

Титульный лист отчета о работе в 2013 г.  
по Программе фундаментальных исследований Президиума РАН  
«Корпусная лингвистика»

Номер и название направления Программы 4: Создание и развитие корпусных ресурсов по языкам мира, раздел «Корпуса малых языков»	
Название проекта <b>Создание электронного корпуса фольклорных текстов диалекта тували языка ифугао, провинция Ифугао, Филиппины</b>	
Научный руководитель проекта (ФИО полностью, уч. ст.) Станюкович Мария Владимировна, к.и.н.	
E-mail, телефон, факс (с кодом города) почтовый адрес руководителя проекта	<a href="mailto:mstan@kunstkamera.ru">mstan@kunstkamera.ru</a>
Полное и краткое название организации – адресата финансирования	ФИО (полностью) руководителя организации – адресата финансирования Чистов Юрий Кириллович
	ФИО (полностью) главного бухгалтера организации – адресата финансирования Савина Татьяна Анатольевна
	Телефон, факс (с кодом города), E-mail организации – адресата финансирования (812) 328 0812, (812) 328 0811, <a href="mailto:info@kunstkamera.ru">info@kunstkamera.ru</a>
Год начала – год окончания проекта	<b>2013—2014</b>
Объем финансирования, полученного в 2013 г.	Объем финансирования, запрашиваемый на 2014 г.
Основные исполнители (ФИО, уч. ст.)	Т.А. Архангельский, к.ф.н.
	С.Б. Клименко
	А.К. Касаткина
	А.А. Чечулин
	А.А. Янковская
Дата сдачи отчета 19 ноября 2013	Подпись руководителя проекта:

**УТВЕРЖДАЕМ**

Координатор Программы

акад. Вяч. Вс. Иванов

Координатор Программы

чл-корр. РАН В.А. Плунгян

«\_19\_» ноября 2013г.

1. Название направления

**Создание и развитие корпусных ресурсов по языкам мира, раздел «Корпуса малых языков»**

2. Название проекта

**Создание электронного корпуса фольклорных текстов диалекта тували языка ифугао, провинция Ифугао, Филиппины**

3. Руководитель проекта

**Станюкович Мария Владимировна, к.и.н., заведующая отделом этнографии Австралии, Океании и Индонезии, МАЭ РАН**

4. Основные участники проекта

**Архангельский Тимофей Александрович, к.ф.н., м.н.с. НИУ ВШЭ**

**Клименко Сергей Борисович, аспирант ИЛИ РАН**

**Касаткина Александра Константиновна, м.н.с. Отдела этнографии Австралии, Океании и Индонезии, МАЭ РАН**

**Чечулин Андрей Алексеевич, н.с. лаборатории проблем компьютерной безопасности Санкт-Петербургского института информатики и автоматизации РАН**

**Янковская Аглая Алексеевна, аспирантка Отдела этнографии Австралии, Океании и Индонезии, МАЭ РАН**

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

В 2013 году участниками проекта были выполнены (с превышением) все задачи, поставленные в начале года и необходимые для создания пилотной версии электронного корпуса фольклорных текстов диалекта тували.

**1. Отбор текстов**

Первой задачей был отбор материалов для создания текстовой базы. Из массива материалов, собранных руководителем проекта в процессе полевых записей и архивных изысканий на Филиппинах с 1995 года, были выбраны неопубликованные фольклорные тексты и интервью, частично выверенные и уточненные в процессе работы в поле в этом (2013) году. Общий объем этих текстов превышает 12000 словоупотреблений (не считая англоязычных фрагментов, время от времени появляющихся в интервью из-за частого переключения кодов). Самым объемным (около 8000 словоупотреблений) является худхуд «Алигуён нак Биненуахен». Неопубликованный текст этого эпического сказания хранится в Архиве Университета Атенео (Манила, Филиппины).

Выбор сказания был обусловлен тем, что данный текст, который по просьбе создателей архива нужно подготовить к публикации, уже был проанализирован руководителем проекта на предмет выявления механизмов сочетаемости личных имен и топонимов; был предложен метод стандартизации текстов по единой схеме для их сопоставления внутри

ареала распространения традиции [Станюкович 2008]. Выполнение этой задачи при создании корпуса текстов поднимается на новый уровень. Методы корпусной лингвистики многократно ускорят и упростят выделение общего и локального в традиции, и в перспективе позволят создать новые инструменты исследования истории миграций и культурных заимствований на Северном Лусоне.

