

Федеральное агентство по делам национальностей
Правительство Республики Коми
Министерство национальной политики Республики Коми
ГОУ ВО «Коми республиканская академия
государственной службы и управления»
(ГОУ ВО КРАГСИУ)
ФГБОУ ВО «Сыктывкарский государственный университет
имени Питирима Сорокина»
ГАУ РК «Дом дружбы народов Республики Коми»

**ЭЛЕКТРОННАЯ
ПИСЬМЕННОСТЬ НАРОДОВ
РОССИЙСКОЙ ФЕДЕРАЦИИ:
ОПЫТ, ПРОБЛЕМЫ
И ПЕРСПЕКТИВЫ**

Сборник материалов
Международной научной конференции

16–17 марта 2017 г.
Сыктывкар

УДК 003.3:004(470+571)
ББК 81.2
Э45

*Печатается по решению
Совета по науке ГОУ ВО КРАГСиУ
(протокол № 6 от 01.03.2017)*

Ответственный редактор:

Федина М.С., канд. филол. наук, доцент, начальник Центра инновационных языковых технологий ГОУ ВО КРАГСиУ.

Редакционная коллегия:

Чуяшкова М.В., начальник отдела государственных языков Министерства национальной политики Республики Коми; **Тюкавина И.А.**, канд. ист. наук, доцент кафедры управления информационно-документационными и социально-политическими процессами ГОУ ВО КРАГСиУ; **Завгородняя Л.П.**, помощник ректора по международным связям ГОУ ВО КРАГСиУ.

Э45 Электронная письменность народов Российской Федерации: опыт, проблемы и перспективы (16–17 марта 2017 г., Сыктывкар): сб. материалов Международной науч. конф. – Сыктывкар: ГОУ ВО КРАГСиУ, 2017. – 308 с.
ISBN 978-5-93206-356-9

В сборнике публикуются тексты докладов и выступлений участников Международной научной конференции «Электронная письменность народов Российской Федерации: опыт, проблемы и перспективы» (16–17 марта 2017 г., Сыктывкар).

Статьи посвящены различным аспектам поддержки и функционирования языков народов России в электронном информационном пространстве: использованию языков в Интернете, теории и практике создания и использования электронных образовательных ресурсов, электронных библиотек, национальных корпусов языков, различных баз данных, электронных словарей, энциклопедий, переводчиков и т.д.

Статьи написаны на русском, коми и английском языках.

УДК 003.3:004(470+571)
ББК 81.2

ISBN 978-5-93206-356-9

© ГОУ ВО КРАГСиУ, 2017

СОДЕРЖАНИЕ

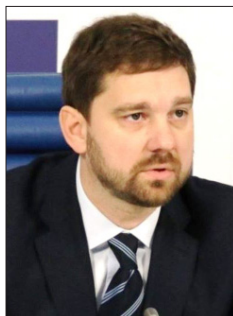
Приветственные слова	6
Алос-и-Фонт Э. Профиль сельских школьников Чувашии, использующих чувашский, татарский, русский или иностранный язык на социальных сайтах и в SMS-сообщениях	11
Баженов И.И. Опыт использования национального корпуса русского языка на учебных занятиях по дисциплине «Информационные технологии»	24
Бартфаи Чаба Причины выбора удмуртского языка в социальных сетях	29
Болтачев А.В. Разработка сервиса для редактирования марийско-русского, русско-марийского словарей	34
Бускунбаева Л.А., Сиразитдинов З.А., Ишмухаметова А.Ш. Состав и структура корпуса публицистики башкирского языка	39
Бусыгина Л.В., Байтерякова Ю.Т., Боталова Н.П. Электронные учебники как средство интенсификации и повышения качества обучения удмуртскому языку	44
Габова М.А., Поберезкая В.Ф. Технологии создания и использования электронных образовательных ресурсов на родном языке в начальной школе	49
Габова М.А., Поберезкая В.Ф., Терентьева С.Н. Технология использования коми-русского электронного образовательного ресурса «Окружающий мир» в начальной школе ..	55
Гагиева А.К. Исторический опыт документирования языка	62
Григорьев Г.Л. Популяризация и распространение языков с помощью мобильных приложений на примере удмуртского языка ..	67
Епанов В.Н. Функционирование коми-пермяцкого языка в социальных сетях	70
Золина Е.Н., Хуснутдинов А.А. О некоторых особенностях структурной организации электронных текстов	73
Каракулов Б.И. Удмуртский язык в электронном информационном пространстве	78
Коваль Е.С. Возможности использования национального корпуса коми языка в научных исследованиях (на примере работы «Сравнительные конструкции в коми языке»)	84
Коровина Н.С., Савельева Г.С. Фольклорное наследие народа коми в современной культуре	88



Кунгурцева В.С., Трашеев С.В. Особенности интеграции средств информационно-коммуникационных технологий в современное образовательное пространство	97
Леонтьев Н.А., Торотов Г.Г. Многопользовательская морфологическая разметка корпуса якутского языка	101
Лобыгин А.Н., Тимерханова Н.Н. Электронное учебное пособие по краеведению на удмуртском языке	104
Миллер Е. Автоматическое выравнивание словарей литературного удмуртского языка и бесермянского диалекта	109
Мусанов А.Г. Геоинформационная система «Географические названия Республики Коми»	112
Муталов Р.О., Муталов М.Р. Микроструктура электронного словаря даргинского языка	117
Нагурная С.В. Новописьменный карельский язык в социальных сетях	122
Нестерова Н.А. Лаборатория будущего	129
Новикова Н.М. О реализации проекта аудиобиблиотеки на коми языке «Ловъя кыв» (Живоекомислово.рф)	133
Остапов Н.В. Учебный мультимедийный проект «Аслам чужанінб туй» / «Дорога домой»: от проектирования до реализации	135
Остапова Е.В. О внедрении в образовательный процесс электронных ресурсов по коми литературе	139
Остапова З.В., Кулышева Л.Г., Рудецкая М.В. Электронные образовательные ресурсы этнокультурной направленности в дошкольном образовании Республики Коми	145
Пишлѐгер К. Удмуртский язык в социальной сети «ВКонтакте»: Квантитавные и (возможные) квалитативные исследования	154
Попова И.А. Ресурсы облачных интернет-технологий в системе дистанционного преподавания курсов русского языка и литературы в основной школе	163
Рыбка К.В. Заимствованная лексика в сфере информационных технологий как стилистическое средство	166
Саая С.Д. Способы создания компьютерного интерфейса на тувинском языке (на основе социальной сети «ВКонтакте»)	173
Сабанцев Г.Л., Чемышев А.В. Яндекс.Переводчик и языки России	178
Силин В.И., Акулова Л.И. Создание электронных справочно-библиографических баз данных по истории географического изучения и краеведению регионов (на примере Республики Коми) ...	182



Сиразитдинов З.А., Бускунбаева Л.А., Шамсутдинова Г.Г.	
Структура словарей в лексикографическом подфонде машинного фонда башкирского языка	190
Соловар В.Н. Актуальные вопросы совершенствования письменности хантыйского языка	195
Сорокина А. Система для сравнительно-исторического корпусного анализа пермских языков	200
Тимерханова Н.Н. «По родным просторам» с родным языком	203
Тимерханова Н.Н., Кудрявцев А.Ф. Энциклопедия по географии на удмуртском языке	212
Федина М.С., Левченко Д.А. Из опыта создания коми медиатеки ...	220
Ферхеес С. Эвиденциальность и перфект в рутульском языке (на материале говора с. Кина)	228
Фридман Г.И., Королькова А.В. Шрифт как основа электронной письменности	236
Хазова Е.В., Матвеева Л.Д., Исакова О.Л. Функционирование коми языка в официальных интернет-источниках	242
Васильева А.А. Кызды велӧдны тӧдтӧм кыв Интернет отсӧгӧн (ас опыт серти)	247
Макарова Е.А., Кожевина Е.В. Блогосфераын коми кывлӧн петкӧдчӧм	249
Сычева Ю.С. «Арт» журналлӧн «Коми гижысьяс лыддьӧны ассыныс кывбурьяс» медводдза мультимедиа приложение йылысь	255
Hakkarainen J.-P. You can't have one without the other: the role of (electronic) libraries in the linguistic eco-system	260
Partanen N. Challenges in OCR today: Report on experiences from INEL	263
Pirinen T.A., Tyers F.M., Trosterud T., Vincze V., Riebler M. ACL SIGUR – Association of Computational Linguistics’ Special Interest Group on Uralic languages	274
Riebler M. Language Documentation meets Language Technology	279
Rueter J. Fieldwork and Early Literary Texts (FELT)	288
Trosterud T. Language technology in Russia	294
Dr. Wilbur Joshua The Pite Saami lexicographic backbone From a FileMaker Pro database to published digital results	299



Игорь Баринов

*руководитель
Федерального агентства
по делам национальностей*

«Язык — это основа идентичности. Отношение к родному языку, по сути, отношение к самому себе, к своим предкам и потомкам. Язык — это духовная и культурная летопись народа, ведь он несёт в себе больше информации, чем самые объёмные энциклопедические издания. В то же время язык — живой организм, в котором запечатлены все исторические вехи, спаяны времена и эпохи.

Уверен, что наши совместные усилия должны быть направлены на повышение качественного уровня владения русским языком, родными языками, их сохранение, развитие и популяризацию среди молодёжи. Очень важно при этом, какие средства коммуникации с молодёжью мы выбираем. Конечно, для подрастающего поколения, идущего в ногу со временем, удобнее получать знания с использованием доступных информационных технологий, поэтому те вопросы, которые обсуждаются на конференции, чрезвычайно важны и востребованы сегодня».



Л.В. Максимова

*Первый заместитель
Председателя Правительства
Республики Коми*

Уважаемые гости, участники и организаторы конференции! Искренне приветствую всех вас от имени Правительства Республики Коми на нашей гостеприимной многонациональной земле Коми.

Сыктывкар является одной из главных площадок республики, где проводятся международные и всероссийские научные и общественные форумы. Знаковым событием этого года стала международная конференция «Электронная письменность народов Российской Федерации: опыт, проблемы и перспективы».

В эпоху Интернета, мощного развития IT-технологий и создания единого информационного пространства проблема сохранения языков встаёт острейшим образом не только в сфере их традиционного использования, но и с точки зрения наличия их электронной письменности. Создание и развитие электронной письменности – важный фактор выживания языков народов России в условиях процесса глобализации, залог сохранения культурного и языкового разнообразия.

В Республике Коми работа по развитию коми электронной письменности в последние годы ведётся особенно активно. Это постоянная забота не только о сохранении, но и прежде всего о расширении среды функционирования национального языка. На базе Коми республиканской академии государственной службы и управления 5 лет успешно функционирует Центр инновационных языковых технологий, включающий в себя Бюро офици-

ального перевода и Межрегиональную лабораторию информационной поддержки функционирования финно-угорских языков. Профессиональной команде Центра за очень короткие сроки удалось разработать и апробировать целый спектр электронных продуктов на коми и иных языках.

Правительство Республики Коми исходит из понимания того, что язык – это основа духовности, средство познания подрастающим поколением себя, общества и родины, средство сохранения и развития этноса. Поэтому мы и впредь будем поддерживать всё, что связано с развитием родных языков.

Уверена, что сегодняшняя конференция послужит объединению исследователей, использующих компьютерные подходы в работе с языками Российской Федерации, даст импульс развитию компьютерной лингвистики в высших учебных заведениях и научно-исследовательских организациях различных регионов Российской Федерации, упрочит взаимодействие науки, власти, бизнеса и институтов гражданского общества для продвижения языков народов России в электронном информационном пространстве.

Благодарю всех участников конференции и гостей, что нашли время посетить нашу республику. Желаю плодотворных дискуссий, налаживания контактов и обретения нового опыта!



Е.В. Савтенко

*Министр
национальной политики
Республики Коми*

Дона войтыр!

Уважаемые коллеги! Рада приветствовать вас, участников, гостей и организаторов Международной конференции «Электронная письменность народов Российской Федерации: опыт, проблемы и перспективы»!

Вопросы сохранения и развития языков народов России активно обсуждаются на федеральных и региональных площадках, где признаётся, что сбалансированная политика в отношении языков имеет стратегическое значение для развития гармоничных межнациональных отношений и обеспечения гражданского единства.

Республика Коми достигла определённых результатов в деятельности, направленной на поддержку государственных языков – коми и русского. Деятельность эта многогранна и разнообразна, в неё вовлечены многие государственные и общественные институты. Немаловажное значение придаётся научно-практическому аспекту этой деятельности. Начиная с 2009 г. стало доброй традицией собирать в г. Сыктывкаре учёных, экспертов, учителей для обсуждения места родных языков в системе образования, развития билингвизма в условиях современного общества.

Сегодня мы открываем новую площадку, объединившую исследователей и разработчиков электронной письменности на родных языках народов нашей страны.

Этнокультурное и языковое многообразие должно быть представлено в мировом электронном информационном пространстве.

Это прежде всего продиктовано необходимостью расширения области функционирования национальных языков, обеспечения их равноправия в *IT*-среде. Эта нелёгкая задача требует приложения сил науки, власти, бизнеса и институтов гражданского общества.

Мы надеемся, что конференция актуализирует тематику и направления научных исследований по компьютерной лингвистике, определит приоритетные области внедрения результатов таких исследований, укрепит межрегиональное и международное сотрудничество.

Желаю всем эффективного диалога, новых идей и полезных контактов! Мы же со своей стороны готовы к сотрудничеству со всеми, кому небезразлична судьба родных языков, в том числе на просторах глобального информационного пространства.



Э. Алос-и-Фонт,
*внешний научный сотрудник
Исследовательского центра
социолингвистики и коммуникации
Университета Барселоны*

ПРОФИЛЬ СЕЛЬСКИХ ШКОЛЬНИКОВ ЧУВАШИИ, ИСПОЛЬЗУЮЩИХ ЧУВАШСКИЙ, ТАТАРСКИЙ, РУССКИЙ ИЛИ ИНОСТРАННЫЙ ЯЗЫК НА СОЦИАЛЬНЫХ САЙТАХ И В SMS-СООБЩЕНИЯХ*

В Чувашии с начала 1960-х гг. распространена образовательная система, в которой преподавание на чувашском и татарском языках ведётся, как правило, только в сельских школах и только с первого по четвёртый класс включительно. Данная система, среди прочего, формирует у учащихся привычку читать на русском, но мало прививает навыки чтения и письма на других языках. Социолингвистическое исследование среди школьников Моргаушского и Аликовского районов Чувашии показало, что, хотя им, как правило, более комфортно говорить по-чувашски, однако при этом им скорее легче читать и писать по-русски [1]. По мере того, как увеличивался контакт с русским языком в сельской местности в последние десятилетия через телевидение, радио и Интернет, становилось больше детей, приехавших из городов в сёла, дискомфорт при общении на русском языке снижался, как и, следовательно, при чтении и письме на нём. В результате при такой же поддержке родному нерусскому языку, как и десятилетия назад, навыки чтения и письма на нём снижались. Их можно оценивать посредством анализа функционирования чувашского и татарского языков в социальных сетях.

Уровень использования чувашского языка в SMS-сообщениях и соцсетях среди школьников Чувашской Республики исследован в разных ситуациях: в городах [2], районных центрах [3] и Моргаушском и Аликовском районах Северной Чувашии [4]. Результаты показали, что чувашский язык используют гораздо реже письменно, чем устно. В качестве примера можно привести Моргаушский район (97% населения – чуваша), где из 762

© Алос-и-Фонт Э., 2017

* Статья написана при поддержке стипендии Мемориального фонда им. Г.Й. Рамстедта Финно-угорского общества (Финляндия) для проекта «Ситуация чувашского и татарского языков в Южной Чувашии».



опрошенных-школьников 71,6% говорят с друзьями на чувашском языке и 53,1% – в основном на нём, 30,6% посещающих соцсети заявляют, что пишут в них на нём, и 2,7% – в основном на нём. Проверка реального использования чувашского языка в соцсетях у почти 40 респондентов показала, что они реже пишут на нём, чем заявляют [5]. Тем не менее до сих пор не изучен профиль представителей населения, использующих чувашский язык.

С целью прояснения ситуации с чувашским и татарским языками в южных районах Чувашии в феврале-апреле 2016 г. опрошены 913 учащихся 7–11 классов 19 случайно выбранных школ Комсомольского и Батыревского районов (31% учащихся этих классов). В данных районах 69% населения – чуваша, 27% – татары, 3% – русские. Выборка соответствует доле распределения населения по району, типу населённого пункта (райцентр или не райцентр) и национальному составу (более подробно о выборке см. источник [6]). Анализ данных владения чувашским и татарским языками показал, что в районных центрах родители-чуваша недавно перешли в основном на русский язык в общении с детьми, и такое явление начинает распространяться в ближайших к с. Комсомольское деревнях. Среди татар передача родного языка хорошо сохраняется [7].

Из 902 учеников – участников выборки, проживающих в изученных районах, 67% опрошенных (600 человек) – чуваша, 28% (254 человека) – татары. С друзьями на чувашском языке говорят 55,3%, 45,1% – в основном на нём. С ними на татарском говорят 25,9%, 23,7% – в основном на нём. Что касается использования языков в соцсетях, 40,2% посещающих их заявляют, что пишут в них на чувашском, 11,3% – в основном на нём, 21,6% – на татарском, 9,6% – в основном на нём. Посещают соцсети 99% опрошенных. В изученных районах Южной Чувашии отмечено значительно более частое использование чувашского языка среди чувашей, чем в северном Моргаушском районе. При этом использование татарского среди татар наблюдается значительно чаще, чем чувашского среди чувашей изученных здесь районов (см., например, различие, с одной стороны, между 67 и 40,2% и, с другой стороны, между 28 и 21,6%). Также заметно, что чуваша и татары мало используют язык соседней национальности вопреки тому, что часто учатся в одной школе: только 6 чувашей говорят об использовании татарского языка в соцсетях, 9 татар – чувашского (хотя 53 татарина заявляют, что свободно пишут по-чувашски).

Всё это даёт мало информации о том, кто выбирает один или другой язык в соцсетях, поэтому нами был использован многовариантный статистический анализ, позволяющий уточнить влияние разных факторов в одном явлении. Поскольку, как часто бывает в социологических анализах, много факторов не следуют нормальному распределению, нами использо-



ваны логистические регрессии. В данной статистической модели событие (зависимая переменная) должно принять только две стоимости («да» или «нет», «0» или «1» и т.д.), а признаки (независимые переменные) могут быть разнообразными, в том числе нет необходимости, чтобы они следовали нормальному распределению (о подробностях статистической модели см. [8]). В качестве зависимых переменных выбраны СоцСетиЧув (респондент пишет на чувашском языке в соцсетях) и СоцСетиТат (респондент пишет на татарском языке в соцсетях). Для того чтобы лучше определить влияние факторов, в исследовании к этим двум основным переменным добавлены ещё 6:

- СоцСетиИно: респондент пишет на иностранном языке в соцсетях;
- СоцСетиТолькоРус: респондент пишет *только* на русском языке в соцсетях;
- SMSЧув: респондент пишет на чувашском языке в SMS-сообщениях;
- SMSTат: респондент пишет на татарском языке в SMS-сообщениях;
- SMSИно: респондент пишет на иностранном языке в SMS-сообщениях;
- SMSTолькоРус: респондент пишет *только* на русском языке в SMS-сообщениях.

Частотность этих событий указана в *табл. 1*.

