта (ФИО *полностью*, ученая степень, должность, место работы):  
Перкова Наталья Викторовна, аспирант, университет Стокгольма;  
Воробьев Григорий Михайлович, аспирант, СПбГУ – университет Мюнстера;  
Кириянов Денис Павлович, студент, СПбГУ;  
Панова Татьяна Александровна, студент, СПбГУ
5. Содержание фактически проделанной работы, полученные результаты, их новизна, научная и практическая значимость, актуальность, соответствие мировому научному уровню (до 5 стр.)  
Проект по созданию корпуса идиша начался несколько лет назад по совместной инициативе Института Языкознания (Москва) и университета Регенсбурга (Германия). Был разработан «черновой» вариант корпуса: в нем содержались лишь некоторые современные периодические нью-йоркские издания, парсер корпуса был адаптирован только к орфографии YIVO, в то время как, помимо вышеупомянутой, в идише существует огромное количество различных орфографий.  
Подключение к проекту Петербургской группы под руководством М. Л. Кисилиера позволило развивать этот проект интенсивнее. Во-первых, в корпус, в котором, как уже было сказано выше, прежде содержались в первую очередь современные нью-йоркские газетные издания, были добавлены некоторые, пока небольшие, работы классиков литературы на языке идиш: Ицика Мангера, Ицхока-Лейбуш Переца, Шолом-Алейхема, Шолома Аша, Менделе-Мойхер Сфорима, Исаака-Башевиса

Зингера, и др.

Во-вторых, 26 октября 2012 в Институте Языкознания в Москве состоялась встреча ведущих участников всех трёх «веток» проекта – Московской, Регенсбургской и Санкт-Петербургской. На этой встрече были определены важнейшие тактические и стратегические вопросы, актуальные в настоящий момент для корпуса, намечены пути их решения. Самым важным результатом является выработка стратегических приоритетов касательно литературы для пополнения корпуса, а также начало работы над адаптацией парсера под различные орфографические стандарты: разнообразие орфографий, которыми пользовались даже крупнейшие деятели культуры идиша, велико, поэтому решение этой задачи – крупный шаг вперед для данного проекта.

В-третьих, участниками проекта была проведена экспедиция в республики Молдова и Приднестровье (историческая юго-восточная Бессарабия) с целью записей материала для последующего создания устного корпуса языка идиш. Несмотря на то, что юго-восточная Бессарабия является предметом пристального внимания некоторых идишистов, устных записей на идише оттуда до сих пор записано не было, в отличие от многих других территорий бывшего СССР; записанный материал, таким образом, вкупе со сделанными ранее записями в других регионах, смогут в будущем лечь в основу корпуса устного языка идиш, который планируется открыть к концу 2014 года.

В-четвертых, состав корпуса – за счет вышеупомянутых произведений классиков литературы на идише - пополнен до пяти миллионов словоупотреблений. Произведен сбор текстов для корпуса. Грамматический словарь корпуса достигает сейчас 15000 слов, завершена работа над системой грамматической разметки.

Итогом вышеописанной работы стала возможность в ноябре 2012 года открыть корпус для всеобщего пользования (адрес ресурса в сети интернет - <http://corpustechnologies.com:8080/YNC/search/>)

Новизна вышеперечисленных результатов заключается в том, что впервые составляется представительное электронное собрание текстов на языке идиш, позволяющее анализировать разноплановое и разножанровое творчество на идише. Подобного собрания текстов на языке идиш ранее не существовало.

Научная значимость работы заключается в том, что в случае осуществления проекта ученые получают инструмент, позволяющий находить и проверять гипотезы в отношении языка идиш, в отношении его взаимовлияния с другими языками (генетически и ареально близкими), а также исследовать диахроническую вариативность. До сих пор не существовало таких работ, в основе которых лежали бы эмпирические данные, а также (за редким исключением) таких, которые рассматривали бы данные идиша как в типологическом аспекте, так и в свете новейших теорий грамматики.

Прикладная значимость: принципиально новые возможности для преподава-

ния языка идиш, поскольку мы можем лучше понять, как используют язык его носители, адаптировать либо создать учебное пособие. Последние зачастую создавались следуя традиции и иногда, как следствие, – игнорируя языковые реалии.

Актуальность работы над корпусом заключается в том, что в данный момент корпусные исследования вышли на передний план в современной лингвистике; для того чтобы сделать достойные внимания выводы о языке, необходимо наличие корпуса. С этой точки зрения у любого языка мира должен быть корпус. Кроме того, для идиша очевидным образом назрела необходимость адекватного грамматического и лексикографического описания, которое в настоящее время немислимо сделать без использования корпуса.

Соответствие мировому научному уровню заключается в том, что с помощью этого корпуса стало возможным решать широкий круг лингвистических проблем, связанных с идишем, самым тщательным образом.