Как и остальные тексты этого собрания, худхуд Алигуёна нуждается в дальнейшей проверке оригинального текста и полной переработке перевода, который сделан совершенно неудовлетворительно. Начатая в экспедиции 2013 г. работа по уточнению и прояснению «темных мест» должна быть продолжена в дальнейшем.

Интервью, заложенные в пилотную версию электронного корпуса, были собраны М.В.Станюкович во время зимней экспедиции 2012 года. Для сопоставительного анализа были выбраны интервью с солистами (Аппин Гуманган), наиболее знающими (Рубен Гуманган) и рядовыми (Каталина Думангин, Антонио Хуммиуат, Пабло Хабойбой) участниками хора, исполняющего эпические сказания. Последний, 82-летний Пабло Хабойбой, родившийся в деревне Пало муниципалитета Кианган (зона распространения диалекта тували ифугао) и переехавший к жене в яттука-язычную деревню, является одним из наиболее выдающихся мумбаки – жрецов. Все интервью были записаны руководителем проекта в первой (из трех) ее экспедиции 2012 года, в феврале 2012 в селении Боко, деревня Нунгауа, муниципалитет Асипуло, провинция Ифугао. Запись интервью предшествовала исполнению похоронного эпического сказания.

Пограничье муниципалитетов Кианган и Асипуло, где были записаны все заложенные в корпус тексты, характеризуется очень высокой степенью многоязычия. Во всех текстах на диалекте тували ифугао имеются иноязычные элементы. В тексте сказания значительная часть лексики взята из диалектов яттука и келей-и южнокордильерского языка каллахан (также известного как калануйа); есть и элементы заимствованной из этих диалектов аффиксации. В интервью живая, некодифицированная речь показывает еще более высокий уровень смешения форм. Это понятно, если учесть, что помимо тували ифугао все интервьюируемые свободно владеют несколькими языками и диалектами (яттука, келей-и, илоканский, самый молодой из интервьюируемых говорит также на тагальском и английском).

В такой ситуации корпус текстов становится неоценимым и, возможно, единственным действенным инструментом как для анализа архаичного языка сказаний, так и для сопоставления его с языком других фольклорных жанров, не говоря уже об описании существующих в настоящее время разговорных языковых норм в регионе.

## **2. Расшифровка и перевод текстов**

Расшифровка, правка и перевод текстов осуществлен с использованием словаря Хохулиных и во взаимодействии с языковыми помощниками из провинции Ифугао Джозефин Патауэг и Марилин Гимбатан как в поле, так и по переписке.

## **3. Обработка текстов для базы данных**

Оригинальный текст расшифрованных аудиозаписей на диалекте тували языка ифугао был построчно выровнен с переводом на английский язык. Исходные тексты без разметки и их переводы хранятся в файлах формата txt в кодировке UTF-8. Кроме того, была составлена таблица с метаинформацией для каждого текста (название, жанр, исполнители, время и условия записи, тема беседы для интервью).

Следующей задачей, стоявшей перед нами, была разметка этих текстов. Возможности и область применения корпуса в первую очередь определяются тем, какими видами разметки он снабжён. Поскольку предполагается, что создаваемым нами корпусом будут пользоваться исследователи с разными интересами и задачами — лингвисты, фольклористы, антропологи, — было решено внести в корпус несколько уровней разметки. Первоначальным образцом создаваемой нами базы данных был корпус шорских и телеутских фольклорных текстов Функа и Шаховцова [Функ et al. 2012], ориентированный в первую очередь на фольклористов и антропологов. Участие в проекте профессиональных лингвистов

Т.А. Архангельского и С.Б. Клименко позволило нам внести морфологическую разметку. Уровни разметки, которыми на первом этапе будет снабжён корпус, таковы:

1. Метаразметка (информация о тексте).
2. Морфологическая разметка.
3. Информация об исполнении сказания (солист/хор).
4. Разметка важных для культуры и фольклора объектов с высоким семиотическим статусом.
5. Личные имена
6. Топонимы

### **1. Метаразметка (информация о тексте).**

«Паспорт текста»:

- диалект и язык текста. В настоящее время все тексты, которые мы вносим в базу, принадлежат к диалекту тували языка ифугао, однако корпус будет расширяться. Во время исполнения сказания всегда делают перерывы для еды, питья и отдыха, во время которых поют вставные развлекательные песенки, иногда на других языках — например, салидуммай на языке калинга.
- «авторы» текста (основной исполнитель фольклорного текста, участники хорового сопровождения; для интервью и записи устных рассказов - рассказчик, интервьюируемый):
  - имя, пол, дата рождения или примерный возраст, родной язык/диалект, знание других языков/диалектов; время записи; место записи; объем текста
- кто записал, каким способом (данные о собирателе, способ записи (видео-, аудиозапись, запись под диктовку, самозапись сказителя, запись местного краеведа и пр.)