Т а б л и ц а 1

Частотность стоимости «1» («правда») у зависимых переменных по национальности респондентов, в %

	Все	Чуваши	Татары
СоцСетиЧув	40,2	56,7	—
СоцСетиТат	21,6	—	74,1
СоцСетиИно	5,3	—	—
СоцСетиТолькоРус	37,6	—	—
SMSЧув	36,6	52,8	—
SMSTат	20,3	—	69,3
SMSИно	3,5	—	—
SMSTолькоРус	41,8	—	—

Следует отметить, что при проведении многовариантного анализа необходимо комплексно интерпретировать появление независимых переменных в регрессиях. Например, уже известно, что чувашский язык лучше сохраняется в Батыревском районе, чем в Комсомольском [9]. Следовательно, очевидно, что в Батыревском районе респонденты пишут на нём чаще,



чем в Комсомольском. Тем не менее основная причина состоит не в месте проживания, а в различных частотах общения на языке дома в каждом районе. Одновременное изучение обоих факторов в многовариантном анализе укажет, существуют ли специфические для района явления. Например, есть ли в одном районе группы школьников, использующих чаще или реже один язык за пределами той ситуации, которую можно предполагать в зависимости от домашнего языка (а также какой у родителей уровень образования и т.д.)?

Важным условием для точного определения роли разных факторов в многовариантном анализе является их независимость друг от друга. Например, у нас есть информация о численности и расстоянии до районного центра населённых пунктов, где проживают респонденты, а также об их языках домашнего общения. Можно предполагать, что эти 3 фактора играют определённую роль в выборе языка общения респондентов. Проблема в том, что они сильно коррелируют друг с другом. Выбранная статистическая модель не позволяет это определить, напротив, необходимо выбрать факторы (независимые переменные) без сильных корреляций между собой. В данном случае мы выбрали язык домашнего общения, потому что, на наш взгляд, он является основным объясняющим фактором.

Важным представляется понимание роли школьного образования и семьи. Проблема в том, что они также сильно коррелируют между собой: дети, которые дома говорят на чувашском, татарском или русском языке, чаще всего учились на том же языке в начальной школе. Поскольку оба фактора нельзя включать в одну регрессию, мы изучали 2 регрессии – с языком домашнего общения и языком преподавания в начальной школе.

Другой вопрос заключается в том, что, как указано выше, использование чувашского и татарского языков в большей степени ограничено представителями соответствующей национальности. Чтобы лучше определить факторы, связанные с использованием этих языков, мы сделали такие же регрессии для этих языков в общей и «национальной» выборках.

В результате для чувашского и татарского языков вычислены по 8 регрессий в зависимости от: 1) типа общения (соцсеть или SMS); 2) выборки (общей или «национальной»); 3) сферы использования языка (дома или в начальной школе).

Для русского и иностранных языков выборка была сделана немного иначе: использована только общая выборка, но кроме включения средней оценки респондента по 8 предметам (так же, как сделано для чувашского и татарского языков) отдельно проанализирована возможная связь с оценкой по предметам «Русский язык» и «Иностранный язык». Следовательно, также в этих двух случаях вычислены по 8 регрессий.



В итоге в качестве независимых переменных выбраны следующие (для которых затем описываются основные величины и предполагаемое влияние на зависимые переменные):

БАТЫРЕВО: район проживания респондента: «0» – Комсомольский, «1» – Батыревский. 594 (66%) респондентов живут в Батыревском, 308 (34%) – в Комсомольском.

ДЕВОЧКА: пол респондента для обработки был кодирован: «0» – мальчик, «1» – девочка. В выборке 510 (57%) девочек и 392 (43%) мальчика.

РАБОТПАПЫ: есть ли работа у отца? Для получения показателя социально-экономического уровня семьи респондентов спрашивали, есть ли работа у родителей. В некоторых случаях стало известно из личного разговора с ними, их письменных ответов или объяснений преподавателей, что одного или обоих родителей нет. Мы решили кодифицировать эти известные нам случаи для более точного понимания ситуации, хотя значительная часть отсутствия ответов, несомненно, относится к отсутствию отца. Таким образом, определены 3 стоимости переменной. У 19 (2%) респондентов нет отца, у 251 (29%) отец не работает, у 587 (68 %) – работает (проценты вычислены от общего числа ответивших на вопрос). В отношении 45 отцов этой информации нет.

РАБОТМАМЫ: есть ли работа у матери? У 6 (1%) респондентов нет матери, у 285 (32%) матери не имеют работы (или работают домохозяйками, а также на домашней фермерской работе), у 605 (67%) – имеют работу. 6 респондентов не ответили на вопрос.

ОБРАЗПАП: уровень образования отца. Образование родителей влияет на выбор языка общения с детьми, поэтому предполагалось, что оно может влиять на уровень владения письменным языком и, следовательно, на выбор языков письменного общения. Уровень образования кодифицирован по 7 различным уровням, начиная с «неполного среднего или ниже» и кончая «послевузовским». Отцы 170 (20%) опрошенных имеют высшее или послевузовское образование, а 271 (33%) – общее среднее или ниже. По 69 отцам нет таких данных.

ОБРАЗМАМ: уровень образования матери. У матерей 262 (30%) респондентов есть высшее или послевузовское образование, а у 224 (26%) – общее среднее или ниже. В отношении 41 матери таких данных нет.

КНИГИ: количество книг дома. Дополняет информацию об уровне образования родителей. У 430 (48%) опрошенных есть от 26 до 100 книг. Нет таких данных в 5 случаях.

Оценки: средняя оценка по предметам «Русский язык», «Русская литература», «Иностранный язык», «История», «Алгебра», «Геометрия», «Физика», «Химия». Респонденты написали свои оценки («2»–«5») для каждого



предмета за последнюю четверть (средние классы) или первое полугодие (старшие классы). Оценки по отдельным предметам сильно коррелируют между собой, поэтому мы выбрали среднюю величину. Предполагалось, что лучшие школьники склонны расширять свой языковой репертуар, особенно, если это касается иностранных языков. Средняя величина для всех респондентов – 4,1.

ОценкаРус: средняя оценка по предмету «Русский язык». Средняя величина для всех респондентов – 4,0.

ОценкаИно: средняя оценка по предмету «Иностранный язык». Средняя величина для всех респондентов – 4,2.

ШколаЧув: респонденты указали, где они учились в начальных классах, основной язык обучения, а также языки, на которых они общались с преподавателями и одноклассниками. На основе этих данных и информации, полученной от представителей школ, мы установили языки обучения в начальных классах. Как правило, в данных районах обучение ведётся на русском языке в районных центрах, а также в отдельных школах Комсомольского района. В остальных школах преподавание ведётся на чувашском или татарском языке с 1 по 4 класс. При чувашском языке обучения переменная кодирована «1», при русском – «0», при обоих языках обучения (из-за изменения школы) – «0,5». В итоге 50% респондентов обучались на чувашском языке в начальных классах, в том числе 73% чувашей. Для 18 респондентов нет данных по использованию языков в начальной школе.

ШколаТат: 25% респондентов обучались на татарском языке в начальных классах, в том числе 87% татар.

ШколаРус: 24% респондентов обучались на русском языке в начальных классах.

СемьяЧув: респонденты заявили, на каких языках они говорят с разными членами семьи и в какой степени используют их. Для общения с каждым человеком определён индекс использования чувашского: «5» – используется только чувашский язык, «1» – используются только другие языки, «3» – используются чувашский и другие языки в одинаковой степени. В качестве показателя уровня общения на чувашском языке в семье выбрана средняя величина индексов для матери, отца и родных братьев/сестёр. В среднем для всей выборки индекс «СемьяЧув» – 3,0, среди чувашей – 3,9.

СемьяТат: в среднем для всей выборки индекс использования татарского языка в семье «СемьяТат» – 2,0, среди татар – 4,7.

СемьяРус: в среднем индекс использования русского языка в семье «СемьяРус» – 1,9.



Далее представлены результаты 32 логистических регрессий, подсчитанных при помощи статистической программы R 3.3.1.

Результаты корреляций в логистических регрессиях по использованию чувашского языка в соцсетях и SMS-сообщениях (см. *табл. 2*) показывают, что семья и школа являются основными факторами, определяющими выбор языка общения. Кроме того, школьники с лучшими оценками более склонны писать по-чувашски, особенно в SMS-сообщениях, где этот язык используется реже. В двух регрессиях появляется возможное отрицательное влияние более высокого уровня образования мам в использовании чувашского. На первый взгляд, результат недостаточно достоверен, поскольку корреляции слабы и ограничены SMS-сообщениями, но результаты для русского также указывают на подобное явление (см. далее).

Более высокие стоимости показателя Псевдо R^2 Макфаддена для полной выборки в корреляциях с «СемьяЧув» по сравнению с «ШколаЧув» показывают, что семья играет большую роль, чем школа, в выборе чувашского как языка сообщений.

Результаты корреляций в логистических регрессиях по использованию татарского языка (см. *табл. 3*) также показывают, что семья и школа являются основными факторами, что логично. Заметно, что данные факторы появляются значительно реже в корреляциях татарской выборки. Такой результат объясняется тем, что практически все татары изученных районов говорят дома по-татарски, а для подавляющего большинства татарский был языком обучения в начальных классах (именно тот момент, что он не был таким для всех, объясняет, что в татарской выборке школьной фактор играет одну из определяющих ролей в выборе языка общения).

Так же, как и в ситуации с чувашским языком, школьники с более высокими оценками более склонны использовать татарский язык. Кроме того, более высокий уровень образования отца оказывает положительное влияние на выбор татарского языка для общения в соцсетях, но при этом такое влияние не проявляется в общении при помощи SMS-сообщений. Это неожиданный результат. Как правило, в литературе отмечается, что матери оказывают большее влияние на язык детей, чем отцы, или нет различий в уровне влияния родителей [10]. Возможно, дальнейшее изучение роли отцов и матерей в передаче татарского языка сможет пролить свет на этот вопрос.

При анализе использования иностранных языков необходимо быть осторожными, потому что малое число респондентов указали на их использование: 47 (5,3%) в соцсетях, 31 (3,5%) в SMS-сообщениях. Респондентами отмечено использование корейского языка в соцсетях в 5 случаях (учениками разных школ), японского, эстонского и древнерусского – по 1.



Т а б л и ц а 2

*Достоверность корреляций в логистических регрессиях
для использования чувашского языка в соцсетях и SMS-сообщениях*

	СоцСетиЧув				SMSЧув			
	Всего		Чуваши		Всего		Чуваши	
Случаи	40,2%		56,7 %		36,6%		52,8%	
БАТЫРЕВО			?	+				
ДЕВОЧКА								
ОБРАЗПАП								
ОБРАЗМАМ							–	?
РАБОТПАПЫ								
РАБОТМАМЫ								
КНИГИ								
Оценки			+	+		?	++	++
ШКОЛАЧув	+++		+++		+++		+++	
СЕМЬЯЧув		+++		+++		+++		+++
N	816	816	537	537	800	800	521	521
Псевдо R ² Макфадде- на, %	22,5	26,6	13,1	14,1	25,0	26,8	14,5	13,9

Примечание. В столбцах представлены результаты значимости зависимых переменных (построчно). Знаки «+» или «-» обозначают соответственно положительную или отрицательную корреляцию. Один такой знак показывает достоверность на 95%, два – на 99%, три – на 99,9%. Знаки «?+» и «?-» показывают соответственно положительную или отрицательную корреляцию с достоверностью на 90% (признаваемую сомнительной для определения корреляции). Серым фоном отмечены переменные, не включённые в регрессии. Строка «N» показывает количество ответов, включённых в регрессии. «Псевдо R² Макфаддена» – показатель уровня пояснительности регрессии: чем больше число, тем больше уровень.

Кроме отдельных случаев, представляется маловероятным, чтобы школьники изучали эти языки настолько хорошо, чтобы иметь возможность более-менее свободно общаться на них. Скорее всего, знание лишь отдельных слов или выражений является достаточным, чтобы испытывать гордость по поводу владения этими языками. Можно предполагать, что это справедливо и для английского, который изучают в школе. Поэтому, вероятно, в большинстве случаев отмечается лишь минимальное использование иностранного языка.



Т а б л и ц а 3

*Достоверность корреляций в логистических регрессиях
для использования татарского языка в соцсетях и SMS-сообщениях*

	СоцСетиТат				SMSTат			
	Всего		Татары		Всего		Татары	
Случаи	21,6%		74,1%		20,3%		69,3%	
ДЕВОЧКА								
ОБРАЗПАП	++	+	+	+	?+			
ОБРАЗАМАМ								
РАБОТПАПЫ								
РАБОТАМАМЫ								
КНИГИ								
ОЦЕНКИ	?+	+	+	+		+		?+
ШКОЛАТАТ	+++		+		+++		+++	
СЕМЬЯТАТ		+++				+++		+
N	816	816	235	235	800	800	235	235
Псевдо R ² Макфаддена, %	56,2	60,6	14,1	13,0	58,9	60,3	13,5	11,0

Результаты исследования используют иностранные языки показывают, что это скорее учащиеся, говорящие дома по-русски и обучавшиеся на нём в начальных классах (см. табл. 4). Лучшие оценки по иностранному языку коррелируют с использованием иностранных языков в соцсетях, но не сильно (оценки сильнее коррелируют с использованием чувашского и татарского языка). Такое влияние не проявляется в SMS-сообщениях, что свидетельствует о более элементарном использовании иностранных языков в этом контексте. Заметны сильная корреляция с количеством книг дома и отсутствие влияния уровня образования родителей. У нас нет объяснений для этого результата. Также интересно, что учащиеся Комсомольского района несколько чаще используют иностранные языки, чем школьники Батыревского. Возможно причина в том, что Комсомольский – более русифицированный район, или в нём по какой-то причине более развито определённое игровое общение на иностранных языках.

Из всего этого самым любопытным результатом является то, что русскоязычные чаще используют иностранные языки. Поскольку школьные оценки играют небольшую роль, мы склонны полагать, что иностранные языки помогают респондентам выделиться. Возможно, эту функцию чаще выполняет родной язык среди учащихся-билингвов.



Таблица 4

*Достоверность корреляций в логистических регрессиях
для использования иностранного языка в соцсетях и SMS-сообщениях*

	СоцСетиИно				SMSИно			
Случаи	5,3%				3,5%			
БАТЫРЕВО	–	– –	?	–	– –	– –	–	–
ДЕВОЧКА								
ОБРАЗПап								
ОБРАЗМам								
РАБОТПапы								
РАБОТМамы								
КНИГИ	+++	+++	+++	+++	+	+	+	+
ОЦЕНКИ								
ОЦЕНКАИно		+		+				
ШКОЛАРус	++	++						
СЕМЬЯРус			+++	+++			+	+
N	816	816	816	816	800	800	800	800
Псевдо R ² Макфаддена, %	22,0	23,5	23,2	24,8	16,4	16,6	17,4	17,7

Изучение информации о том, какие респонденты используют русский язык представлялось малоинтересным, потому что на нём пишет подавляющее большинство. Более информативным дополнением были данные о тех, кто пишет только по-русски (это примерно 40% респондентов).

Как предполагалось, русскоязычные-монолингвы – это, как правило, говорящие дома по-русски и обучавшиеся по-русски в начальной школе (см. табл. 5). В то же время бросается в глаза сильная отрицательная корреляция между школьными оценками (в частности, по русскому языку) и русскоязычным монолингвизмом. Это представляет собой яркий контраст с положительными корреляциями между школьными оценками и использованием чувашского или татарского языков. Дальнейшие исследования должны объяснить связь между школьной успеваемостью и моно/би/трилингвизмом в изученных районах, но мы считаем целесообразным связать данные корреляции со старательностью учащихся. Писать на чувашском или татарском языках, как правило, труднее, т.к. менее привычно, поэтому трудолюбивые учащиеся более склонны к этому, а более ленивые – склонны использовать «лёгкий» язык школьного обучения.



Т а б л и ц а 5

*Достоверность корреляций в логистических регрессиях
для использования только русского языка в соцсетях и SMS-сообщениях*

	СоцСетиТолькоРус				SMSTолькоРус			
Случаи	37,6%				41,8%			
БАТЫРЕВО								? ⁺
ДЕВОЧКА	? ⁺				? ⁺	+		
ОБРАЗПАП		? ⁻						
ОБРАЗАМАМ	? ⁺	? ⁺			? ⁺	? ⁺		
РАБОТПАПЫ								
РАБОТАМАМЫ								
КНИГИ								
Оценки	---		---		---		---	
ОЦЕНКАРУС		---		--		---		---
ШКОЛАРУС	+++	+++			+++	+++		
СЕМЬЯРУС			+++	+++			+++	+++
N	816	816	816	816	800	800	800	800
Псевдо R ² Макфаддена, %	13,5	13,2	14,6	14,3	14,5	14,7	14,5	14,9

Такие факторы, как пол респондента и уровень образования матери, не появляются с необходимой достоверностью в регрессиях. Интересно отметить, что данные возможные корреляции появляются в регрессиях с языком школьного обучения, а не с языками домашнего общения. Известно, что родители-чуваши с более высоким образованием чаще говорят по-русски с детьми [11], поэтому корреляция с уровнем образования матерей при отсутствии переменной «СемьяРус» представляется предположительным показателем роли более образованных матерей-чувашек в передаче русского языка детей. Напоминаем, что такая возможность была указана и в регрессиях по использованию чувашского языка, хотя с недостаточной уверенностью. Исследования по семейной передаче чувашского языка в данных районах должны подтвердить или опровергнуть такое предположение.

Следует отметить, что наличие работы у родителей не появилось в ни одной регрессии. Возможно, этот показатель недостаточно хорошо характеризует социально-экономическую ситуацию семей. Пол респондентов не играет практически никакой роли: нет значительных различий в выборе языков письменного общения по полу кроме не утверждённой тенденции к русскому монолингвистическому письменному общению.



Таким образом, посредством ряда логистических регрессий на результатах опроса 913 учащихся 7–11 классов Комсомольского и Батыревского районов Чувашии изучены факторы, связанные с выбором чувашского, татарского, русского и иностранных языков при общении в социальных сетях и SMS-сообщениях. Как правило, чувашский и татарский языки используют те, кто обучался на них в начальных классах школы и говорит на этих языках дома. Использующие только русский язык – это скорее те учащиеся, которые обучались на русском в начальной школе и используют его в домашнем общении. Среди татар семейная передача родного языка всеобщая, а обучение на родном языке очень распространено, при этом на уровне семейного общения влияние языка школьного образования особенно выделяется.

Чувашский или татарский языки выбирают, как правило, учащиеся с более высокими школьными оценками, а у ограничивающихся только русским языком оценки часто ниже средних. Полагаем, это связано с тем, что преимущественно русскоязычное образование прививает навыки чтения и письма на русском. Менее трудолюбивые учащиеся более склонны к «лёгкому» языку. Пишущих на иностранном языке мало: результаты исследования показывают редкое и элементарное использование иностранных языков. Корреляция с лучшими оценками, в частности по иностранным языкам, слаба. Их чаще используют русскоязычные учащиеся, возможно из-за того, что иностранные языки помогают выделиться среди сверстников: в отличие от чувашского или татарского, которыми многие одноклассники владеют лучше.

* * *

1. *Алос-и-Фонт Э.* Этноязыковая ситуация в сельских районах Чувашской Республики (по данным социолингвистического опроса школьников Моргаушского и Аликовского районов) // Исследование языковой ситуации в Чувашской Республике. – Чебоксары: ЧГИГН, 2015. – С.136–163.

2. *Алос-и-Фонт Э.* Отношение городского населения Чувашской Республики к использованию государственных языков (по данным социолингвистического опроса городских учащихся старших классов школ) // Исследование языковой ситуации в Чувашской Республике. – Чебоксары: ЧГИГН, 2015. – С.48–89.

3. *Алос-и-Фонт Э.* Этноязыковая ситуация в районных центрах и крупнейших сельских населённых пунктах Чувашской Республики (по данным социолингвистического опроса учащихся старших классов школ) // Исследование языковой ситуации в Чувашской Республике. – Чебоксары: ЧГИГН, 2015. – С.90–125.