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

6.1. количество монографий – 0, не планировалось

6.2. количество сборников статей – 0, не планировалось

6.3. количество статей – 0, не планировалось.

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.)

18-31 августа 2012 на средства гранта была проведена международная экспедиция на территории республики Приднестровье (Тирасполь, Бендеры, Рыбница) и республики Молдова (Кишинев, Орхей, Бельцы). Руководитель экспедиции – член-корр. РАН В. А. Плунгян. Участники - Иосиф Гордон, Денис Кирьянов, Татьяна Панова, Илья Учитель (все - СПбГУ), Елена Лучина (МГУ), Янина Вурбс (Университет Потсдама, Германия), Алек Бурко (семинария JTS, Нью-Йорк).

Целью экспедиции являлась запись нарратива на идише с целью последующей расшифровки и создания корпуса устной речи на языке идиш. В результате экспедиции было записано около 18 с половиной часов речи носителей языка; полученный материал позволяет:

- 1) к концу 2014 года создать подкорпус устной речи на языке идиш объёмом в 10 часов звучания;
- 2) создать ещё один подкорпус (еще около 8 часов звучания), в который будут выложены те записанные в ходе экспедиции тексты, в которых носители активно и, как правило, неосознанно прибегают к переключению кодов (идиш-русский). Такой материал представляет собой большой интерес для социолингвистов; от разработчиков он потребует тщательного составления разметки.

Помимо вышеупомянутых аспектов, практическая ценность экспедиции заключается в том, что:

- 1) на условной карте идиша на территории бывшего СССР оказалось закрыто еще одно «белое пятно»: несмотря на немалое количество еврейских антропологических экспедиций, проведенных после распада Союза, в том числе и тех, чьи участники работали с информантами на идише, записей на идише из юго-восточной Бессарабии (в отличие от, например, Белоруссии), не существовало. Таким образом, при создании устного корпуса языка идиш - дополнив сделанные в этом году записи уже существующими - теперь стало возможным охватить все диалекты, существующие на территории бывшего СССР (а это практически все три основных диалекта языка: северный, центральный и южный).
- 2) Появилась возможность систематического описания южного диалекта: несмотря на хорошо известные фонетические отклонения от стандартного идиша, мы не очень много знаем об изменениях на остальных уровнях языка.
- 3) По результатам экспедиции стало возможным описать не только особенности южного диалекта, но и социолингвистическое положение идиша в республиках Молдова и Приднестровье на сегодняшний день; подобное описание планируется в докладе с рабочим названием «Sociolinguistical notes on contemporary Yiddish in Bessarabia» (Д. Кирьянов, Е. Лучина), тезисы которого поданы на конференцию «Variation within and across Jewish Languages», которая состоится в университете Антверпена в конце июня 2013 г.

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)

Объём словоупотреблений в корпусе языка идиш увеличен до пяти миллионов, завершена работа над системой грамматической разметки, грамматический словарь увеличен до 15000 слов. В корпус, в котором прежде абсолютное большинство текстов составляли современные американские газеты, добавлены произведения классиков литературы на идише: Айзика-Мейер Дика, Шолом-Алейхема, Шолома Аша, Менделе Мойхер-Сфорима и других. Проведена экспедиция в Бессарабию для сбора материала для корпуса устного идиша, в ходе которой записано 18 часов речи на идише. В будущем возможно открыть два подкорпуса – на «чистом» идише и с переключением кодов. Помимо этого, выработаны новые принципы отбора литературы для пополнения корпуса, а также прин-

ципы адаптации парсера корпуса языка идиш под различные орфографические стандарты.

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

Объем словоупотреблений в корпусе языка идиш увеличен до пяти миллионов, завершена работа над системой грамматической разметки, грамматический словарь увеличен до 15000 слов. Выработаны новые принципы отбора литературы для корпуса и адаптации парсера под различные орфографические стандарты. В корпус добавлены произведения классиков литературы на идише: Шолом-Алейхема и др. Проведена экспедиция в Бессарабию, в ходе которой записано 18 часов речи на идише.

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 планируется добавить в корпус 2,5 миллиона словоупотреблений, закончить работу по созданию единых орфографических стандартов, а также начать расшифровку записанных в ходе экспедиции 2012 г. текстов. Кроме того, на конференции «Variations within and across Jewish languages», которая состоится в Антверпене 26-28 июня 2013 года, планируются выступления участников проекта.

Подпись руководителя проекта

М.Л. Кисилиер

**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 языка идиш	ИЛИ РАН	к.ф.н. М. Л. Кислиер, (+4)		увеличение объема корпуса до 7,5 млн. словоупотреблений, конец работы по созданию единых орфографических стандартов, увеличение грамматического словаря до 20000 слов. Презентация корпуса на международной конференции (?)