Тип текста: фольклор, интервью, рассказ

Для фольклора:

- жанр (сказание жанра худхуд с указанием на подтип — худхуд риса, похоронный худхуд и т.д.; вставные песенки аапо/лиулиуа, заговор и т.д.)
  - обстоятельства записи (во время обряда, исполнение под запись, исполнение отрывка в виде примера в ходе интервью и др.; в случае исполнения во время обряда – время дня или ночи)
  - длительность исполнения текста
  - длительность обряда и место записанного текста в общем сценарии обряда
  - отношение говорящего к данному жанру («так велели предки», «это наша культура/история», «без этого рис не вырастет/ребенка не вылечить», «пустое развлечение», «это слова дьявола», «наш священник это запрещает», «очень опасно», «я христианин, но это мой долг перед родственником», «раньше это только женщины пели», «мы это пели Королеве Софии из Испании» и т.д.)
- Для интервью и бытового рассказа:
- тематика (рассказ о себе, рассказ о жанре фольклора, рассказ о лучших исполнителях прошлого, наивная лингвистика, наивная этнонимика и т.д.).
  - контекст (запись во время паузы между исполнением частей сказания в обряде, запись рассказа- пояснения перед или после обряда, воспоминания по случаю, запись во время плановой работы с информантом и т.д.).
  - хронотоп текста (приблизительное указание на место и время описываемых в тексте событий — для воспоминаний или этиологических рассказов).

### **2. Морфологическая разметка**

Помимо самих текстов в нашем распоряжении была только версия словаря, не предназначенная для автоматической обработки текстов. При выборе способа разметки текстов мы принимали во внимание два соображения. С одной стороны, объём имеющихся у нас текстов не настолько велик, чтобы было необходимо создавать полноценную систему морфологического анализа — тексты можно было частично или полностью разметить вручную.

С другой стороны, составление формализованного описания морфологии ифугао и создание средства для автоматической разметки представляет самостоятельный научный интерес из-за обилия нетривиальных морфологических явлений в этом языке. В итоге было решено придерживаться смешанного подхода, а именно создать систему морфологической разметки с ограниченным словарём и дополнить её ручным разбором наиболее частотных словоформ, а в перспективе ещё и ручным снятием омонимии. В ходе работы участниками проекта были созданы пилотная версия словаря и грамматики в формате, используемом морфологическим анализатором UniParser, успешно использованном ранее для разметки корпусов на других языках.

Для создания электронного словаря в формате UniParser был использован словарь [Nohulin, Nohulin 2010] (недавно он был вывешен на сайте Летнего Института Лингвистики – см [Nohulin, Nohulin 2013]). Был написан скрипт, который способен переводить статьи из этого словаря (в формате doc) в машиночитаемый вид. Конечно, из-за того что в нашем распоряжении была только эта версия словаря, а также из-за ошибок в тексте словаря, в результирующем электронном словаре содержатся ошибки, часть которых устранена вручную, а часть ещё предстоит устранить.

Формальное описание грамматики было создано на основе статьи [Nohulin, Burquest 2007]. Из-за сложной морфологии языка ифугао (допускающей, например, одновременную редупликацию и инфиксацию) составление такого описания представляло собой сложную практическую задачу, для решения которой были созданы специальные вспомогательные программные средства. Информацию о редупликациях и геминациях было решено хранить в файле словаря, перечисляя для каждой лексемы все возможные варианты её основы, которые порождаются специальным скриптом, если известен её класс.

Имеющиеся словарь и грамматика, а также вручную размеченный файл с наиболее частотными словоформами были применены для морфологического анализа текстов. В результате в настоящий момент количество разобранных словоформ в худхуде превышает 60%, а в текстах интервью — 50%. Тем не менее, часть частотной лексики в худхуде остаётся неразобранной, поскольку в речи солиста нередко встречались заимствования и фрагменты на диалекте яттука и, возможно, диалекте келей-и языка калануйа (каллахан), а также из илоканского языка.