4. *Алос-и-Фонт Э.* Этноязыковая ситуация в сельских районах Чувашской Республики (по данным социолингвистического опроса школьников Моргаушского и Аликовского районов)...

5. *Алос-и-Фонт Э.* Этноязыковая ситуация в сельских районах Чувашской Республики (по данным социолингвистического опроса школьников Моргаушского и Аликовского районов)...

6. *Алос-и-Фонт Э.* Устное владение языком своей национальности чувашами и татарами Комсомольского и Батыревского районов Чувашской Республики // Ашмаринские чтения: сб. трудов X Международных науч.-практ. конф. – Чебоксары: Чувашский университет, 2016. – С.89–98.

7. *Алос-и-Фонт Э.* Устное владение языком своей национальности чувашами и татарами Комсомольского и Батыревского районов Чувашской Республики...

8. *Товар-Гарсиа Э.Д.* Образовательные траектории школьников Ярославской области: дискриминантный анализ в сравнении с логистической регрессией // Социология: 4М. – 2013. – № 36. – С.65–93.

9. *Алос-и-Фонт Э.* Устное владение языком своей национальности чувашами и татарами Комсомольского и Батыревского районов Чувашской Республики...

10. *De Houwer A.* Parental language input patterns and children's bilingual use // Applied Psycholinguistics. – 2007. – Vol. 28. – № 3. – P.411–424.

11. *Алос-и-Фонт Э.* Этноязыковая ситуация в сельских районах Чувашской Республики (по данным социолингвистического опроса школьников Моргаушского и Аликовского районов).; *Alòs i Font H., Tovar-García E.D.* Factors affecting the knowledge of Chuvash language amongst school students in Russia // Sociolinguistic Studies. – 2015. – Vol. 9. – № 4. – P.345–372.



И.И. Баженов,

*канд. физ.-мат. наук,
доцент, профессор кафедры
информационных систем,
математики и естественно-
научных дисциплин
ГОУ ВО КРАГСиУ*

ОПЫТ ИСПОЛЬЗОВАНИЯ НАЦИОНАЛЬНОГО КОРПУСА РУССКОГО ЯЗЫКА НА УЧЕБНЫХ ЗАНЯТИЯХ ПО ДИСЦИПЛИНЕ «ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ»

Среди сетевых образовательных порталов особое место занимают ресурсы, содержащие базы знаний с вычислительными алгоритмами. Они интересны для организации исследовательской и справочно-информационной деятельности учащихся, студентов и педагогов. Несмотря на широкую известность, эти ресурсы нечасто используются в образовательном процессе в силу разных причин. Особое место среди сервисов с «умными вычислениями» занимает поисково-информационная система «Национальный корпус русского языка» (далее – НКРЯ).

Корпус русского языка (<http://ruscorpora.ru>) – это информационно-справочная система, представляющая собой собрание русских текстов в электронной форме. Корпус содержит множество типов письменных и устных текстов, представленных в русском языке. Основной корпус сегодня разбит на 2 группы: середина XVIII – середина XX в. (ранние тексты) и середина XX – начало XXI в. (современные письменные тексты). Сегодня НКРЯ имеет значительный объем (265,5 млн словоформ в основном корпусе и 364 млн словоформ в общем объеме всех корпусов). Корпус содержит дополнительную информацию о свойствах входящих в него текстов (так называемую разметку). Разметка – это главное отличие корпуса от простых коллекций (или библиотек) текстов. В Национальном корпусе русского языка в настоящее время используется 5 типов разметки: метатекстовая, морфологическая (словоизменительная), синтаксическая, акцентная семантическая. Система разметки постоянно совершенствуется.

Корпус предназначен для всех, кто интересуется самыми разными вопросами, связанными с русским языком: профессиональных лингвистов,



преподавателей языка, школьников и студентов, иностранцев, изучающих русский язык. Конечно, в первую очередь корпус нужен для обеспечения научных исследований лексики и грамматики языка, а также тонких процессов языковых изменений, происходящих в языке на протяжении сравнительно небольших периодов – от одного до двух столетий. Другая задача корпуса – предоставление всевозможных справок, относящихся к указанным областям (лексика, грамматика, акцентология, история языка). Современные компьютерные технологии многократно упрощают и ускоряют процедуры лингвистической обработки больших массивов текстов. С этой точки зрения основными потребителями национальных корпусов являются исследователи-лингвисты самого разного профиля. Однако круг пользователей корпуса вовсе не ограничивается профессиональными исследователями языка. Надёжные статистические данные о языке определённой эпохи или определённого автора могут интересовать литературоведов, историков и представителей многих других областей гуманитарного знания. Большое значение национальный корпус имеет и для преподавания языка в качестве родного или иностранного. Корпус является хорошей информационной системой и инструментом для проведения исследований школьников и студентов.

Автор данной статьи на протяжении 4 последних лет активно использует в образовательном процессе возможности НКРЯ в рамках преподавания учебной дисциплины «Информационные технологии», которая входит в учебный план всех направлений подготовки в Коми республиканской академии государственной службы и управления. Основной задачей дисциплины является развитие навыков применения интернет-технологий в профессиональной и повседневной деятельности обучающихся, формирование устойчивого опыта работы с различными сетевыми ресурсами, создание персонального интернет-пространства.

Форма проведения занятий с использованием НКРЯ может быть разнообразной. Это могут быть как простые упражнения на поиск данных, например частоты использования отдельных словообразований во временной шкале, так и более содержательные исследования, связанные с проверкой различных гипотез, например попытка связать историко-культурные явления в жизни российского государства с отражением их в письменности. Многие примеры заданий для проведения занятий могут быть получены при знакомстве с содержанием раздела «Публикации» на сайте НКРЯ (<http://ruscorpora.ru>).

В рамках курса «Информационные технологии» студентам предлагается выполнить лабораторную работу. При подготовке лабораторной работы автор использовал методические материалы проведённых ранее



авторских мастер-классов и семинаров (например, сетевой семинар «30х90» [1] и мастер-класс в образовательной онлайн-конференции на портале «Галактика Интел» [2]), а также материалы мастер-класса А. Матюшкина (Петрозаводск) [3]. Упомянутые мастер-классы и сама лабораторная работа сопровождались отдельными сайтами со свободным доступом. Мероприятия проводились как в очном, так и в дистанционном режиме.

В качестве примера приведём некоторые задания, входящие в лабораторную работу.

1. Проверим несколько гипотез, связанных с введением нормы, решающей использовать слово «кофе» в среднем роде:

- первая гипотеза: слово «кофе» стало часто использоваться в текстах в среднем роде. Исследуйте вопрос о частоте использования слова «кофе» в мужском и среднем роде;

- вторая гипотеза: мужской род слова «кофе» объясняется широким использованием в России в разные периоды двойственного слова «кофий». Проверьте, насколько жизненным является это предположение. Для этого постройте графики частоты использования слов «кофе» и «кофий» в течение двух последних столетий.

2. Сравните популярность использования в русском языке слов «компьютер» и «ЭВМ» в последнее столетие. Для этого постройте график частоты использования этих словоформ в указанный период. Используя график, сделайте вывод о динамике изменений и основных тенденциях. Постарайтесь найти объяснения обнаруженным явлениям.

3. Исследуйте популярность отдельных профессий (например, «лётчик»), используя частотный анализ по годам. Постарайтесь найти объяснения «пикам» популярности, связав их с историческими событиями в стране, культурными явлениями или другими причинами. Такой анализ популярности слов можно провести не только для профессий, но и самых различных словоформ, например, проанализировать динамику вторжения в язык заимствованных слов и выражений.

4. Проверьте, насколько оправдано решение об использовании преимущественно сине-голубых оттенков при экранизации романа Л.Н. Толстого «Анна Каренина» британским режиссёром Джо Райтом (экранизация 2012 г. с Кирой Найтли в главной роли). Связано ли выделение синего цвета в качестве доминирующего в кадрах этого фильма с оригинальным текстом романа? Для этого определите частоту использования слов «синий», «голубой» в романе Л.Н. Толстого по сравнению со словами, обозначающими другие цвета.

5. Проверьте, насколько оправдано представление о Ф.М. Достоевском как писателе, пишущем о бестелесном, духовном, психологическом,



мало говорящем о бытовом и земном. Для этого выделите подкорпус с его произведениями и проведите частотный анализ семантических признаков, выбрав полярные значения «еда и напитки» (предметные признаки) и «психическая сфера» (непредметные признаки). Аналогичное исследование проведите для других известных писателей.

6. Познакомьтесь с коллекцией исследований, примеры которых приведены в специальном разделе портала НКРЯ. Эта коллекция содержится на странице корпуса в разделе Публикации и *Studiorum* (последний раздел в настоящее время выведен из главного меню сайта).

Отчёт о выполнении исследований по сформулированным выше заданиям создаётся каждым студентом в виде отдельной страницы сайта (блога), который ведётся в рамках всего курса «Информационные технологии» как площадка представления отчётных материалов. На страничке сайта обучающиеся размещают скриншоты экрана монитора, подтверждающие результаты и формулируемые выводы. В лабораторной работе могут быть использованы коллективные формы работы, а также публичное (открытое для сетевого сообщества студентов) обсуждение результатов. Автором также планируется использование НКРЯ при проведении онлайн-викторин – командных соревнований по поиску информации.

В заключение следует сказать, что практически все студенты встречаются с этим сетевым ресурсом впервые. Большинство студентов высоко оценивают возможности и функционал корпуса. Работа над заданиями вызывает у них интерес, а отчётные материалы содержат развёрнутые и подробные исследования. Отзывы студентов о корпусе в основном положительные, многие считают, что полученные навыки работы в НКРЯ могут пригодиться в будущем.

Автором были также проведены мастер-классы для учителей, в том числе русского языка и литературы. Педагогов впечатляют возможности и объём базы знаний корпуса, однако обсуждение вопросов применимости электронного ресурса на практике в образовательном процессе вызывает у них определённый пессимизм. Это связано, по мнению автора, прежде всего с двумя факторами. Первый – профессиональный консерватизм, нежелание что-то менять в наработанных схемах и сложившихся методиках преподавания языковедческих дисциплин. Вторая причина – отсутствие необходимой компетенции учителя в организации образовательного процесса с использованием интернет-технологий. Учителя также часто говорят о плохой материально-технической базе, низкой скорости Интернета и т.п. Но в данном случае следует отметить, что «Национальный корпус русского языка» – это отечественный продукт, размещающийся на российских серверах, он удобен для работы и не требует высокоскоростного канала связи.



* * *

1. Авторский сетевой семинар «30х90» (ведущий И.И. Баженов). – URL: <https://sites.google.com/site/30x90no21>
2. Тренинг: 10 шагов по сервисам Web 2.0 (ведущий И.И. Баженов). – URL: <https://goo.gl/FEbMA7>
3. Мастер-класс: Вслед за словом (Электронные инструменты обучения исследованию художественного текста) (ведущий А. Матюшкин). – URL: <https://sites.google.com/site/vsledzaslovom>



Бартфай Чаба,
аспирант
Университета Сегеда
(Венгрия, г. Сегед)

ПРИЧИНЫ ВЫБОРА УДМУРТСКОГО ЯЗЫКА В СОЦИАЛЬНЫХ СЕТЯХ

Удмуртский язык относится к пермской ветви уральской языковой семьи. Его ближайшим родственником является коми язык. Удмуртский является пятым по величине уральским языком: в стране насчитывается по данным переписи России 2010 г., 552 299 этнических удмуртов. Большинство из них живут в Удмуртской Республике, где составляют 28,0% (410 584 человек) населения. Есть также небольшие диаспоры удмуртов в других регионах России, таких как Кировская область, Пермский край, Башкортостан, Татарстан, Марий Эл. Официальными языками в Удмуртской Республике являются русский и удмуртский, но на практике правительственные чиновники не говорят на удмуртском, и есть только несколько школ, где учат этому языку [1]. Однако удмуртский язык является одним из наиболее активных языков меньшинств в цифровом мире. Существует большое количество веб-страниц и блогов на удмуртском языке, видео на *Youtube*, а также немало пользователей, применяющих этот язык, не только в российской социальной сети *ВКонтакте*, но и *Facebook* и *Twitter*. И самый важный момент онлайн-присутствия языка – наличие его на *Web 2.0*, что обеспечивает возможность для пользователей создавать контент без каких-либо специальных знаний [2].

В настоящее время важность использования языка в цифровом мире выросла, т.к. методы и средства коммуникации резко изменились за последние несколько десятилетий [3]. Выживание языка сегодня зависит не только от количества, плотности и т.д. говорящего на нём населения, но и от того, насколько активно он применяется в сети и как он используется в повседневной жизни. А. Корнай отмечал, что существуют 2 вида смерти языка: обычная смерть и цифровая смерть языка. Сегодня наблюдаются оба эти явления, происходящие в основном независимо друг от друга. При этом мёртвый язык может быть активным в цифровой форме, например, латинский, идиш и древнегреческий можно использовать в онлайн-общении. И напротив, мёртвый в цифровой форме язык может быть далёк от выми-



рания в реальном мире, например, языки, которые не имеют письменности, но используются относительно большими группами людей. Тем не менее очевидно, что языкам, не имеющим цифровой формы, суждено рано или поздно исчезнуть [4].

Оригинальные факторы выбора языка до сих пор остаются неизменными, т.к. категория «особенности общения» не изменилась, только количество возможных значений этих свойств множилось. Четыре основных фактора выбора языка, по мнению Ф. Грожан, следующие: ситуация, тема, участники и функции сообщения [5]. В своём исследовании мы сосредоточились именно на этих факторах.

Таким образом, главной целью нашего исследования было найти какую-либо корреляцию между выбором языка и содержанием сообщения. Также во внимание были приняты тип сообщения (новости/запросы), цель сообщения и отправитель.

Методология.

Данная работа следует традиции исследований, основанных на анализе контента с точки зрения теоретической независимости. Описательный подход к исследованию позволяет сосредоточиться на выборе языка как естественного явления интернет-коммуникации и создания контента [6].

Процесс сбора материалов в основном по эксплуатации. После нахождения наиболее активных сообществ удмуртов в сети «ВКонтакте», посты были сохранены с даты создания сообщества. Наилучшим форматом для сохранения является *PDF*, поскольку он позволяет не только сохранить форматирование исходного материала, но и дополнительное визуальное содержание страниц, например, фотографии, ссылки на видео, аудио, заголовки, опросы и т.д. Это также удобно для копирования и вставки текста для дальнейшего анализа.

Фактический анализ контента был выполнен вручную с помощью базы данных *Microsoft Excel*, поскольку нет хорошего двуязычного инструмента анализа контента для русского и удмуртского языков. Были записаны следующие параметры для каждого поста:

- *ID*: специальный идентификатор поста; содержит имя сайта, название группы, период времени поста и номер (например: vk_udmurt_2016_04_001);
- *отправитель*: имя пользователя, который создал пост;
- *язык поста*: русский, удмуртский или смешанный;
- *область релевантности*: Удмуртия, Россия или весь мир;
- *главная тема*: культура, политика, экономика, религия, общественный интерес;
- *подраздел*: образование, язык, искусство, театр и т.д.;



- *тип поста*: новости или запрос;
- *тип контента*: текст, аудио, видео, фото;
- *источник*: оригинал или репост;
- *комментарии*: здесь упоминаются неясные моменты, информация, не подпадающая под вышеобозначенные параметры.

Материалы.

Для исследования были отобраны 5 наиболее активных удмуртских связанных сообществ путём сравнения их в различных аспектах, например, они должны были быть легко доступными для поиска *ВКонтакте* и поисковых системах, таких как *Google* или *Yandex*. Чем легче найти сообщества, тем удобнее пользователям присоединиться к ним. Кроме того, они должны были быть относительно активным в исследуемом периоде – в апреле 2016 года. Были проанализированы 326 постов. Материалы были собраны и архивировались в мае 2016 г., но все посты сообществ были сохранены с начала их работы для статистического анализа. Отмечается, как правило, один пост в день в каждой группе.

1. **УМОО «ШУНДЫ»**. Эта группа является активной с января 2008 г., посты в ней могут создавать только администраторы, при этом, помимо оригинального размещения содержания, они также делятся материалами с других страниц. Их собственные посты почти всегда написаны на удмуртском и посвящены культуре и общественным интересам. Иногда администраторы делятся информацией на русском языке, но только если новость содержит информацию, относящуюся к удмуртам. Почти каждый пост содержит вложение, изображение или видео, часто смайлы. Создатели группы пытаются создать в ней удмуртскую атмосферу, но если размещаемая информация имеет решающее значение для удмуртов, то язык источника не важен, при этом наблюдаются попытки уменьшить влияние русского языка путём написания нескольких слов во введении на удмуртском языке.

2. **«УДМУРТЛЫК»** – русско-удмуртское двуязычное сообщество пользователей, охватывающее все виды удмуртских этнически, культурно и лингвистически связанных контентов и пользователей. В этой группе состоят не только этнические удмурты, но и русские, которые живут в Удмуртской Республике, а также те, кто интересуется удмуртской культурой. Страница содержит в основном новости и информацию об Удмуртии, но здесь можно найти и другие различные типы постов: от театральных программ до объявлений о продаже котят. Это является следствием свободного редактирования: каждый член сообщества имеет право создавать контент в группе. Эта группа также является одним из первых и самых популярных удмуртских сообществ в «*ВКонтакте*».

Язык постов – в основном русский, но присутствует также много постов на удмуртском. Соотношение составляет около 2/3 постов на русском



и 1/3 – на удмуртском, но при этом отсутствует корреляция между темой и выбором языка.

3. «**Удмурт Кенеш**». Это сообщество было создано в феврале 2009 г. в качестве официальной страницы в *ВКонтакте* Удмуртского Кенеша, являющегося независимой группой удмуртских активистов, которые пытаются улучшить ситуацию удмуртов. По этой причине администраторы постоянно делятся новостями о своей деятельности, а также важной информацией для удмуртской этнической общины. Страница функционирует как совокупность удмуртских новостей и постов в социальных сетях, поэтому темы всех постов очень похожи.

4. «**Эктоника [UDM]**» – страница *ВКонтакте* популярной удмуртской музыкальной группы. Удивительно, но большинство постов здесь созданы на русском, наблюдается также смешанный язык. Все посты тесно связаны с деятельностью музыкальной группы (туры, выступления, фотографии и т.д.).

5. «**УДМУРТ ДУННЕ ГАЗЕТ / Газета УДМУРТСКИЙ МИР**» – страница в *ВКонтакте* самой долго издаваемой газеты на удмуртском. В исследуемый период «УДМУРТ ДУННЕ» был самым активным сообществом (в среднем 4 поста в день).

Содержание материалов здесь в основном оригинальное, и лишь в редких случаях на русском языке. Администраторы пытаются устранить любые влияния русского языка, т.к. нет таких тем (или так кажется), которые не могут быть изложены на удмуртском языке.

Таким образом, результаты исследования показывают, что нет корреляции между содержанием поста языком, на котором он создан. Каждая тема может быть выражена на обоих языках. Также отсутствует тенденция, которая позволила бы предположить, что некоторые темы будут часто публиковаться в этих группах на русском или удмуртском языках. Даже язык источника не имеет никакого отношения к выбору языка поста. Самым важным фактором при выборе языка, на наш взгляд, является сознательный выбор индивида. Это явление также подтверждается тем фактом, что существует лишь небольшое количество смешанного использования языков или сбалансированного двуязычия. Наиболее частыми являются 2 крайности: использование либо только русского, либо только удмуртского языка. Причиной этого может быть формат и характер текстов, т.к. в социальных сетях использование языка может быть похоже на разговорную речь, но в этом случае – скорее на хорошо сформированные тексты [7].

Число постов и активность сообщества говорят о том, что удмуртский – не вымирающий язык, по крайней мере, в цифровом мире [8]. Тем не менее остаётся нерешённым вопрос о том, сколько пользователей актив-



но в этих сообществах, т.к. число подписчиков не равно числу пользователей, которые читают каждое сообщение. Кроме того, неясно, активна ли группа только в Интернете или в повседневной жизни тоже.