Тексты с морфологической разметкой хранятся в XML; кроме того, был создан скрипт, порождающий таблицы Excel с размеченными текстами, содержащие несколько выровненных друг относительно друга слоёв (оригинал, лемма, разбиение на морфемы и клитики, глоссирование и перевод). Представление текстов в виде таблиц позволит вручную редактировать результаты морфологического анализа и снимать омонимию.

На последнем этапе был создан скрипт, который соединяет разобранные тексты, перевод и XML с разметкой объектов и исполнителей и создаёт из них единый XML-документ в формате, используемом корпусной платформой EANC.

### **3. Информация об исполнении сказания (солист/хор).**

Худхуды исполняются солистом и хором; каждая строка начинается солистом и продолжается хором. Вся повествовательная часть содержится в партии солиста, которая характеризуется очень высокой вариативностью. Партия хора, напротив, состоит исключительно из клише — формул, содержащих развернутое имя персонажа (обычно с патронимом, для женщин — с андронимом), пространственных или временных формул. Внутри одного сказания партия хора обладает крайне низкой вариативностью; сопоставление вариантов-формул хора в будущем даст возможность выделения локальных вариаций. Солист руководит хором, который лишь допеваает до конца предложенную солистом группу развернутых формул. Вместе с тем, без квалифицированных участников хора исполнение сказания невозможно. В зависимости от «спетости» коллектива и манеры пения солистки она либо задает формулу хору и отдыхает, обдумывая следующую строку, либо пропеваает всю партию хора вместе с ним.

#### **4-6. Разметка важных для культуры и фольклора объектов с высоким семиотическим статусом, личных имен и топонимов**

Разметка служит нескольким целям. Одна из них — создание в будущем ономастикона эпических сказаний, который в свою очередь позволит решить некоторые вопросы культурных взаимодействий и этногенеза народов Северного Лусона. Разметка объектов с высоким семиотическим статусом позволит создать сетевые отображения связей между персонажами и материальными объектами и функции последних в традиционной религиозно-мифологической системе. На основании разметки в будущем будет осуществлена привязка фотоизображений музейных коллекционных предметов из фондов МАЭ РАН к фольклорным текстам.

И разметка исполнителей, и разметка объектов была внесена в тексты в виде XML и таким образом легко может быть перенесена в базу данных поисковой системы, в которой будут выложен корпус.

Главным достижением является выработка принципов, создание и апробация инструментов для уникальной лингвоантропологической базы данных, в которой языковые данные будут существовать в культурном контексте. Корпус текстов, который нам удалось создать, в перспективе пополнится аудио-, видео- и фотоматериалами, как собранными участниками проекта, так и историческими фотографиями, зафиксировавшими исполнение обрядовых жанров фольклора в далеком прошлом (фотографии столетней давности из архива МАЭ РАН). Несмотря на то, что в исследовании филиппинских языков широко применяются мультимедийные технологии (см. [Станюкович 2012]), единственным существующим в мировой практике аналогом нашей базы является «Говорящий словарь языка Кхинина-анг бонток» выдающегося лингвиста Лоренса Рида [Reid, Kikusawa 2013]. Отличие нашего корпуса от уникального словаря Рида — в его принципиальной направленности не на словарную, а на корпусную линию. Если у Л.Рида основной единицей является отдельное слово, то в нашей базе данных основной единицей является фольклорный текст в обрядовом контексте.

#### **Библиография**

- [Hohulin, Hohulin 2010] — Hohulin, Richard M. and E. Lou Hohulin. Tuwali Ifugao Dictionary [Hohulin, Hohulin 2013] Tuwali Ifugao Dictionary and Grammar Sketch. . [URL] <http://www-01.sil.org/asia/philippines/works-ifk.html>
- [Hohulin, Burquest 2007] — Burquest, Donald & E. L. Hohulin. 2007. Morphophonology in Tuwali Ifugao. *Journal of Philippine Linguistics* 38(1). 1-38.
- [Reid, Kikusawa 2013] — Reid, Lawrence A., compiler, & Ritsuko Kikusawa, project coordinator. Talking Dictionary of Khinina-ang Bontok. The Language Spoken in Guina-ang, Bontoc, Mountain Province, the Philippines. [URL] <http://htq.minpaku.ac.jp/databases/bontok/>
- [Станюкович 2008] — Станюкович М.В. Эпическое сказание ифугао «Алигуён, сын Биненуахена»: сюжет, персонажи и топонимы // Станюкович М.В., отв. ред. и сост. Индонезийцы и их соседи. Festschrift Е.В.Ревуненковой и А.К.Оглоблину. Маклаевский сборник, Вып. 1. Издательство МАЭ РАН. С. 223-243
- [Станюкович 2012] — Станюкович М.В. Филиппинский фольклор: проблемы сбора, хранения и исследования полевых материалов. // Мультимедийные и цифровые технологии в собирании, сохранении и изучении фольклора / Сост. В.Л. Кляус, Е.В. Миненок. Под ред. В.М. Гацака. М., ИМЛИ, 2012. С. 141-147.
- [Функ et al. 2012] — Д. А. Функ, Н. А. Мамонтова, К. Г. Шаховцов. Электронный корпус фольклорных текстов на языках малочисленных народов Сибири (на материалах шорского, телеутского и эвенкийского языков): принципы создания и структура // Мультимедийные и цифровые технологии в собирании, сохранении и изучении фольклора / Сост. В. Л. Кляус, Е. В. Миненок. Под ред. В. М. Гацака. М., 2012. С. 162-179.

## **6. Общее число опубликованных в 2013г. по проекту работ**

2013 год – пилотный этап, первый год работы нашего проекта. Первоначальные выводы были представлены в докладах участников на 1 российской и 5 международных конференциях (1 в России, 1 во Франции, 3 в Англии, 1 в Финляндии). Статьи по докладам готовятся к публикации.

## **10. Экспедиции, проведенные в рамках проекта (регион, руководитель, сроки, тематика исследований, полученные результаты, их значимость)**

Софинансирование: ФФЛИ, МАЭ РАН

Регион – провинция Ифугао, Филиппины

Руководитель – М.В. Станюкович, исполнитель – С.Б. Клименко

Общие сроки экспедиции : с 3 ноября по 11 января (индивидуальные сроки пребывания участников различны)

Тематика исследований:

Работа с информантами с целью выверить перевод и грамматический разбор терминов из частотного списка; работа с информантами по расшифровке и переводу записей; опрос информантов для выявления сказительских творческих биографий и сказительских школ; запись новых сказаний и интервью; сбор данных для метаразметки, морфологической разметки, интервьюирование и сбор коллекционных предметов -семиотически значимых объектов, играющих важную роль в сказаниях.

Полученные результаты, их значимость: на настоящий момент в результате работы с информантами в поле выверен частотный список, что позволило завершить морфологическую разметку. Работы в поле продолжаются, результаты в полной мере можно будет оценить после завершения полевого сезона.

## **11. Конференции, организованные в рамках проекта (название, место и сроки проведения, обсуждаемые проблемы, результаты)**

Начата подготовка к проведению в рамках Маклаевских чтений секции по корпусной лингвистике. Конференция состоится в МАЭ РАН в апреле 2014 года, председатель оргкомитета – М.В. Станюкович. На секции предлагается обсудить методы создания фольклорных корпусов малых коренных народов мира.

## **12. Важнейшие научные результаты работы по проекту (не более 0,5 стр. для публикации на сайте Программы)**

Создана интерактивная база данных малого филиппинского языка, находящегося под угрозой исчезновения, предназначенная как для лингвистов, так и для фольклористов и антропологов. Расшифровано несколько фольклорных текстов и интервью, записанных М.В. Станюкович во время зимней экспедиции 2012 года. В базу заложены тексты на диалекте тували языка ифугао общим объёмом более 12000 словоупотреблений, в том числе худхуд (эпическое сказание) «Алигуён нак Биненуахен». Частотный список слов и ряд текстов выверены в экспедиции. Кроме метаинформации, тексты были снабжены несколькими видами метаразметки: информацией об исполнении сказания (отмечены фрагменты, исполняемые солистом и хором), разметкой персонажей и морфологической разметкой. Для создания морфологической разметки на основе словаря [Nohulin, Nohulin 2010] и статьи [Nohulin, Burquest 2007] было создано формализованное описание грамматики и лексики диалекта тували ифугао в формате UniParser. Словарь в настоящий момент содержит более 4000 лексем. Кроме того, отдельно был создан и вручную размечен список из наиболее частотных словоформ. Имеющиеся словарь и грамматика, а также вручную размеченный частотный список были применены для морфологического анализа текстов. В результате в настоящий момент количество разобранных словоформ в худхуде превышает 60%, а в текстах интервью — 50%. Размеченные тексты хранятся в виде XML и с помощью специально написанного скрипта могут быть преобразованы в таблицу Excel с глоссиро-