Целью нашей работы было дать некоторое представление о использовании удмуртского языка в сети «ВКонтакте», однако остаётся много вопросов, связанных с существованием удмуртского языка в виртуальном пространстве. Существует большое количество онлайн-контента, интернет-сообществ, блогов и других *Web 2.0* страниц, видеороликов в *Youtube* (например, есть более тысячи телепередач, сотни эпизодов ток-шоу, а также множество программ от телеканала «Мынам Удмуртия») на удмуртском языке. С этим количеством материала было бы интересно исследовать удмуртский виртуальный лингвистический ландшафт [9]. Более того, было бы интересно сравнить результаты исследования с традиционным лингвистическим ландшафтом удмуртских городов в реальной жизни, т.к. это может дать представление о том, как язык может развиваться в цифровом пространстве, и в то же время быть под угрозой в реальном мире [10].

* * *

1. *Salánki Zsuzsanna*. Az udmurt nyelv mai helyzete. – PhD Thesis, 2007.
2. *Pischlöger Christian*. Udmurtness in Web 2.0: Urban Udmurts Resisting Language Shift // *Finnisch-Ugrische Mitteilungen*. – 2014. – № 38. – P.143–162.
3. *Crystal David*. Internet Linguistics: a Students Guide. – Routledge, 2011. – P.19–57.
4. *Kornai András*. Digital Language Death. 2013. PLoS ONE 8(10): e77056. doi:10.1371/journal.pone.0077056
5. *Grosjean Francois*. Life with two languages. – Cambridge: Harvard University Press, 1982. – P.135–145.
6. *Drisko James, W.-Maschi Tina*. Content Analysis. – Oxford University Press, 2016.
7. *Crystal David*. Op. cit.
8. *Kornai András*. Op. cit.
9. *Ivkovic Dejan, Lotherington Heather*. Multilingualism in Cyberspace: Conceptualising the Virtual Linguistic Landscape // *International Journal of Multilingualism*. – 2009. – № 6/1. – P.17–36.
10. *Landry Rodrigue–Richard Bourhis*. Linguistic landscape and ethnolinguistic vitality // *Journal of Language and Social Psychology*. – 1997. – № 16. – P.23–49.



А.В. Болтачев,
независимый исследователь
(г. Киров)

РАЗРАБОТКА СЕРВИСА ДЛЯ РЕДАКТИРОВАНИЯ МАРИЙСКО-РУССКОГО, РУССКО-МАРИЙСКОГО СЛОВАРЕЙ

Определение и структура словаря.

Словарь представляет собой инструмент для понимания слов (и коротких фраз) на одном языке путём объяснения их на другом. Словари обычно базируются на словнике (списке слов); при этом каждой вокабуле (слову из словника) соответствует словарная статья.

Словарные статьи обычно имеют иерархическую структуру, например:

- 1) слово – перевод;
- 2) слово – несколько переводов;
- 3) слово – несколько переводов с примерами использования.

Более сложные статьи могут включать:

- 1) пояснения (например, «биологический термин»);
- 2) перекрёстные ссылки между статьями;
- 3) дополнительную информацию, такую как позиция ударения в слове;
- 4) полисемию, омонимию/омографию (когда одно слово имеет много значений, информация о них представляется в иерархической структуре).

Создание словарей.

Словари обычно создаются в рабочих группах из нескольких человек.

Возможна следующая организация работы группы:

1. *Выбор источников.* Определение набора ранее опубликованных словарей, текстовых корпусов и т.д., которые можно взять за основу. Словари могут иметь, в том числе, обратное направление перевода.

2. *Определение словника.* Словник может создаваться отдельно и использоваться как основа для многих словарей, например для перевода с русского на многие финно-угорские языки.

3. *Разработка структуры,* т.е. определение того, насколько сложной статья может быть.



4. *Разделение работы между членами группы.* Возможны 2 стратегии:

- по буквам, когда один работник берёт все слова, начинающиеся с одной буквы;

- по темам, когда один работник берёт, например, все термины национальности. Эта стратегия реализуется сложнее, т.к. списки терминов должны быть заранее подготовлены;

5. *Вычитка и корректура.* Эта стадия направлена на улучшение качества словаря.

Работая над словарём, каждый работник использует как предопределённые условия и инструкции, так и собственный опыт и знания для того, чтобы формировать словарные статьи. В любом случае, процесс создания словаря не может быть автоматизирован полностью, и доверие к его материалу основывается на доверии к его авторам и редакторам.

Инструменты.

Поскольку под словарём понимается обычно печатная книга, инструменты (компьютерные программы), применяемые для создания словарей, могут быть разделены на 2 группы:

1. Программы, типичные для создания книг, включая визуальные редакторы текста, такие как *Microsoft Office Word*.

а) достоинства:

- высокая распространённость;
- свобода действий – фактор простоты в работе;

б) недостатки:

- свобода действий – фактор, ведущий к большому разнообразию статей различных авторов, а также к большому объёму работы.

2. Специализированные сервисы:

а) достоинства:

- возможность большой специализации для конкретных целей;
- богатый функционал;
- исключение некоторых ручных стадий;

б) недостатки:

- отсутствие свободы действий – чувство ограниченности.

В настоящее время продолжается работа по созданию сервиса для марийско-русского, русско-марийского словарей.

Марийско-русские и русско-марийские словари.

Некоторые из марийско-русских и русско-марийских словарей были созданы в Марийском научно-исследовательском институте языка, литературы и истории им. В.М. Васильева. Среди них есть 10-томное издание марийского словаря, который переводит слова с лугово-восточномарийского и горномарийского языка на русский.



Словарь вначале появился в форме 10 книг. Позже он был оцифрован, для чего был произведён примерно следующий объём работы:

- 1) разделение всех книг на отдельные страницы;
- 2) сканирование каждой страницы;
- 3) распознавание текста на каждой странице;
- 4) объединение текста (на этой стадии словарь стал одним файлом в формате «текст с форматированием», легко конвертируемом в *HTML*);
- 5) вычитка;
- 6) ручной перевод в формат разметки, похожий на *DSL* (*DSL* – формат, используемый популярным программным обеспечением для словарей, таким как *ABBYY Lingvo* и *GoldenDict*).

Работа по оцифровке словаря является частью большего процесса (по сути, его завершающим этапом) и фокусируется на:

- 1) анализе словаря и формата его содержимого;
- 2) разработке сервиса создания и редактирования словарей.

Структура и анализ содержимого.

Формат, похожий на *DSL* (далее называемый *FULabDSL*), состоит в основном из тегов в квадратных скобках. Пример статьи представлен на *рис. 1*.

```

[udar] ó [/udar]
I.
[trn]    1[/trn]
[mhr]    1          1[/mhr] [rus]          1[/rus]
[mhr]    2          1[/mhr] [aut]          2
[/aut] [rus]          2[/rus]
II.
[trn] [p]    .[/p]          [/trn]

```

*Рис. 1. Пример формата *FULabDSL**

Разработка сервиса.

Разработка парсера.

Первая задача состояла в том, чтобы разработать парсер для формата *FULabDSL*, в который исходный материал поступает после процесса ручной обработки. Для программирования был выбран современный язык программирования *Clojure*, т.к. он имеет хорошее быстродействие и удобен для разработки систем обработки данных. Разработанный парсер состоял из двух частей:

- 1) токенизатора, который переводил отдельные символы, такие как «II.» или «[trn]», в токены;



2) самого парсера, который применял структуру («грамматику») ко всем статьям, чтобы проверить их соответствие данной структуре.

Для парсера грамматика играет одну из центральных ролей. Грамматика – это один из формальных языков (а не грамматика марийского или русского). Пример одной из грамматик словаря представлен на *рис. 2*.

Синтаксис – *Lisp*-подобный, символы имеют следующие значения:

- 1) *Plus* – одно и более включений;
- 2) *Seq* – последовательность;
- 3) *Char* – терминальный символ;
- 4) *Star* – ноль и более включений;
- 5) *MayBe* – ноль или одно включение.

```
(Plus
  (Seq
    [
      (Char "trn")
      (Star (Seq [
        (Char "mhr")
        (MayBe (Char "aut"))
        (Char "rus")
      ]))])])
```

Рис. 2. Пример грамматики FULabDSL

Таким образом, разработанный парсер очень близок к широко известному парсеру регулярных выражений. Разница в том, что в отличие от парсера регулярных выражений в данном парсере токены могут иметь нагрузку.

Грамматику следует читать таким образом: словарная статья состоит из одного или более включений переводов, где за каждым переводом следует один или более примеров, где примеры включают марийский текст, (возможно) автора и русский текст.

Парсер был далее разделён на две части, из которых каждая была опубликована на *GitHub*[^] и *Clojars*:

1. *RegExpForObj* – парсер для «объектифицированных» регулярных выражений библиотека общего назначения [?];

2. *fulabdsl* – парсер для словарей естественных языков в формате *FULabDSL* сам парсер [?].

Парсер сейчас является основной завершённой частью работы, на которой основываются последующие разработки.



Веб-интерфейс.

Далее, началась работа над проектом веб-интерфейса. Были сразу определены следующие требования к нему:

- 1) многопользовательский;
- 2) полная история редактирования; ничего не должно исчезать;
- 3) «богатый» интерфейс; используется перетаскивание мышью и подобные функции.

На данный момент разработана только часть, отвечающая за сохранение истории редактирования; редактор статьи находится в работе.

Перспективы.

Основной перспективой проекта является завершение сервиса, что позволит возможным сделать редактирование словарей в онлайн-режиме:

- 1) возможность пользователям иметь свои собственные версии статей;
- 2) возможность «редактирования толпой»;
- 3) возможность хорошо организованным группам качественно выполнять свою работу.

Одним из расширений, также находящимся в разработке, является программа, которая создаёт черновую копию русско-марийского словаря из марийско-русского. Эта программа называется «перевёртыш», она использует русский *HunSpell* как грамматический движок и формирует черновик обратного словаря, собирая все включения конкретной вокабулы и её производных слов в переводы и примеры оригинального словаря.



Л.А. Бускунбаева,

*канд. филол. наук, научный сотрудник
лаборатории лингвистики
и информационных технологий
Института истории, языка и литературы
Уфимского научного центра РАН
(Республика Башкортостан, г. Уфа)*

З.А. Сиразитдинов,

*канд. филол. наук, ст. научный сотрудник,
зав. лабораторией лингвистики
и информационных технологий
Института истории, языка и литературы
Уфимского научного центра РАН
(Республика Башкортостан, г. Уфа)*

А.Ш. Ишмухаметова,

*мл. научный сотрудник лаборатории
лингвистики и информационных
технологий отдела языкознания
Института истории, языка и литературы
Уфимского научного центра РАН
(Республика Башкортостан, г. Уфа)*

СОСТАВ И СТРУКТУРА КОРПУСА ПУБЛИЦИСТИКИ БАШКИРСКОГО ЯЗЫКА

Лаборатория лингвистики и информационных технологий (далее – Лаборатория) отдела языкознания Института истории языка и литературы Уфимского научного центра РАН активно развивает одно из приоритетных направлений современной лингвистики – корпусную лингвистику. На сегодняшний день сотрудниками Лаборатории разработаны и размещены в Интернете 3 лингвистических корпуса – корпус прозы, корпус публицистики и корпус фольклора.

Работа над корпусом публицистики башкирского языка была начата в 2013 г. и продолжается по сей день, его база данных постоянно пополняется новыми материалами.

В процессе выполнения данного проекта была проделана колоссальная работа по подборке текстов СМИ, выходящих в Республике Башкортостан на башкирском языке, редактированию и вводу публицистических текстов в базу данных корпуса.



Материалы, предоставленные редакциями республиканских, районных и городских газет и журналов, сотрудниками Лаборатории переводятся в нужный формат для дальнейшей работы в корпусе и редактируются. Каждая статья сохраняется в отдельном файле, удаляются те части, которые не несут информационной нагрузки, такие как объявления, программы телепередач, пословицы, поговорки либо изречения, состоящие из одного-двух предложений и употреблённые прежде всего для заполнения газетного/журнального пространства.

После ввода текста в базу данных корпуса каждой статье приписываются следующие **метаданные**:

- 1) название СМИ;
- 2) дата издания (*для газетных текстов – год, месяц, число, для журнальных – год, месяц, номер выпуска*);
- 3) автор статьи;
- 4) название статьи;
- 5) тематика текста: искусство, культура и литература; красота и здоровье; кулинария; медицина; наука и техника; образование; политическая и социальная жизнь; природа и путешествия; психология; религия; сельское хозяйство; спорт; философия; частная жизнь; экономика;
- 6) жанр текста: интервью, беседа; обзор печати; письма; поздравления; рецензия; советы; статья, очерк, репортаж, обозрение; художественные произведения (эссе, фельетон, рассказ, стихи, эпиграммы);
- 7) объём текста (число словоформ);
- 8) тип носителя (газета, журнал).

Ограничив область поиска и отметив отдельно либо в совокупности нужные параметры, по первым шести пунктам предоставляется возможность получения интересующейся информации (см. *рис. 1*).

На данный момент корпус публицистики включает в себя 11 441 статью общим объёмом в 6 697 226 словоупотреблений из следующих газет и журналов:

- «Ағизел» (2010–2013), 1148 текстов;
- «Ашказар» (2009–2013), 2799 текстов;
- «Башкортостан кызы» (2004–2011), 944 текста;
- «Йэшлек» (2010), 2008 текстов;
- «Киске Өфө» (2009–2011), 3 866 текстов;
- «Табын» (2009–2013), 676 текстов.

Публицистический корпус башкирского языка также оснащён лингвистическими разметками – морфологической и частично семантической.

Морфологическая разметка даёт информацию о морфологических характеристиках слова (лемма, часть речи, грамматические категории) и по-

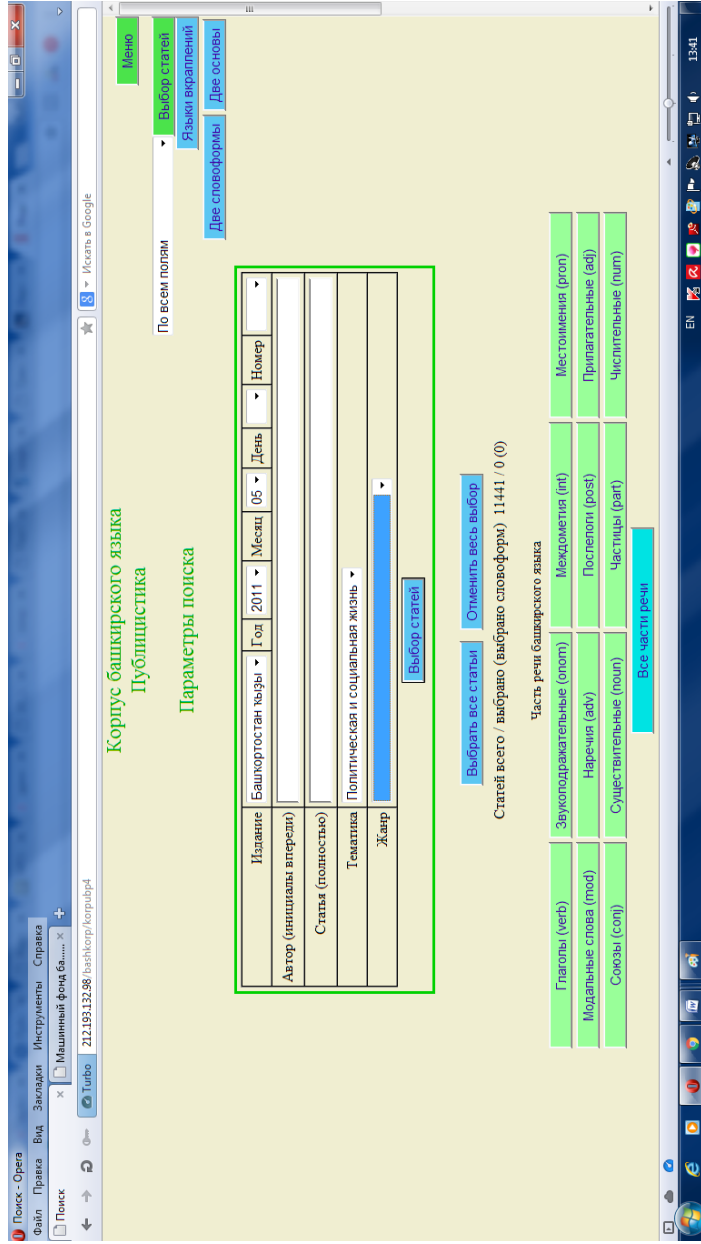


Рис. 1. Интерфейс корпуса публицистики башкирского языка

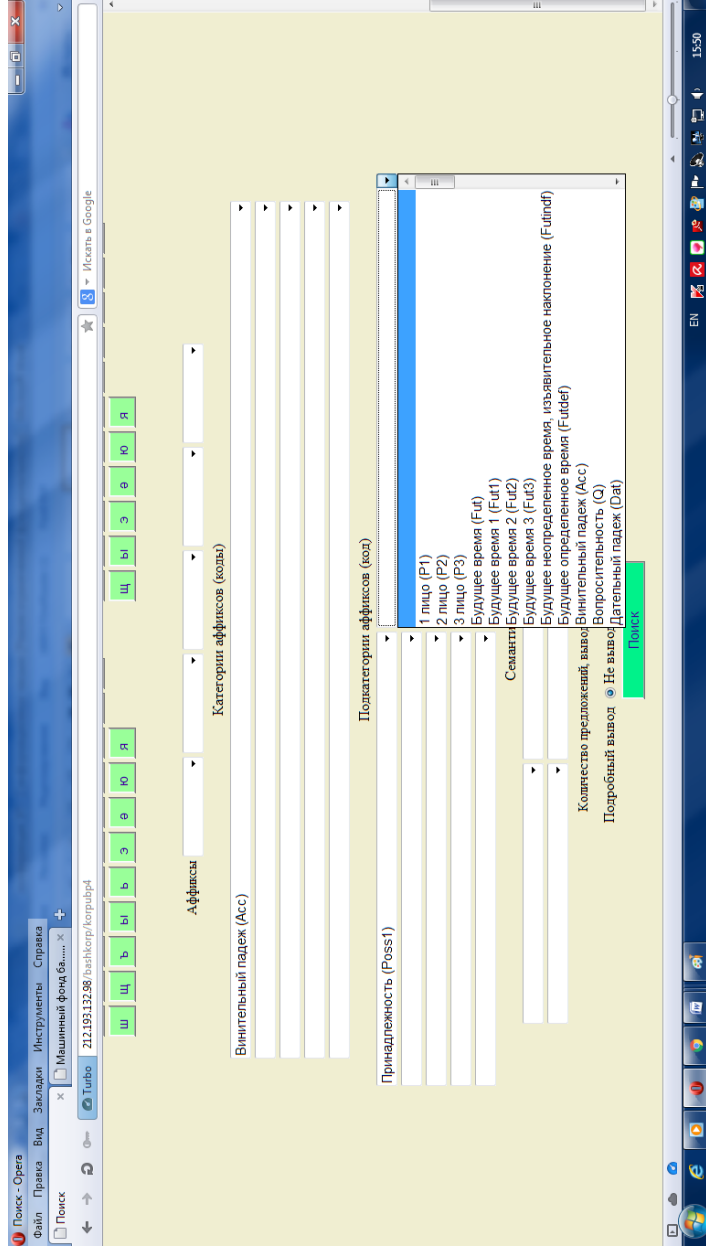


Рис. 2. Выбор морфологических характеристик слова



звляет производить выбор нужных пользователю слов, форм слов и конструкций [1] (см. рис. 2).