ванным текстом; такая таблица будет использоваться для ручной проверки морфологической разметки и снятия омонимии. Формат XML-документов соответствует стандартам корпусной платформой EANC, и в дальнейшем все файлы могут быть выложены в этой платформе.

### **13. Наиболее значимый научный результат проекта**

Создана интерактивная база данных фольклора малого филиппинского языка, находящегося под угрозой исчезновения. Расшифрованные и выверенные в экспедиции тексты общим объемом более 12000 словоформ снабжены метаинформацией, разметкой исполнителей, разметкой персонажей и морфологической разметкой. Для последней было создано описание диалекта тували языка ифугао в формате UniParser и словарь в том же формате объемом более 4000 лексем. Файлы с этой информацией хранятся в XML и могут быть выложены в корпусной платформе EANC.

14. Краткий финансовый отчет за 2013 г. (основные статьи расходов по проекту, сумма)

15. Запрашиваемый на 2014 г. объем финансирования, с кратким обоснованием расходов корпусом текстов будет выполнена в любом случае.

### **16. Краткое обоснование научных работ на 2014 г., ожидаемые результаты**

В 2014 г. основные направления работ будут следующими:

1. Пополнение текстовой базы. К корпусу будет добавлено ещё несколько сказаний и интервью, собранных и/или выверенных в экспедиции, а общий объем корпуса достигнет 30 тысяч словоупотреблений.

2. Усовершенствование морфологической разметки. Большинство статей в словаре, созданном в этом году на основе [Nohulin, Nohulin 2010], требуют некоторой правки, в первую очередь в переводах — их нужно сделать короче (для глоссирования в большинстве случаев достаточно одного-двух слов) и точнее. Кроме этого, для улучшения словаря на основе имеющихся у нас текстов будет нужно скорректировать информацию о сочетаемости глагольных основ с ТАМ-морфемами и внести информацию о наиболее частотных заимствованиях из диалекта яттука языка калануйя, которые часто встречаются в художках. В ряде текстов предполагается вручную снять омонимию.

3. Создание онлайн-корпуса. Размеченные тексты будут доступны в Интернете в корпусной поисковой платформе EANC. Для этого будет нужно внести в неё небольшие изменения, например, для подсветки фрагментов, исполняемых солистом. К сайту с онлайн-корпусом предполагается добавить аудио и видеоматериалы, которые позволят услышать интонацию, видеть мимику и артикуляцию сказителей и информантов, фотографии объектов с высоким семиотическим статусом *in situ* (полевые фотографии) и студийные фотографии собранных предметов из коллекций МАЭ РАН.

Мы планируем провести мини-конференцию по теме проекта в МАЭ РАН в апреле 2014 года, а также опубликовать результаты в виде статей в отечественных и зарубежных журналах и сборниках.

Подпись руководителя проекта



**Форма 2**  
**Планируемое содержание работ на 2014 г.**

№	Название проекта	Организация-исполнитель и учреждения-соисполнители	Руководитель проекта (+ кол-во исполнителей)	Запрашиваемый объем финансирования на 2014 г. (тыс. руб.)	Ожидаемые в 2014 г. результаты
	Создание электронного корпуса фольклорных текстов диалекта тували языка ифугао провинции Ифугао, Филиппины	МАЭ РАН, ИЛИ РАН, ИИиА РАН (все-Санкт-Петербург), ВШЭ (Москва)	Станюкович М.В.+5		Текстовая база будет пополнена, общий объём корпуса достигнет минимум 30 тысяч словоупотреблений. Будет усовершенствована морфологическая разметка. Будет создан онлайн-корпус. Все размеченные тексты будут доступны в Интернете в корпусной поисковой платформе EANC. При условии достаточного финансирования будет проведена экспедиция, к сайту с онлайн-корпусом будут добавлены аудио, видео и фотоматериалы.