Семантическая разметка в корпусе публицистики разработана частично и включает в себя тематический класс и словообразовательные характеристики лексемы. Данная разметка позволяет выдавать примеры употребления конкретного слова в определённом значении, что существенно облегчает поиск нужной информации в корпусе публицистики, содержащего многомиллионные словоупотребления [2].

Корпус публицистики содержит богатый материал для исследования башкирского языка, позволяющий оценивать современное состояние языка, фиксировать неологизмы, диалектные особенности той или иной территории.

Данные корпуса публицистики активно привлекаются при написании дипломных и диссертационных работ, подготовке научных статей, разработке различного рода словарей, в том числе и десятичного «Академического словаря башкирского языка» [3], и позволяют многократно упростить и ускорить процесс сбора и обработки информации.

Рассмотренный корпус позволяет оснастить словарные статьи богатым иллюстративным материалом. Например, при запросе примеров на слово *теи* (зуб) корпус публицистики выдаёт 802 контекста. Из всего многообразия примеров составителю остаётся подобрать наилучший вариант к тому или иному значению конкретного слова.

* * *

1. *Сиразитдинов З.А., Бускунбаева Л.А., Ишмухаметова А.Ш., Ибрагимова А.Д.* Информационные системы и базы данных башкирского языка. – Уфа, 2013. – С.49–51.

2. *Бускунбаева Л.А., Сиразитдинов З.А.* О проблемах создания национального корпуса башкирского языка // Современное казахское языкознание: актуальные вопросы прикладной лингвистики: материалы Международной науч.-теорет. конф., посвящённой 75-летию известного учёного профессора Жубанова Аскара Кудайбергенулы. – Алматы, 2012. – С.54–58.

3. Академический словарь башкирского языка. – Уфа, 2011. – Т.I; 2011. – Т.II; 2012. – Т.III; 2012. – Т.IV; 2013. – Т.V; 2014. – Т.VI; 2015. – Т.VII; 2016. – Т.VIII.



Л.В. Бусыгина,

канд. филол. наук, ведущий

научный сотрудник БУУР

«НИИ национального образования»

(Удмуртская Республика)

Ю.Т. Байтерякова,

ст. научный сотрудник БУУР

«НИИ национального образования»

(Удмуртская Республика)

Н.П. Боталова,

ст. научный сотрудник БУУР

«НИИ национального образования»

(Удмуртская Республика)

ЭЛЕКТРОННЫЕ УЧЕБНИКИ КАК СРЕДСТВО ИНТЕНСИФИКАЦИИ И ПОВЫШЕНИЯ КАЧЕСТВА ОБУЧЕНИЯ УДМУРТСКОМУ ЯЗЫКУ

В соответствии с новым порядком формирования федерального перечня учебников, а также с требованиями федерального государственного образовательного стандарта начального общего образования, направленными на деятельностный, личностно-ориентированный подход, формирование информационной культуры субъектов образовательного процесса, в настоящее время актуализируется проблема создания электронных форм учебников (далее – ЭФУ) по образовательным предметам, в том числе по предметам «Родной язык» и «Литературное чтение на родном языке». Данная задача решается БУУР «Научно-исследовательский институт национального образования» в рамках государственного задания Министерства образования и науки Удмуртской Республики. На данный момент для уровня начального общего образования созданы электронные учебники «Удмурт кыл» («Удмуртский язык (родной)», автор Ю.Т. Байтерякова), «Зеч-а, бур-а, удмурт кыл!» («Здравствуй, удмуртский язык! (неродной)», автор Н.П. Боталова) и «Лыдзён книга» («Книга для чтения на удмуртском языке», автор Л.В. Бусыгина).

ЭФУ представляют собой электронные издания, соответствующие по структуре, содержанию и художественному оформлению печатной форме



учебника, включающие в себя мультимедийные элементы и интерактивные ссылки, расширяющие и дополняющие содержание учебника. Они выполняют все функции, присущие бумажным учебникам (информационную, мотивационную, воспитывающую и др.). Созданы с учётом таких требований, как преемственность материала учебного предмета, единство методической и дидактической основы, а также единство психолого-педагогических подходов. При этом ЭФУ, вовлекая в процесс обучения иные, в отличие от обычного учебника, возможности человеческого мозга, в частности, слуховую и эмоциональную память, а также используя компьютерные объяснения, способствует достижению тех методических целей, которые не в полной мере могут быть реализованы с помощью печатной формы учебника.

Электронные учебники имеют удобный и интуитивно понятный интерфейс, с ними можно работать в онлайн- и офлайн-режимах. Инструменты и сервисы для работы с электронным учебником позволяют оперативно переходить в разделы меню, легко и быстро находить необходимый материал.

ЭФУ содержат различные опции: «Кыллюкамъёс» («Словари»), «Каталог», «Закладкаос» («Закладки»), «Заметкаос» («Заметки»), «Журнал», «Кулэ баме потон» («Переход на нужную страницу»), «Поиск» (поиск по отдельным словам и фразам), «Учебнике пыртэм пусъёс» (условные обозначения, используемые в учебнике), «Настройки», «Учебник сярысь» («Об учебнике») и «Печатной учебник» (переход на печатную версию учебника).

В раздел «Кыллюкамъёс» включены словари, помещённые в конце печатного учебника. Но поскольку электронные учебники – это открытая система, есть возможность введения в них и дополнительных словарей.

В «Каталоге» собраны сводные данные учебника: презентации, аудио-, видеоматериалы, тренажёры, тесты и др.

Важный элемент ЭФУ – закладки. Данный инструмент даёт возможность сохранять ссылки на избранные разделы электронного учебника в отдельном списке и осуществлять при необходимости быстрый переход к ним.

В разделе к каждому уроку пользователь может оставить какие-либо заметки «Заметкаос». Данный раздел позволяет быстро перейти к любой заметке. В разделе заметок учащиеся могут оставлять свои творческие работы, чтобы впоследствии разместить их в свои портфолио, а также зафиксировать оценку своей работы на занятиях. На основе подобного рода обратной связи учитель сможет диагностировать процесс обучения на начальной и промежуточной, а не только на конечной стадии. Таким образом заметки используются в качестве техники формирующего оценивания.



Раздел «Журнал» предназначен для просмотра результатов пройденных контрольных тестов. Результаты прохождения контроля записываются в виде «5 шонер 8 пѳлысь», т.е. «5 правильных из 8». Отметки ставит учитель в соответствии с критериями, выработанными им совместно с учащимися.

Инструмент «Поиск по номеру страницы» предоставляет возможность выполнять быстрый поиск и переход на определённые страницы учебника.

Опция «Поиск» позволяет выполнять поиск по отдельным словам и фразам.

«Пуктэтѳёс» – настройки приложения. В ЭФУ предусмотрена возможность изменения размера текста учебника. Всего доступно 5 размеров текста: очень маленький, маленький, средний, большой и очень большой.

Основной раздел меню – это раздел «Учебник», позволяющий перейти непосредственно к содержанию учебника.

Электронный учебник представляет собой многослойную структуру.

Первый слой учебника составляет основной материал, соответствующий материалу печатного учебника, далее идут тесты в режиме «Тренажёр» и «Контроль» для проверки и самопроверки знаний и дополнительные материалы (аудиозаписи, видеоматериалы, презентации, галереи картин и др.).

Основной материал усилен дополнительными возможностями: в полноэкранном режиме открываются иллюстрации учебника, увеличиваются шрифты, а также при наведении курсора на новые слова, выделенные в тексте жирным шрифтом и звёздочкой, во всплывающем окне появляются толкования слов.

Для проверки и самопроверки знаний во всех линиях электронных учебников используются 5 типов заданий: «выбор вариантов из списка», «упорядочивание», «соответствие», «ввод значения» и «заполнение пропусков». Все они доступны как в режиме тренажёра, так и в режиме контроля. Включённые в учебники тренажёры к каждому уроку обеспечивают тренировочную учебную деятельность, контрольные тесты дают возможность за короткое время эффективно и объективно проводить проверку знаний учащихся по основным разделам учебника. Результаты ответов, автоматически отражающиеся в журнале электронного учебника, позволяют учащемуся, учителю, родителям наглядно видеть результаты освоения материала учащимися по каждой теме и в случае нежелательных оценок добиваться лучших результатов.

Дополнительные материалы в каждой линии электронных учебников различны. Так, в учебнике «Лыдзѳн книга» («Книга для чтения») (Т.С. Ива-



нова, И.Ф. Тимирязнова)) содержится достаточно много презентаций по биографиям писателей, аудиозаписей произведений, вошедших в учебники (скороговорки, песен, сказок), а также текстов произведений русских и финно-угорских писателей, нацеленных на воспитание культуры межнациональных отношений, видеороликов (сказок и мультфильмов), галерей картин российских художников и т.д. Работа с этими материалами способствуют формированию техники чтения и приёмов понимания текста, развитию устной и письменной речи (в том числе значительному обогащению словаря), повышению интереса к самому процессу чтения, появлению потребности к чтению, а также развитию творческих способностей учащихся.

Данные материалы, оформленные в формате гипертекстовых ссылок, во-первых, отвечают таким предъявляемым к учебникам требованиям, как структурированность, удобство в обращении. Во-вторых, гипертекстовая подача материала позволяет учащимся неоднократно возвращаться к презентациям, аудио- и видеозаписям, всплывающим подсказкам. Такого рода возможности выполняют роль «бесконечно терпеливого учителя», предоставляя практически неограниченное количество разъяснений, повторений, подсказок и др. Благодаря наличию гипертекстовых ссылок электронный учебник значительно экономит время учащегося, затрачиваемое на поиски учебного материала, а также при повторении неизвестных или забытых понятий.

На уроках удмуртского языка работа по развитию речи, обогащению словарного запаса учащихся, формированию орфографически грамотного письма является одной из главных задач обучения. Поэтому вполне естественно появление в дополнительных материалах ЭФУ «Удмуртский язык» интерактивных материалов, объединённых в рубрики «Словарная работа» («Будэты кылузырлык-тэ!»), «Аудиодиктанты» («Гожтомы диктант»), а также различных словарей (орфографический, толковый, словари синонимов, антонимов, родственных слов). Работа в рубрике «Будэты кылузырлык-тэ!» предполагает выполнение разнообразных заданий, в частности, заданий на обогащение словарного запаса учащихся (работа со словарём, уточнение словаря учащихся, усвоение значений новых слов), выработку орфографической грамотности (правописание новых слов), развитие речи (включение новых слов в словосочетания, предложения и тексты). В рубрике «Аудиодиктант» представлен материал для самопроверки знаний, навыков грамотного письма в условиях полной самостоятельности учащихся. Суть аудиодиктантов состоит в записи воспринимаемых на слух слов, словосочетаний, предложений и текстов.

В учебниках «Зеч-а, бур-а, удмурт кыл!» для русскоязычных учащихся сделан акцент на формирование у учащихся произносительных на-



выков. Например, в ЭФУ для 1 класса включено более 280 аудитивных заданий. Учащиеся слушают аудиозаписи, повторяют за голосом диктора удмуртские слова, выражения и предложения, отвечают на вопросы, кратко пересказывают содержание текстов, находят необходимую информацию.

Новое средство обучения позволяет оживить урок благодаря прослушиванию песен. В начальной школе в содержание ЭФУ включено 24 песни для разучивания.

Начиная со 2 класса в ЭФУ введены грамматические презентации. Во всех электронных учебниках «Зеч-а, бур-а, удмурт кыл!» присутствуют и русско-удмуртский, и удмуртско-русский словари.

Таким образом, электронные учебники по удмуртскому языку и литературе дают уникальные возможности интенсификации образовательного процесса:

- незамедлительная обратная связь;
- компьютерная визуализация учебной информации;
- архивное хранение достаточно больших объёмов информации с возможностью её передачи, а также лёгкого доступа и обращения пользователя;
- автоматизация процессов информационно-поисковой деятельности, возможность многократного повторения отдельных элементов ЭФУ.

Интерактивная форма подачи материала в ЭФУ поддерживает на уроке высокую активность каждого ребёнка, вызывает интерес учащихся к предмету, активизирует их познавательную деятельность, даёт возможность решить одновременно несколько задач: изучить новый материал, закрепить, углубить знания, за короткое время провести контроль и самоконтроль знаний. Уже в младших классах обучающийся научится ориентироваться в информационно-коммуникационных технологиях (ИКТ), получать знания с помощью инструментов ИКТ. Более того, сегодня, в условиях введения профессионального стандарта педагога, умение работать с ЭФУ является необходимой составляющей профессионального уровня педагогических работников начального общего образования.



УДК 378.147

М.А. Габова,

*канд. пед. наук, доцент кафедры
начального и дошкольного образования
Сыктывкарского государственного
университета имени Питирима Соркина*

В.Ф. Поберезкая,

*канд. пед. наук, доцент кафедры
начального и дошкольного образования
Сыктывкарского государственного
университета имени Питирима Соркина*

ТЕХНОЛОГИИ СОЗДАНИЯ И ИСПОЛЬЗОВАНИЯ ЭЛЕКТРОННЫХ ОБРАЗОВАТЕЛЬНЫХ РЕСУРСОВ НА РОДНОМ ЯЗЫКЕ В НАЧАЛЬНОЙ ШКОЛЕ

Федеральный государственный образовательный стандарт начального общего образования предписывает, что образовательное учреждение должно быть обеспечено учебниками с электронными приложениями, учебно-методической литературой и материалами по всем учебным предметам основной образовательной программы начального общего образования. Образовательное учреждение должно иметь доступ к печатным и электронным образовательным ресурсам (далее – ЭОР), в том числе к электронным образовательным ресурсам, размещённым в федеральных и региональных базах данных ЭОР [1].

В соответствии с ГОСТ Р 53620-2009, ЭОР – образовательный ресурс, представленный в электронно-цифровой форме и включающий в себя структуру, предметное содержание и метаданные о них [2].

Применение ЭОР в образовательном процессе в сочетании с системами управления обучением и образовательным контентом позволяет эффективно реализовать:

- организацию самостоятельной когнитивной деятельности учащихся;
- организацию индивидуальной образовательной поддержки учебной деятельности каждого учащегося преподавателями;
- организацию групповой учебной деятельности с применением средств информационно-коммуникационных технологий (далее – ИКТ) [3].

Однако практика показывает, что существующих ЭОР для реализации задач обучения, воспитания и развития обучающихся недостаточно,



и у учителей возникает необходимость самим разрабатывать ЭОР. В условиях недостаточности готовых ЭОР на родном языке учитель может сам разработать ЭОР, используя для этого доступные средства ИКТ.

При разработке ЭОР можно придерживаться следующего алгоритма (см. *табл.*).

Т а б л и ц а

Алгоритм работы над ЭОР

Этап	Содержание этапа
1. Предварительная работа	Формулировка исходной идеи. Оформление документации на разработку ЭОР. Оценка существующих элементов. Составление перечня необходимых и наличествующих специалистов
2. Сбор необходимой информации	Анализ потребностей. Выделение главной дидактической цели. Обоснование необходимости и того нового, что внесёт продукт
3. Подготовка содержания	Выделение дидактических подцелей. Составление плана. Представление содержания в форме модулей
4. Дизайн	Разработка концепции. Выбор медиа (звук, изображения, видео). Написание сценария. Детальный дизайн + подключение интерактивности
5. Производство	Программирование и оцифровка содержания. Создание изображений, звука и т.п. Компоновка готовых материалов в модули. Наладка навигации по продукту
6. Тестирование	Тестирование и оценка продукта
7. Юридический	Регистрация и сертификация ЭОР
8. Поддержка	Техническая и методическая поддержка

При самостоятельной разработке ЭОР для учителя доступными являются такие средства ИКТ, которые имеют интуитивно понятный интерфейс, возможность работы в режиме как онлайн, так и оффлайн. Нами были апробированы различные программные продукты, из которых мы рекомендуем к использованию следующие: *MS PowerPoint, Adobe Flash Professional SC6, Zooburst, PhotoPeach, Prezi, Google, PowToon, LearningApps.*

Среди недостатков большинства представленных продуктов можно отметить отсутствие русификации, невозможность сохранить созданный



ресурс на компьютере, необходимость постоянного подключения к Интернету, ограниченность функций в бесплатном режиме доступа. Всё это затрудняет использование этих инструментов учителем начальных классов.

По нашему мнению, для учителя более удобным инструментом станет программа для создания презентаций *MS PowerPoint*. Эта программа имеет возможности создания мультимедийного контента с использованием текстовой, аудио-, видеоинформации, анимации, а также интерактивного рабочего пространства за счёт триггеров, гиперссылок и макросов. Эти приёмы позволяют сделать презентацию интерактивной, отвечающей на действия детей и позволяющей делать осознанный выбор объектов исходя из дидактической цели конкретного задания. Всё это отвечает образовательным потребностям младших школьников в соответствии с возрастными особенностями восприятия, мышления, внимания.

Презентация становится мультимедийной за счёт использования звуковых объектов (образцов речевого высказывания, музыки, песен) и видеофрагментов.

Программа *MS PowerPoint* позволяет создавать интерактивные задания разных типов:

- выбор необходимого объекта среди предложенных;
- составление объекта из частей (слова из букв, предложения из слов, геометрической фигуры из «кусочков» и т.п.);
- классификация объектов (части речи, группировка по обобщающим словам, природные и математические объекты);
- установление соответствия между множествами объектов (предмет и признак, предмет и действие, автор и название произведения, арифметический пример и ответ);
- перемещение объектов (составление словосочетаний, подбор слова к картинке);
- составление таблицы;
- рисование, дорисовывание и др.

С учётом этих положений авторским коллективом преподавателей института педагогики и психологии Сыктывкарского государственного университета имени Питирима Сорокина разработано интерактивное электронное учебное пособие на коми и русском языках «Окружающий мир. 1 класс» [4].

При разработке ЭОР авторы опирались на современные дидактические принципы развивающего обучения, выдвинутые Д.Б. Элькониним, В.В. Давыдовым и Л.В. Занковым [5], а также принцип интеграции образовательных областей. Предварительный анализ программы и учебников



А.А. Плешакова, Г.Г. Ивченко, И.В. Потапова [6] позволил нам сконструировать сценарий пособия на основе тематического принципа.

Пособие содержит 9 сцен: «Я и мой дом», «Дорога в школу», «Моя школа», «Мой класс», «Моя деревня», «Сыктывкар – столица Республики Коми», «Моя республика», «Москва – столица России», «Мой дом – моя планета». В каждой сцене присутствует персонаж – зверёк, который является символом коми промыслового календаря. Белка, Выдра, Лось и другие персонажи предлагают учащимся выполнить интерактивные задания и сопровождают детей в изучении теоретического материала соответствующей сцены.

Совместно с героями-сверстниками Агни и Микитом первоклассники погружаются в увлекательное путешествие по Республике Коми и России, знакомятся с традициями и достопримечательностями нашей Родины, узнают о природных зонах, о живой и неживой природе, учатся правилам поведения в транспорте, на улице, в театре.

Технические возможности программы *MS PowerPoint* позволили объединить в структуре каждой сцены познавательный материал в текстовой и графической форме, интерактивные задания на активизацию мыслительной деятельности учащихся и материал для проверки сформированности знаний.

Анимации и триггеры использованы в заданиях на исключение лишнего, выбор объектов, классификацию. Для заданий с «перетаскиванием» объектов использован макрос *DragAndDrop*.

Использование внутренних и внешних гиперссылок позволяет включать в презентацию другие средства и ресурсы. Например, после изучения темы детям предлагается собрать пазл по теме сцены (два уровня – простой и сложный; уровень сложности выбирает сам первоклассник). Дополнением каждой сцены являются интернет-ресурсы (видеозаписи, официальные сайты), которые помогут учителю более подробно раскрыть содержание изучаемой темы.

Методический аппарат пособия содержит в себе задания, направленные на формирование предметных, метапредметных и личностных результатов основной образовательной программы начального общего образования.

Мультимедийность и интерактивность разработанного пособия позволяют использовать его для реализации различных дидактических задач на основе системно-деятельностного подхода на уроках разного типа.

Учитель может организовать изучение материалов пособия на уроках коми языка и окружающего мира в рамках индивидуальной, фронтальной и групповой работы с применением интерактивной доски, экрана и мультимедиа.



медийного проектора или персональных компьютеров. Пособие может быть использовано для индивидуальной и групповой домашней работы детей, а также в организации внеурочной проектно-исследовательской деятельности.

Разработанное электронное пособие является логически завершённым продуктом. Оно позволит создать условия для успешной социальной адаптации первоклассников. От того, как пройдёт адаптация на первом году обучения, во многом зависит работоспособность и успеваемость в последующие годы.

Отличительной чертой пособия является то, что оно полностью составлено на коми языке. На сегодняшний день в школах Республики Коми представлен достаточно скудный учебный и познавательный материал на коми языке. Поэтому данное учебное электронное пособие могут использовать в своей практике не только учителя начальных классов, но и учителя коми языка как родного и как неродного.

Русскоязычный вариант пособия может быть рекомендован к использованию во всех школах Республики Коми.

Описанный ЭОР представлен в открытом доступе на сайте Межрегиональной лаборатории информационной поддержки функционирования финно-угорских языков по адресу <http://fu-lab.ru/uchposobiya>

Описанная и апробированная авторами технология может быть транслирована на другие образовательные области для разработки ЭОР.

* * *

1. Федеральный государственный образовательный стандарт начального общего образования (утв. приказом Минобрнауки России от 06.10.2009 № 373; в ред. приказов от 26.11.2010 № 1241, от 22.09.2011 № 2357). – URL: http://xn--80abucjiibhv9a.xn--p1ai/%D0%B4%D0%BE%D0%BA%D1%83%D0%BC%D0%B5%D0%BD%D1%82%D1%8B/922/%D1%84%D0%B0%D0%B9%D0%BB/748/%D0%A4%D0%93%D0%9E%D0%A1_%D0%9D%D0%9E%D0%9E.pdf (дата обращения: 23.01.2017).

2. ГОСТ Р 53620-2009. Информационно-коммуникационные технологии в образовании. Электронные образовательные ресурсы. Общие положения. – URL: <http://www.gost-load.ru/Index/50/50209.htm> (дата обращения: 23.01.2017).

3. *Габова М.А.* Электронные образовательные ресурсы в профессиональной подготовке будущих педагогов дошкольного образования // *Детский сад: теория и практика.* – 2012. – № 1. – С.100–112.



4. Окружающий мир. 1 класс: электронное образовательное пособие. – URL: <http://fu-lab.ru/uchposobiya> (дата обращения: 23.01.2017).

5. Обучение и развитие / под ред. Л.В. Занкова. – М.: Педагогика, 1975; *Давыдов В.В.* Теория развивающегося обучения. – М., 1996; *Эльконин Д.Б.* Психология обучения младшего школьника. – М., 1974.

6. *Плешаков А.А.* Окружающий мир. Рабочие программы. Предметная линия учебников системы «Школа России». 1–4 классы: пособие для учителей общеобразовательных организаций. – М.: Просвещение, 2014; *Потапов И.В.* Обучение в 1 классе по учебнику «Окружающий мир» Г.Г. Ивченковой, И.В. Потапова: программа, метод. рекомендации, поурочные разработки. – М.: АСТ; Астрель, 2015.



М.А. Габова,

*канд. пед. наук, доцент кафедры
начального и дошкольного образования
Сыктывкарского государственного
университета имени Питирима Сорокина*

В.Ф. Поберезкая,

*канд. пед. наук, доцент кафедры
начального и дошкольного образования
Сыктывкарского государственного
университета имени Питирима Сорокина*

С.Н. Терентьева,

*доцент кафедры теории
и методики обучения Сыктывкарского
государственного университета
имени Питирима Сорокина*

ТЕХНОЛОГИЯ ИСПОЛЬЗОВАНИЯ КОМИ-РУССКОГО ЭЛЕКТРОННОГО ОБРАЗОВАТЕЛЬНОГО РЕСУРСА «ОКРУЖАЮЩИЙ МИР» В НАЧАЛЬНОЙ ШКОЛЕ

Созданный авторским коллективом Сыктывкарского государственного университета имени Питирима Сорокина электронный образовательный ресурс (далее – ЭОР) «Окружающий мир. 1 класс» предназначен для использования на уроках окружающего мира в 1 классе (по программам А.А. Плешаковой, М.Ю. Новицкой «Окружающий мир» – 1 класс по системе «Школа России» и А.А. Вахрушева, О.В. Бурского, А.С. Раутиана «Окружающий мир» – 1 класс по системе «Школа 2100» [1]) на коми и русском языках.

Этот интерактивный ресурс был создан с целью формирования этнокультурных представлений обучающихся посредством ЭОР в процессе освоения образовательной области «Окружающий мир» на коми и русском языках.

В данном ресурсе авторы учитывали современные дидактические принципы развивающего обучения (Д.Б. Эльконин, В.В. Давыдов, Л.В. Занков [2]) и принцип интеграции разных образовательных областей.

При создании ЭОР авторы опирались на требования федерального государственного образовательного стандарта начального общего образо-



вания, согласно которым вся деятельность педагога должна быть направлена на формирование и развитие универсальных учебных действий учащихся начальных классов.

Предметные результаты предполагают формирование конкретных знаний и умений в предметной области «Окружающий мир» (1 класс). Мы рассмотрели программы А.А. Плешаковой, М.Ю. Новицкой «Окружающий мир» – 1 класс по системе «Школа России» и А.А. Вахрушева, О.В. Бурского, А.С. Раутиана «Окружающий мир» – 1 класс по системе «Школа 2100» и выделили основные содержательные блоки (сцены), которые необходимо изучить первокласснику:

- «Я и мой дом»;
- «Дорога в школу»;
- «Моя школа»;
- «Мой 1 класс»;
- «Деревня»;
- «Сыктывкар – столица Коми»;
- «Моя республика»;
- «Москва – столица России»;
- «Мой дом – моя планета».

Учебников по окружающему миру на коми языке нет, поэтому для успешной адаптации первоклассников, у которых родной язык – коми, мы решили разработать содержание перечисленных сцен на коми и русском языках.

Познавательные метапредметные результаты отражены в содержании интерактивных заданий, способствующих формированию логических действий: сравнение, классификация объектов, установление причинно-следственных связей, подведение под понятие, замещение, моделирование.

Разработаны задания, которые стимулируют осознанное и произвольное построение речевого высказывания на коми языке, как в устной, так и письменной форме. В каждой сцене присутствует рефлексия способов, условий действия, контроль и оценка процесса и результатов деятельности.

Регулятивные метапредметные результаты предусмотрены в заданиях, которые направлены на развитие самоконтроля, коррекции и саморегуляции обучающихся.

Коммуникативные метапредметные результаты заложены в заданиях, разработанных с учётом принципа индивидуализации и дифференциации процесса обучения, поэтому учитель на уроке может организовать как индивидуальную, так и групповую форму обучения (для формирования умений с достаточной полнотой и точностью выражать свои мысли



в соответствии с задачами и условиями коммуникации, овладения монологической и диалогической формами речи в соответствии с грамматическими и синтаксическими нормами родного языка).

Личностные метапредметные результаты реализованы через материалы, направленные на формирование уважительного отношения к людям разных национальностей, пола и возраста, осознание своей этнической принадлежности в Российской Федерации, нравственных норм поведения в транспорте, на улице, в театре, а также основ экологического сознания, что в целом будет способствовать формированию у обучающихся целостной научной картины природного и социокультурного мира, осознанию своего места в обществе.

Представленные в пособии материалы решают пропедевтическую задачу образования – формирование общности образовательных областей через основные понятия курса химии, астрономии, физики, ОБЖ, литературного чтения.

Пособие может быть использовано для индивидуальной работы детей за персональным компьютером на уроках и дома, а также для индивидуальной, фронтальной и групповой работы на уроках и во внеурочной деятельности с применением интерактивной доски.

При создании пособия необходимо было учитывать этнографические особенности коми народа, поэтому было решено, что 9 блоков-сцен будут удачно соотноситься с коми промысловым календарём. Каждой сцене соответствует свой зверёк промыслового календаря, который сопровождает обучающегося при изучении темы. Структура и содержание сцен представлены в *таблице*.

Т а б л и ц а

*Структура и содержание пособия
«Окружающий мир. 1 класс» (на коми языке)*

Сцена	Тема сцены	Герой, который сопровождает ученика	Цель	Особенности прохождения сцен
1	Я и мой дом	Белка	Ученик, открывая новое вокруг нас, сможет рассказать о своей семье, профессиях, семейных традициях и правилах безопасного поведения в быту	Животное (в правом верхнем углу экрана) используется как гиперссылка для перехода на следующий слайд

*Продолжение таблицы*

2	Дорога в школу	Лиса	Овладев знаниями о правилах безопасности дорожного движения, ученик сможет дойти до школы по карте-схеме	Предусмотрены задания на классификацию, установление последовательности и обобщение с перетаскиванием объектов
3	Моя школа	Выдра	Ученик сможет самостоятельно определить назначение кабинетов по их оборудованию – библиотеку, актовый и спортивный залы, медицинский кабинет	Предусмотрены задания с перетаскиванием объектов. Гиперссылки на знаках «!» и «?»
4	Мой 1 класс	Лось	Ученик, зная особенности комнатных растений, сможет определить последовательность развития растения, а также назвать виды комнатных растений	Предусмотрены задания на классификацию с перетаскиванием объектов
5	Деревня	Росомаха	Ученик сможет рассказать о традиционном укладе коми семьи, об архитектурных особенностях жилища, о традиционных праздниках и забавах. Учится классифицировать зимующих и перелётных птиц	Предусмотрены задания на классификацию с перетаскиванием объектов
6	Сыктывкар – столица Коми	Горноста́й	Первоклассник узнает о правилах поведения в общественном транспорте, об архитектурных достопримечательностях, культурных, исторических и промышленных объектах Сыктывкара	Предусмотрен тест с выбором ответа, задания на классификацию с перетаскиванием объектов



Продолжение таблицы

7	Моя республика	Олень	Ученик узнает о природных зонах «тундра», «тайга», о живой и неживой природе, о достопримечательностях и о традиционных праздниках	Красные маркеры на карте Республики Коми – это гиперссылки на слайды о соответствующих городах или районах, порядок может быть любой. Для возврата на карту – гиперссылка на изображении оленя в правом верхнем углу экрана
8	Москва – столица России	Медведь	Ученик узнает об архитектурных, культурных, исторических, спортивных достопримечательностях Москвы и 7 чудесах России, путешествуя на различных транспортных средствах	Предусмотрены вопросы по содержанию видеоролика
9	Мой дом – моя планета	Куница	Ученик узнает о космических телах (звёздах, планетах, спутниках, кометах) и планете Земля: о материках, океанах, разнообразии животного мира	Предусмотрены задания с перетаскиванием объектов

Структура сцен одинакова:

- теоретический материал;
- интерактивные задания;
- материал для проверки сформированности знаний (базового и основного уровней);
- коми-русский словарь;
- пазлы (в конце прохождения сцены ученик собирает тематический пазл, который имеет 2 уровня сложности, отличие – количество деталей картинки).

Большинство сцен содержат ссылки на дополнительные интернет-ресурсы (видеозаписи, официальные сайты организаций т.д.), которые учитель может использовать на уроках и во внеурочной работе.



Раскроем содержание двух сцен ЭОР.

В сцене 6 «Сыктывкар» Сьёдбёж (горноста́й) встречает ребят на вокзале столицы Республики Коми – г. Сыктывкара и они пересаживаются на автобус и совершают путешествие по столице, используя тест, позволяющий актуализировать знания о правилах поведения в общественном транспорте. Авторами разработки предложена как старая карта города (историческая справка о происхождении столицы), так и современная (по которой будет совершаться путешествие). Ученики знакомятся с достопримечательностями города (административные здания, культурные сооружения, памятники), а также с заводом и фабрикой. После посещения Института геологии им. А.А. Чернова им предлагается правильно определить некоторые горные породы, а затем выбрать названия подходящих памятников. Рефлексивный этап данной сцены: детям предлагается собрать пазлы – фотографии Сыктывкара.

После знакомства со столицей республики (Сцена 7 «Республика Коми») первоклассники вместе с Кёр (оленом) отправляются в путешествие по Коми республике и знакомятся с её городами и районами. Авторами ЭОР разработана интерактивная карта Республики Коми, поэтому путешествие можно начинать с любого района или города. После знакомства с городом или районом ученики выполняют задания на знание природных зон, живой и неживой природы, культурного наследия и традиций коренного населения. Рефлексивный этап данной сцены: детям предлагается собрать пазлы – Северное сияние в Республике Коми.

После прохождения всех сцен ученики приходят к выводу: я, моя семья, мой дом, моя школа, моя республика, моя Россия, моя планета – всё едино, взаимосвязано и всё прекрасно.

Использование данного ЭОР позволит учителю реализовать проектную и игровую технологию, сформировать информационно-коммуникационную компетентность обучающихся и приобщить их к этнокультурным ценностям региона.

Так как ЭОР создан на двух языках, то русскоязычный вариант пособия может быть рекомендован к использованию во всех школах Республики Коми.

Данный ресурс представлен в открытом доступе на сайте Межрегиональной лаборатории информационной поддержки функционирования финно-угорских языков по адресу <http://fu-lab.ru/uchposobiya>

Описанная и апробированная авторами технология может быть транслирована на другие образовательные области для разработки ЭОР.



* * *

1. Плешаков А.А. Окружающий мир. Рабочие программы. Предметная линия учебников системы «Школа России». 1–4 классы: пособие для учителей общеобразовательных организаций. – М.: Просвещение, 2014.

2. Обучение и развитие / под ред. Л.В. Занкова. – М.: Педагогика, 1975; Давыдов В.В. Теория развивающегося обучения. – М., 1996; Эльконин Д.Б. Психология обучения младшего школьника. – М., 1974.



А.К. Гагиева,

*д-р ист. наук, доцент, профессор
кафедры управления информационно-
документационными и социально-
политическими процессами
управления ГОУ ВО КРАГСиУ*

ИСТОРИЧЕСКИЙ ОПЫТ ДОКУМЕНТИРОВАНИЯ ЯЗЫКА

В настоящее время много внимания уделяется электронной письменности. В основе научных исследований лежат вопросы языковой технологии и компьютерной лингвистики как языков крупных диаспор, так и малых народов. Не принижая значения данных исследований и их определяющей роли в развитии науки сегодня, хотелось бы отметить следующее. Известно, что в основе любого научного изыскания в области документирования лежит изучение документирования языка.

Сегодня под документированием языка понимается «долговременная многоцелевая фиксация языковых данных», которая связана с полевой лингвистикой, т.е. с изучением языков в среде их естественного существования [1]. Целью документирования, «в идеале, является создание всеобъемлющего корпуса первичных данных, который полностью удовлетворит последующие поколения пользователей, какой бы аспект языка они ни захотели исследовать. Этим занимаются лингвисты-тинологи, изучающие общезязыковые закономерности путём сопоставления далёких и непохожих друг на друга языков; социолингвисты, этнографы, фольклористы и др.» [2]. Их основная задача – собрать, изучить, проанализировать и сохранить сегодняшний материал, сделать его общедоступным, и притом доступным в удобной электронной форме, что значительно повысит эффективность работы каждого отдельного исследователя и позволит избежать многократного повторения одной и той же работы разными людьми [3]. Считаем, что это заключение справедливо. Вместе с тем в контексте данной работы необходимо отметить следующее.

Известно, что язык любого народа формировался в процессе человеческой деятельности. На него оказывали влияние различные факторы (экономические, политические, социальные и т.д.). Сегодняшние языки народов России, особенно языки малых народов, формировались в различные



эпохи – феодальную, капиталистическую, социалистическую. Проживая в составе Русского государства, Российской империи, СССР и включаясь прямо или косвенно в экономические отношения, носители языка были вынуждены приспосабливаться и изменяться. Изменялся и язык. Этот фактор необходимо учитывать, особенно при анализе документированием языка современных носителей. Не учитывая исторического опыта формирования языка, в частности языка делопроизводственного, трудно, а в ряде случаев невозможно определить причины возникновения того или иного лингвистического явления.

Изучение исторического опыта документирования языка имеет давнюю традицию. В работах М.П. Илюшенко, Н.П. Ерошкина, Я.З. Лившица и В.А. Цикулина [4] и др. неоднократно подчёркивалась связь языка с типом и формой документа. Указывалось, что на формирование языка документа влияют как требования, выдвигаемые временем и эпохой, так и принятые правила оформления документации. Мы придерживаемся устоявшейся традиции и попытаемся представить в данной работе основные этапы формирования документирования языка.

Дошедшие до сегодняшнего дня документы показывают, что в период становления Киевского государства и феодальной раздробленности (IX–XV вв.) шло накопление традиций в области формирования языка документов. Создавались устойчивые формуляры, закрепились приёмы установления документов, возникали стадии написания документа. Однако утверждать, что были сформулированы определённые правила документирования языка, нельзя. Отсутствовала система определённых правил и приёмов работы с документами [5].

В XVI–XVII вв. было сформировано «приказное делопроизводство», получившее своё наименование от названия первых государственных учреждений – приказов. В это время формировался особый язык при подготовке документа, который соответствовал его характеру и назначению. Были сформированы системы документации, которые фиксировали отдельные функции управления (дипломатическую, военную, хозяйственную и др.). Так, дипломатические отношения с зарубежными государствами оформлялись договорами, грамотами, а военные задачи – челобитными, приговорами, списками, сказками и др. Язык документов представлял собой сплошной текст. С целью упрощения задачи составления документов появились «образцовые книги», где не только содержались образцы того или иного документа, но и были представлены примеры оформления языка документа. Основной особенностью приказного делопроизводства и языка документа данной эпохи являлось то, что их организация основывалась на нормах обычного права, т.е. на устной форме, передаваемой из поколения в поколение.



В XVIII в. произошло кардинальное изменение делопроизводства; на место приказному пришло «коллежское», которое было законодательно регламентировано. Основным законом стал «Генеральный регламент», утверждённый Петром I в 1721 году. Он определил не только структуру новых государственных учреждений – коллегий, должностные обязанности чиновников, но и порядок составления, контроль над исполнением, язык и хранение документов [6]. Появились также новые виды и формы документов. Челобитные были заменены на прошения [7]. Кроме того, появились донесения, приказы, ордера, табеля и др. Каждому документу было свойственно своё языковое оформление. При этом они обязательно содержали сведения: об адресате, которые начинались с предлога «в»; сведения об авторе указывались под адресатом и начинались с предлога «из»; далее шло наименование документа и краткое содержание. Подпись включала в себя наименование должности, имя, фамилию, титул и следовала за текстом документа. Скрепки проставлялись в правом нижнем углу листа и содержали имя и фамилию писаря, а также подписи челобитчиков. В левом нижнем углу документа проставлялась дата написания. Исходящий регистрационный номер писали вдоль поля документа. К концу XVIII в. на документах, особенно при ответе на крестьянские челобитные, проставлялось большое число делопроизводственных отметок, прежде всего резолюций, которые писали на лицевой стороне первого листа. Можно встретить отметки о рассмотрении дела, о том, кто и когда рассматривал документ и т.д. Все поступающие ответы, особенно на коллективные челобитные, хранились в волостной избе, а также в уездной канцелярии, а после 1775 г. – в нижних земских судах и уездных казначействах. Здесь на постоянной основе работали секретари, канцеляристы, подканцеляристы и т.д. Таким образом, наблюдалось формирование штата чиновников, профессионально работавших с документами. В связи с отменой столбцов и введением листовой формы документа произошли коренные изменения формуляра документа. Были разработаны технологии регистрации документов, контроля их исполнения, текущего и архивного хранения. Язык документа приобрёл западноевропейский характер.

В XIX – начале XX в. наступил новый этап в развитии делопроизводства и языка документа. Так, 8 сентября 1802 г. был издан Манифест о создании первых восьми министерств; 25 июня 1811 г. было издано «Общее учреждение министерств», которое регламентировало министерское делопроизводство. Был установлен порядок оформления, подписания, хранения, регистрации документов. Порядок оформления был назван «Образ производства дел» и имел 5 составных частей, где указывалось прохождение документа. Если подготовительная работа по оформлению документа



и принятия решений велась в канцелярии коллегии, то теперь все чиновники канцелярии принимали участие в его оформлении и подготовке решения. Язык документа этого периода отличался формализацией использования принятых в министерском делопроизводстве оборотов речи, утверждёнными правилами расположения и оформления реквизитов. Характерным явлением этого периода стало стремление к сокращению объёма документооборота путём делегирования части управленческих решений нижестоящим учреждениям, а также рационализации делопроизводства. Так, в 1911 г. было издано «Положение о письмоводстве и делопроизводстве», где были прописаны правила составления документов, язык документа, а также этапы прохождения документа в делопроизводстве [8].

В 1917 г., после установления Советской власти и формирования советской администрации, были изданы законы, вносившие изменения в деятельность делопроизводственных служб. Начался советский этап в системе делопроизводства. Были введены новые правила оформления документов, изменился алфавит, язык документа, структура делопроизводства и т.д. Все изменения и в языке документа, и в организации делопроизводства находились под контролем государства. Примером может служить Единая государственная система делопроизводства, одобренная в 1973 г., которая до настоящего времени сохраняет свою актуальность. Основные её положения – научно упорядоченный комплекс правил, нормативов и рекомендаций, определивших и регламентировавших организацию делопроизводственных процессов в учреждениях, организациях, на предприятиях. Целью данной системы являлось усовершенствование на единой основе организации делопроизводства путём внедрения единых форм и методов работы с документами. Это была система, носившая общегосударственный характер, она была комплексной и законченной. Заметную роль в повышении качества подготовки документов и утверждении единых норм языка документа сыграли государственные стандарты – ГОСТы. Сегодня они также востребованы и необходимы. В этих документах установлена единая терминология, применяемая в делопроизводстве и архивном деле, а с внедрением в управление средств вычислительной техники в ГОСТах утверждено требование о необходимости придания юридической силы документам на машинописном носителе [9].

Сегодня, в период внедрения языковых технологий и компьютерной лингвистики, как никогда востребован исторический опыт формирования языка документа, который позволит не только избежать ошибок, но и с уверенностью смотреть в будущее.



* * *

1. *Архипов А.В.* Документирование малых языков: научные и технические аспекты // Языковое разнообразие в киберпространстве: российский и зарубежный опыт. – М., 2008. – С.76–83.

2. *Архипов А.В.* Указ. соч.

3. *Архипов А.В.* Указ. соч.

4. *Илюшенко М.П.* История делопроизводства в дореволюционной России. – М.: РГГУ, 1993. – С.18–34; *Ерошкин Н.П.* История государственных учреждений дореволюционной России. – 3-е изд. – М.: Высшая школа, 1983. – С.71–137; История делопроизводства в СССР: учеб. пособие / под ред. Я.З. Лившица, В.А. Цикулина. – М.: МГИАИ, 1974. – С.35–65.

5. *Илюшенко М.П.* Указ. соч.

6. *Ерошкин Н.П.* Указ. соч.

8. *Илюшенко М.П.* Указ. соч.

7. *Гагиева А.К.* Крестьянские челобитные второй половины XVIII в. как исторический источник по истории волостного делопроизводства // Управление документацией: прошлое, настоящее и будущее: материалы II Международной науч.-практ. конф. (Москва, 19–20 марта 2015 г.). – М., 2015. – С.166–172.

9. История делопроизводства в СССР: учеб. пособие / под ред. Я.З. Лившица, В.А. Цикулина... – С.35–65.



Г.Л. Григорьев,
инженер-программист
ООО «КРОК»

ПОПУЛЯРИЗАЦИЯ И РАСПРОСТРАНЕНИЕ ЯЗЫКОВ С ПОМОЩЬЮ МОБИЛЬНЫХ ПРИЛОЖЕНИЙ НА ПРИМЕРЕ УДМУРТСКОГО ЯЗЫКА

1. Экранные клавиатуры, раскладки.

Телефоны, планшеты вошли в современную жизнь очень прочно, без них уже сложно представить наш мир. Поэтому важным моментом в развитии любого языка является его распространение с помощью мобильных приложений. Одним из самых важных инструментов для распространения языков в электронной среде является клавиатура, раскладка. Было бы хорошо с помощью одной единственной клавиатуры набирать текст на любом языке мира, но т.к. буквы и символы языков различны, такой «универсальной» клавиатуры не существует. Поэтому для каждого языка встаёт **проблема** создания своего программного обеспечения для набора текста. Этими программными средствами являются различного рода экранные клавиатуры, а также раскладки клавиатур. К сожалению, не все программы поддерживают языки малых народов. Существует несколько путей решения данной проблемы:

- писать обращения к разработчикам программного обеспечения (далее – ПО) с просьбой добавить тот или иной язык в их ПО;
- разрабатывать своё ПО.

Нами был выбран второй путь как самый быстрый способ достижения цели. Конечно же, в первую очередь необходимо обратить внимание на наиболее распространённые операционные системы, это *Windows (desktop)* и *Android (mobile)*. К счастью, для этих операционных систем существует множество возможных путей решения данной проблемы. Для *Windows* очевидным решением стала утилита **Microsoft Keyboard Layout Creator**. С помощью неё можно создать любую раскладку для клавиатуры с необходимыми символами и буквами (с возможностью набирать дополнительные символы с помощью клавиш *Alt*, *Ctrl*). Но создаваемая с помощью этой утилиты раскладка не всегда позволяет пользователю легко установить её. Поэтому нами было принято решение собрать собственный



установочный пакет с помощью *Inno Setup*. Так был создан установочный пакет **UdmurtPackage**, который позволяет легко установить на ПК пользователя инструменты для набора удмуртского текста. В этот набор вошли:

- 2 вида раскладки;
- словари для системы правописания для браузеров *Google Chrome*, *Firefox*, *Opera*, офисного пакета *LibreOffice* и *OpenOffice*;
- удмуртско-русский переводчик «Электрокыллюкам»;
- свободно распространяемые шрифты Юникод;
- конвертер.

Под операционную систему *Android* было решено создать языковой пакет для программы-клавиатуры *AnySoftKeyboard* от разработчика *Menny Even Danan*. Данное решение было принято в связи с тем, что разработчик предоставлял инструменты для создания собственного языкового пакета. Были созданы языковые пакеты для следующих языков: удмуртский, коми, марийский, тувинский, алтайский. Однако данное решение было не совсем удобным, т.к. для установки клавиатуры необходимо устанавливать дополнительно 2 приложения, и при этом нет возможности добавлять в клавиатуру свой функционал. В связи с этим было решено создать собственное приложение-клавиатуру. За основу была взята «штатная» клавиатура *Android (AOSP)*. Она содержит в себе большой набор базовых функций; одна из важных её особенностей – это возможность набирать текст с помощью «жестов»: непрерывно скользя пальцем по буквам можно легко набирать целые предложения. Так как *Android* является открытой операционной системой, то мы без проблем получили исходный код клавиатуры. Подробно изучив исходный код этого приложения, мы выяснили, что словарь построен следующим образом:

word=tay,f=144
word=али,f=130
word=вал,f=120,

где word=«слово», f=«частотность», соответственно чем выше частотность, тем чаще слово предлагается для набора. Также в словарь можно добавлять сокращения. Например, для слова «понедельник» можно добавить сокращение «пнд», и тогда, при наборе на клавиатуре последовательности букв «пнд» программа автоматически определит, что это сокращение и вместо него напишет полное слово «понедельник»:

word=пнд,f=90,shortcut=понедельник

Однако мы решили использовать этот механизм несколько иначе: вместо сокращений можно записывать переводы слов (с русского на уд-



муртский), тогда при вводе русского слова клавиатурой автоматически будет предложен перевод этого слова на удмуртский. Пример из словаря:

```
word=март,f=90,shortcut=южтолэзь  
word=апель,f=90,shortcut=оштолэзь
```

В этом случае, если пользователь наберёт слово «март», то ему будет предложено поменять это слово на «южтолэзь».

2. Установщики словарей для GoldenDict.

Существует довольно много словарей, которые распространяются в формате *dsl* (*ABBY Lingvo dsl*) и могут быть легко открыты с помощью программы *GoldenDict*. Однако не всегда удобно распространять эти словари, т.к. требуется совершать много предварительных действий для того, чтобы начать пользоваться ими: скачивание, копирование в нужную директорию, установка *GoldenDict* (или другой совместимой оболочки). Этот процесс можно облегчить и ускорить с помощью программы-помощника, который совершит все эти действия автоматически. Так были созданы приложения для устройств *Android* для следующих словарей: удмуртско-русский, русско-коми-русский, марийско-русский, алтайско-русский и др. Также в удмуртско-русский словарь был добавлен морфологический словарь, с помощью которого оболочка начинает понимать словоформы.



В.Н. Епанов,
*научный сотрудник
Центра инновационных
языковых технологий
ГОУ ВО КРАГСиУ*

ФУНКЦИОНИРОВАНИЕ КОМИ-ПЕРМЯЦКОГО ЯЗЫКА В СОЦИАЛЬНЫХ СЕТЯХ

Социальные сети стали неотъемлемой частью жизни многих людей по всему миру. Ежедневно миллиарды людей используют их как непосредственно для общения, так и для других целей (прослушивание музыки, просмотр видео, размещение рекламы и т.д.). Наиболее популярными языками, представленными в социальных сетях, остаются английский, русский, испанский, китайский, французский. В России функционируют и другие языки, кроме русского, в частности коми-пермяцкий.

Вплоть до начала 2010-х гг. коми-пермяцкий язык был слабо представлен в Интернете: имелось незначительное количество контента, отсутствовали сообщества, использующие этот язык, хотя он применялся в переписке, особенно в мессенджерах, например в ICQ.

В 2010 г. появилась первая раскладка клавиатуры для коми языка, что позволило полноценно применять коми-пермяцкий язык в Интернете.

В марте 2012 г. в социальной сети «ВКонтакте» появилась крупная информационно-развлекательная страница «Тупи-Тап». Всего за один год она смогла набрать 3 000 подписчиков не только из Пермского края, но и других регионов Российской Федерации. Во многом благодаря подаче материала развлекательного характера (картинок, мемов, демотиваторов). Большую часть её аудитории составляет молодёжь, из которой 70% – лица женского пола. Одна пятая часть всех посетителей страницы является коми-зырянами, поэтому наблюдались случаи исправления в комментариях коми-пермяцких слов на коми-зырянские, т.к. литературные стандарты этих народов во многом совпадают и различаются незначительно.

В ноябре 2014 г. появился первый блог на коми-пермяцком языке, так и оставшийся единственным. Одной из причин этого может быть недостаточный уровень владения литературным коми-пермяцким языком, т.к. в школах родной язык изучают только до 9 класса и лишь 21% учащихся



Коми-Пермяцкого округа. Всё это приводит к тому, что в переписке можно чаще встретить диалектную речь, чем литературную.

В июле 2016 г. начался перевод социальной сети «ВКонтакте» на коми-пермяцкий язык. К настоящему времени переведена половина «ключей», но из-за отсутствия коми-пермяцкой терминологии перевод затянулся на неопределённый срок.

В личной информации на страницах «ВКонтакте» владение коми-пермяцким языком указали 6 000 человек, а коми-зырянским – 15 500. Эта цифра не совсем соответствует действительности, т.к. разработчики изначально ввели путаницу, используя два схожих наименования языка: «коми» и «коми кыв». Спустя длительное время они исправились и заменили обозначение «коми» на «коми кыв», а «коми кыв» на «перем коми кыв», но не всеми пользователями это исправление было замечено.

Наиболее популярными социальными сетями среди коми-пермяков являются: «ВКонтакте», «Одноклассники», «Инстаграм». На долю первой приходится большая часть аудитории. По состоянию на февраль 2017 г. она представлена девятью группами на коми-пермяцком языке, самыми большими из которых являются «ВЕСТИ Кудымкар» и «Тупи-Тап» (список сообществ представлен в *таблице*).

Т а б л и ц а

*Сообщества социальной сети «ВКонтакте»
на коми-пермяцком языке*

№	Название сообщества	Адрес	Кол-во подписчиков, человек
1	«Тупи-Тап»	vk.com/tupi_tap	5 000
2	«Кывбуррез»	vk.com/kyvburrez	500
3	«Ас кыв дор»	vk.com/askyvдор	20
4	«Учим коми-пермяцкий!»	vk.com/komi_kyv	1 100
5	«ПЕРЕМ КОМИ КЫВЧУКӦР»	vk.com/peremkomikyvchukor	230
6	«МЕНАМ КАГА – мой малыш»	vk.com/kaga_mam	500
7	«ВЕСТИ Кудымкар» (частично на коми- пермяцком)	vk.com/vestikud	5 800
8	«Кыв менам дона»	vk.com/komikyv	120
9	«ЮГӦР_чукӧр»	vk.com/club73203474	50



Основными проблемами существования коми-пермяцкого языка в социальных сетях остаются:

- использование русского языка в переписке между комиязычными пользователями;
- употребление русизмов и диалектизмов и снижение практической грамотности;
- недостаточное использование раскладок для коми языка. По этой причине вместо кириллических символов «Ӧӧ» и «Ӧӧ» применяют либо латинские варианты, либо прибегают к употреблению «Э, О, 0» взамен «Ӧӧ» и «И» вместо «Ӧӧ».

В целом уровень функционирования коми-пермяцкого языка в социальных сетях довольно низок, несмотря на наличие определённого количества контента и использования этого языка в переписке. Необходимы меры по популяризации и повышению активности применения коми-пермяцкого языка в социальных сетях.



УДК 811.161.1

Е.Н. Золина,

*канд. филол. наук, доцент
кафедры иностранных языков
и лингвистики ФГБОУ ВО
«Ивановский химико-
технологический университет»
(г. Иваново)*

А.А. Хуснутдинов,

*д-р филол. наук, профессор
кафедры русского языка
и методики преподавания
ФГБОУ ВО «Ивановский химико-
технологический университет»
(г. Иваново)*

О НЕКОТОРЫХ ОСОБЕННОСТЯХ СТРУКТУРНОЙ ОРГАНИЗАЦИИ ЭЛЕКТРОННЫХ ТЕКСТОВ

Познавательная ценность текста определяется не только той информацией, которая заключается в его содержании. Для полноценного описания текста важна его характеристика с разных сторон: принадлежность текста к тому или иному типу, особенности его построения, способы представления, сфера использования, взаимодействие с адресатом и т.д. Поэтому лингвистика текста относится к числу активно развивающихся направлений филологии, изучающих текст в разных аспектах (см. известные работы М.М. Бахтина, В.В. Виноградова, И.Р. Гальперина и др. [1]). Исследование текста (текстов) важно и с точки зрения тех изменений, которые происходят в сфере человеческого общения. Эти изменения в значительной степени обусловлены бурным развитием в последние десятилетия информационных технологий, значительно расширяющих возможности обработки, сохранения, передачи и преобразования информации. Изучение художественных и нехудожественных текстов нашего времени, независимо от того, в каком аспекте (собственно лингвистическом, литературоведческом, культурологическом и т.д.) оно осуществляется, так или иначе способствует уяснению особенностей отражения индивидуального и общего языкового сознания, «нового мышления», того специфического восприятия современным человеком действительности, определяемого пока – за неимением другого и более точного слова – как «постмодернистское мышление», суть которого



в самом общем виде можно определить как моделирование (виртуализация) действительности. В данном случае мы ставим перед собой задачу указать на некоторые особенности логико-понятийной и композиционно-смысловой структуры современных электронных текстов и используемых в них языковых средств и ресурсов (о некоторых особенностях текста в свете обсуждаемых проблем см., например, [2]).

Традиционно сложившийся, классический вариант построения текста определяется обычно как линейный. Информация в таком тексте располагается последовательно, в определённом порядке и в соответствии с избранной автором сюжетной (в широком смысле) линией. Такой текст с точки зрения его структурной организации представлен тремя разновидностями, которые в школьной практике развития речи определяются как описание (характеристика объекта путём указания на его признаки), повествование (последовательное изложение каких-либо фактов, событий) и рассуждение (логически последовательно построенный ряд суждений и умозаключений на какую-либо тему). Данный текст имеет трёхчастную структуру: вводную часть (зачин), основную часть, заключение (концовку, вывод). Во вводной части даётся информация, необходимая читателю для восприятия основного текста; в основной части излагаются в том или ином виде сведения, ради которых создаётся текст; в концовке ключевые идеи текста представляются в концептуальном виде. Если текст значителен по объёму и сложен по содержанию, то его тема делится на микротемы, а сам текст – на соответствующие структурные элементы, которые также строятся по указанной трёхчленной схеме. Информация вводится в линейный текст в определённой последовательности и в определённую его часть. Это касается и разного рода дополнительных и побочных сведений, авторских отступлений и т.д. – все они в линейных текстах имеют своё место, определяемое содержанием и авторским замыслом. Эта общая схема линейного текста в различных модификациях реализуется в разных сферах функционирования языка (сравним тексты публицистические, научные, официально-деловые).

Особенностью линейных текстов является и то, что «законодателем» в них является автор: он определяет объём и содержание сообщаемой информации, способ и форму её подачи, порядок её изложения, языковые средства и приёмы их использования для реализации авторского замысла и воздействия на читателя и т.д. Читатель здесь выступает в основном в роли «потребителя»: его основная задача – усвоить содержание текста, увидеть заключённые в нём смыслы и идеи, стремясь по возможности приблизиться к авторскому видению. «Авторство» читателя, если под этим понимать глубину постижения текста в читательском восприятии, зависит от уровня его (читателя) образования, знания предмета, отношения к тексту и автору, целевых установок и т.д.



Особую разновидность текстового построения представляет собой гипертекст – текст, в котором объём информации расширяется за счёт включения в него сведений, содержащихся в других текстах (источниках). Такое расширение достигается с помощью ссылок на другие источники информации. Обращение по ссылке к другому тексту как бы включает информацию из него в содержание основного текста, обогащая его дополнительными сведениями. Объём информации в пространстве гипертекста может быть практически безграничным, т.к. источники информации, объединяемые в гипертекст с помощью отсылок, могут сами представлять собой гипертексты, а они, в свою очередь, могут тоже содержать ссылки и т.д.

Гипертекстуальность как особый способ представления информации зарождается, по нашему мнению, уже в рамках линейного текста (об этом применительно к художественным текстам и словарям см., например, работы [3]). Однако своё развитие гипертекст получает с распространением информационных технологий, обеспечивающих быстрый доступ к широкому кругу источников и возможность обработки значительных массивов разнообразной информации.

По сравнению с линейным текстом гипертекст даёт читателю большие возможности для конвертирования текста: обращаясь или не обращаясь к предложенным ссылкам, читатель может сам определять объём конечного текста и его содержание; изменяя последовательность обращения к ссылкам, читатель может изменять последовательность представления информации и т.д. Таким образом, при обработке гипертекста читатель выступает в известной степени как «соавтор», т.к. имеет возможность непосредственно влиять на «конечный продукт», участвуя в структурировании «своего» по форме и содержанию варианта текста из предложенных автором материалов и источников. Иначе говоря, автор создаёт текстовое пространство, а путь чтения избирает читатель. Индивидуальность траектории пути, избранной читателем, определяет субъективность прочтения текста и вариативность его интерпретации.

Современному пользователю сети приходится иметь дело с различной по форме представления информацией (текст, число, графика, аудио, видео, анимация и т.д.), а также с сочетанием разных способов её представления (мультимедиа). Широкие возможности в выборе формы представления информации, допускающие также преобразование информации из одной формы в другую, существенным образом влияют на способы подачи тех или иных сведений потребителю. Предлагаемые потребителю информационные материалы передаются в виде контента, включающего в себя разные по форме представления сведения: тексты, картинки, фотографии, аудио- и видеозаписи и др. Разнообразные по форме материалы объединя-



ются в одно информационное пространство общностью темы: все они так или иначе, в разной форме, с разных сторон и с разной степенью детализации передают одно содержание, раскрывают одну идею, мысль, сообщают об одном событии, явлении и т.д. Иначе говоря, все виды информации (текстовая, графическая, звуковая и др.) подбираются таким образом, чтобы они в максимально доступной для потребителя форме преподносили отвечающие авторскому замыслу сведения, факты, комментарии и т.д. Потребителю при этом предоставляется право выбрать удобную для себя форму представления информации и степень детальности знакомства с содержащимися в материалах сведениями.

Контентный способ подачи информации, по нашему мнению, начинает существенным образом влиять и на структурную организацию современного текста: удельный вес текстов, смысловая и композиционная структура которых организована по принципу информационного наполнения ресурса (контента веб-сайта), постепенно возрастает. Ярким примером такого типа построения текста может служить структурная организация особого жанра *PR*-текстов – пресс-релиза. Пресс-релиз представляет собой текст с актуальной информацией, выпускаемый и распространяемый какими-либо учреждениями, компаниями, организациями, органами для возможного опубликования в печати. Текст пресс-релиза объединён только одной темой – обычно он содержит информацию о каком-либо новостном событии. Эта тема является сквозной, она раскрывается в каждом элементе (обычно абзаце) текста, причём в каждой части текста одна и та же тема освещается с новой стороны, каждая часть текста дополняет общую картину новыми сведениями, данными, фактами и т.д. Текст такого типа строится таким образом, чтобы устранение какого-либо его элемента не наносило существенного ущерба его общей теме и основной идее. (В тексте пресс-релиза элементы текста располагаются по степени их важности. В первом абзаце всегда помещается наиболее важная информация, последовательность следующих элементов определяется значимостью содержащихся в них сведений: чем важнее сведения, тем выше место абзаца в тексте. Такой принцип размещения информации удобен при сокращении текста публикации: удаляются конечные элементы исходного текста, в которых содержится менее значимая информация.)

Контентный способ построения текста требует специального изучения. Главной его особенностью является конструирование текста, затрагивающее как формальную, так и содержательную сторону текста: текст «собирается» из отдельных элементов, причём элементы обладают такими свойствами, что из одного набора элементов можно составить несколько вариантов текста, приспособив каждый из них к условиям определённой



ситуации общения. Следует также отметить, что в таком конструировании существенную роль играет читатель (роль, может быть, не менее важную, чем роль автора текста). Всестороннее исследование таких текстов является задачей будущего. Особый интерес в этом отношении представляют наблюдения учёных за особенностями текстов, относящихся к литературе постмодернизма. Однако круг таких текстов должен быть расширен. В информационных сетях содержится достаточно разнообразных по области применения, назначению, жанру и типу текстов, которые могут быть использованы как материал для исследования специфики отражения в речи особенностей нашего «нового мышления». Обращение к этому материалу расширит и углубит наши представления о содержании и построении текстов различных типов и разного назначения, что позволит не только найти решение ряда обсуждаемых в настоящее время проблем, связанных с изучением текста, но и увидеть новые задачи и перспективы исследования.

* * *

1. *Бахтин М.М.* Эстетика словесного творчества. – М.: Искусство, 1973; *Виноградов В.В.* О теории художественной речи. – М.: Высшая школа, 1971; *Гальперин И.Р.* Текст как объект лингвистического исследования. – Изд. 4-е, стереот. – М.: КомКнига, 2006.

2. *Чернявская В.Е.* Лингвистика текста: Поликодовость, интертекстуальность, интердискурсивность: учеб. пособие. – М.: Директ-Медиа, 2014.

3. *Золина Е.Н.* В поисках формы: к поэтике «профессорского» романа // Языковое образование в России: прошлое, настоящее, будущее: материалы науч.-практ. конф. / отв. ред. Н.К. Иванова. – Иваново: ИГХТУ, 2005. – С.43–47; *Золина Е.Н., Малышенко Е.* Исторические и литературные реминисценции в повести Вильгельма Раабе «Гусиный бунт в Бютцове» // Вестник гуманитарного факультета ИГХТУ. – Иваново, 2006. – Вып. 1. – С.173–177; *Золина Е.Н.* К истокам современной истории культуры (по материалам немецкой гуманитарной мысли второй половины XIX века) // Вестник гуманитарного факультета ИГХТУ. – Иваново, 2007. – Вып. 2. – С.164–172; *Хуснутдинов А.А.* В.И. Даль и русская фразеогрфия: аспекты изучения // В.И. Даль в парадигме идей современной науки: язык – словесность – культура – словари: материалы Всероссийской науч. конф. (3–5 апреля 2003 г.). – Иваново, 2004. – С.36–44; *Хуснутдинов А.А.* Как устроен толковый словарь В.И. Даля: в помощь учителю и пользователю // В.И. Даль в парадигме идей современной науки: язык – словесность – культура – образование: материалы VI Всероссийской науч. конф. (Иваново, 27–28 февраля 2014 г.). – Иваново: Ивановский гос. ун-т, 2014. – С.208–223.



Б.И. Каракулов,

*канд. филол. наук, доцент кафедры
иностранных языков и удмуртской
филологии ФГБОУ ВО «Глазовский
государственный педагогический
институт имени В.Г. Короленко»
(Удмуртская Республика, г. Глазов)*

УДМУРТСКИЙ ЯЗЫК В ЭЛЕКТРОННОМ ИНФОРМАЦИОННОМ ПРОСТРАНСТВЕ

В социолингвистической энциклопедии «Письменные языки мира» (Москва: Academia, 2000), созданной совместно учёными университета им. Лавала (Канада, Квебек) и учёными соответствующих регионов, функционирование удмуртского языка в электронном информационном пространстве, связанном компьютерными сетями, как и других языков России, ещё не обозначено [1].

На рубеже тысячелетий вопросы судьбы удмуртского народа, проблем его сохранения через подъём национального самосознания и развитие языка получают освещение как в выступлениях на различных конференциях, радио, в публикациях на страницах печати, так и научных трудах учёных – статьях и отдельных монографиях. Они помогают глубже видеть и понимать внутринациональные проблемы удмуртского народа.

Для этого периода как никогда подходит замечание, сделанное Фердинандом де Соссюром: «... если какое-либо внешнее потрясение, происшедшее в истории народа, ускоряет языковую эволюцию, то это значит только, что язык вновь обрёл состояние свободы и следует своему нормальному течению» [2].

В 90-е гг. XX в. начинают расширяться функции удмуртского языка: во многих районах создаются экспериментальные удмуртские газеты, увеличиваются часы радиопередач и т.д. Так, например, в Якшур-Бодьинском районе 6 декабря 1990 г. выпущена в количестве 100 экземпляров газета «З:ардон». В Кезском районе выходит приложением к русской газете газета «Камсин». В Дебесском районе эфирное время передач на удмуртском радио доходило до 6 часов. В перестроечный период национальное возрождение выразилось в пропаганде национальной культуры и расширении



обучения истории, краеведению, языку нерусских народов. Уже в 1986 г. возобновлено изучение удмуртского языка в Глазовском пединституте, что связано с открытием удмуртской группы на факультете педагогики и методики начального обучения, а в 1993 г. – удмуртской группы на филологическом факультете и кафедры удмуртской филологии. В 1994 г. в республике число школ, в которых преподаётся удмуртский язык и родная литература, увеличилось на 29 единиц. В 17 школах г. Ижевска открылись удмуртские классы, впоследствии подобные классы стали функционировать в городах Воткинск и Глазов. В 2006 г. в Балезино в 6 детских садах изучали удмуртский язык, но из них в школу набирали только один класс с изучением родного языка. Удмуртский язык начали изучать студенты, обучающиеся в Удмуртском госуниверситете на факультетах романо-германской и русской филологии (в течение одного семестра), факультета филологии и социологии (в течение двух семестров), в финно-угорском отделении высшего колледжа по социально-политическим наукам (4 года), в течение двух лет на русском отделении факультета журналистики.

Согласно Конституции Удмуртской Республики (1994) удмуртский язык получил статус государственного наравне с русским. А на 21 сессии Государственного Совета Удмуртской Республики в ноябре 2001 г. был принят Закон «О государственных языках народов Удмуртской Республики и иных языках народов Удмуртской Республики». Провозглашая в официальном порядке удмуртский язык наравне с русским государственным, власть гарантирует заботу о его сохранении и развитии (и в электронном информационном плане), обеспечивая активное употребление языка в политической, культурной жизни. В этих же целях Совет Министров Удмуртской Республики принимает Государственную программу сохранения и развития удмуртского языка и языков народов, компактно проживающих в Удмуртской Республике, и мероприятий по её реализации [3]. Хотя следует признать, что многие смелые и трезвые мысли по сохранению и развитию удмуртского языка, отмеченные в этой Программе и возложенные на Министерство образования Удмуртской Республики, Министерство труда, Институт истории языка и литературы УрО РАН, остались невыполненными.

Для России XX в. стал веком стремительного развития технологий. Уже в 2014 г. удмуртский язык стал одним из самых распространённых финно-угорских языков России в Интернете: на нём создано около 45 блогов [4]. Первопроходцами удмуртского национального Интернета считают удмуртов-селян – братьев Петровых из деревни Курчум-Норья. Далее появляются частные сайты Чиркова из Игры, Д. Сахарных из Ижевска, Ф. Миннигараева (<http://pedor.vukogurt.ru>) и т.д.



Значительным событием в продвижении удмуртского языка в Интернете стало создание Удмуртской Википедии в 2005 году. Её основатель Д. Сахарных важным препятствием для «свободной энциклопедии» развития считает проблему перевода источников информации на удмуртский язык [5]. Википедисты большое внимание уделяют качеству перевода статей и сохранению нормы удмуртского литературного языка, поэтому переводчиков выбирают через конкурс (см. напр.: <https://vk.com/feed>).

Удмуртских пользователей теперь можно найти в социальных сетях *Instagram*, *Odnoklassniki* и *Twitter*. В 2010 г. популярная российская социальная сеть «ВКонтакте» получила интерфейс и на удмуртском языке. С 2008 г. её переводом занималась группа энтузиастов, которая организовала в Интернете целое движение за распространение и сохранение удмуртского языка.

Ещё в январе 2011 г. блоггер Р. Романов заметил, что «острая потребность в существовании удмуртскоязычного интернета наблюдается только у кучки людей, число которых не превышает 150–200 человек. 100 человек в сутки – стандартная посещаемость самого популярного удмуртского ресурса – группы Удмуртлык ВКонтакте. И это число уже год как не растёт, несмотря на рост группы. Возможно есть молодежь в сёлах и деревнях, до которых ещё не дошёл ВКонтакте» [6].

Однако к середине декабря 2013 г., по подсчётам К. Пишлэгера, в Удмуртской ВКонтакте уже насчитывалось около 100 групп, в каждой из которых среднее число участников составляло 700 человек [7]. Следует заметить, что в удмуртских группах ВКонтакте пишут разные пользователи, в том числе с недостатками речи, на смешанном русско-удмуртском языке и т.д., но в их единении отмечается и положительный факт: эти молодые люди возвращаются к своим корням, не стесняются быть удмуртами, имеют свой менталитет.

В 2007 г. в Интернете по заказу Министерства национальной политики Удмуртской Республики открылся официальный сайт на удмуртском языке. У жителей Удмуртии появилась возможность поздравить своих родных и близких виртуальной открыткой на удмуртском языке: в декабре 2010 г. впервые разработаны в электронном виде поздравительные открытки и размещены на сайте minnac.ru. Использование удмуртского языка можно проследить и на сайтах других организаций, в частности Электронной библиотеки Национальной библиотеки Удмуртской Республики, факультета удмуртской филологии, театра эстрадной песни «Шулдыр жыт» и т.д.

Но государственные сайты среди молодых пользователей удмуртского Интернета не популярны. Так, эксперт по интернет-рекламе, блоггер Р. Романов замечает: «На уровне государства удмуртские версии госсайтов



создаются зачастую только по принципу “У татар есть татарские сайты, пусть у нас тоже будут удмуртские”. А о главной цели сайта (донесение актуальной информации в режиме онлайн) никто и не задумывается. Ставится галочка “Сайт сделан” и про него забывают. Всё это печально» [8].

Творческое объединение удмуртов, проживающих в России и за её пределами, в конце 2008 г. отпраздновало первый день рождения своего молодёжного сообщества «Удмуртлык». Через Интернет они общаются на удмуртском языке, как сообщает газета «Герд» за 2008 г. (№ 10–11).

Наконец, следует заметить, ведётся работа по созданию удмуртского электронного словаря и удмуртско-русского Яндекс-переводчика: <https://translate.yandex.ru/translator/Russian-Udmurt>

Электронный словарь составляется на основе Удмуртско-русского словаря (Ижевск, 2008), в котором уже обнаружили недостатки, например отсутствие компьютерной терминологии. Для Яндекс-переводчика не выбран корпус русско-удмуртских переводных текстов. Некоторые программисты, не изучив вопроса, в этом случае неправомерно говорят о «крайней неразвитости удмуртской корпусной лингвистики» [9]. На небольшом материале можно увидеть, что этот переводчик не распознаёт глагольные формы и не выявляет субъект действия:

1. а) мон вож кыз улын **сылысько** – я **стоял** (*нужно: стою*) под зелёной ёлкой; б) Удмурт кылын **вераськыны** – **Говорят** (*нужно: говорить*) на удмуртском языке; в) **Лыдзоно** гожтэт **бадзым** – Большие письма **читать** (*нужно: надо читать*); Она **посещает** детский сад – **Нылпи** садэ **ветло** (*нужно: ветлэ*) **соос** (*нужно: со*).

2. а) Я **учусь** в школе – Мон школаын **дышетскыкум** (*нужно: дышетскисько*); б) Мама глубоко **задумалась** – Анай мур **малпаськыса** (*нужно: малпаськиз*); в) Мынэмез **уг поты** – Не хочу (*нужно: не хочет*) идти; **Сиеме потэ** – Хочется быть (*нужно: я хочу кушать*).

Пользователь должен знать особенности обозначения удмуртских букв, над которыми ставятся точки, без этого получаются такие ошибки: **Зичы кионэз поям** – Волк Зичы паять (*нужно: лиса обманула волка*); **Пияш валэз кутэм** – Мальчик был возвращён (*нужно: мальчик поймал лошадь*).

Если пользователь вместо поднятых точек ставит по привычке двоеточие, переводчик его не понимает: **Ж:о:к** – Здравствуйте:о:к. (*нужно: стол*); **Ж:ог** – Здравствуйте:друзья (*нужно: быстро*).

Качество машинного перевода можно повысить только на большом корпусе качественно выполненных текстов. Сегодня мы не обойдёмся без постредактирования, что сводит на нет преимущества машинного перевода. В настоящее время качество выполненных текстов может оценить и пред-



видеть случаи возможных ошибок при переводе только лингвист, знающий удмуртский язык. А для этого нужно пренебрегать небрежно выполненными переводами [10], выявлять лексику основного словарного фонда и других слоёв, тщательно исследовать контекст употребления слов, что связано и с многозначностью слов и морфологической невыделяемостью.

Представляется, что продвижение использования удмуртского языка в компьютерных технологиях расширит не только сферу использования удмуртского литературного языка, но и горизонты его исследования, а также укрепит дидактические возможности. Внедрение его в удмуртское языковое пространство может сильно повлиять на развитие удмуртского языка и изучающей его науки – удмуртского языкознания.

Как же компьютерная технология может расширить горизонты исследования удмуртского языка? По сообщению Р.Ш. Насибуллина, только у удмуртов сегодня отсутствует лексическая картотека, выполненная на корпусе текстов [11]. К тому же, отсутствуют словари сочетаемости слов, новые толковые, словообразовательные словари, исторический словарь, этимологический словарь, частотный словарь употребляемости слов. Этот пробел можно решить быстрее и качественнее при помощи компьютерной технологии.

* * *

1. Письменные языки мира: Языки Российской Федерации. Социолингвистическая энциклопедия. – Москва: Academia, 2000. – Кн.1. – С.516.

2. *Соссюр де Фердинанд*. Труды по языкознанию. – М.: Прогресс, 1977. – С.183.

3. Постановление Совета Министров Удмуртской Республики от 14.06.1994. – URL: <http://docs.cntd.ru/document/960100226>

4. *Пишлӧгер К.* Записки из муржол underground: superудмурты и cyberspace // Фроловские чтения: материалы IV Международной науч.-практ. конф., посвящённой 80-летию удмуртского поэта Флора Ивановича Васильева. – Глазов: Глазов. гос пед. ин-т, 2014. – С.57; *Романов Р.* Мон Мыльсы-кыдысын. – URL: <http://kazak-of-sky.ru>

5. *Сахарных Д.* Зачем была, есть и будет нужна удмуртская Википедия. – URL: <https://vk.com/udmwiki10ar>

6. *Романов Р.* А есть ли удмуртский интернет? Интервью Ветру. – URL: <http://kazak-of-sky.ru/category/interview>

7. *Пишлӧгер К.* Указ. соч. – С.57.

8. *Романов Р.* А есть ли удмуртский интернет? Интервью Ветру...



9. *Болтачев А.В.* Средства машинного перевода с русского на удмуртский и обратно. Электронные инструменты обучения удмуртскому: перспективы развития // Языки меньшинств в компьютерных технологиях: опыт, задачи и перспективы: материалы Международной конф. – Йошкар-Ола, 2011. – С.76.

10. *Каракулов Б.И.* Переводческая деятельность в истории удмуртского литературного языка // Национальные языки России: региональный аспект: материалы Международной науч.-практ. конф. – Пермь, 2005. – С.116–120.

11. *Насибуллин Р.Ш.* Выступление в дискуссиях – URL: <https://vk.com/udmwiki10ar>



Е.С. Коваль,
*специалист межрегиональной
лаборатории функционирования
информационной поддержки
финно-угорских языков
ГОУ ВО КРАГСиУ*

**ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ
НАЦИОНАЛЬНОГО КОРПУСА КОМИ ЯЗЫКА
В НАУЧНЫХ ИССЛЕДОВАНИЯХ
(на примере работы
«Сравнительные конструкции в коми языке»)**

Любой человек, занимающийся научной деятельностью или косвенно относящийся к ней, будь то студент, аспирант или кандидат наук, понимает, что любая работа начинается со сбора фактического материала. Больше всего с текстами приходится работать филологам. Если раньше сбор материала осуществлялся вручную, то сейчас появилась возможность быстрого поиска нужного текста. В частности, корпус коми языка является доступным и удобным для этого средством.

Что такое корпус? Корпус коми языка – это информационно-справочная система, основанная на собрании коми текстов в электронной форме с возможностью поиска языковых единиц по различным параметрам. Общий объём корпуса текстов составляет более 35 млн словоупотреблений [1].

Перед тем как выяснить возможности корпуса языка, обратимся к теме работы, на примере которой будет показан потенциал корпуса, – «Сравнительные конструкции в коми языке».

Сравнение – это всегда сопоставление двух предметов, явлений, признаков. Таким образом, сравнительная конструкция (далее – СК) – это четырёхкомпонентная структура, выражающая отношения сопоставления двух компаратов на основании общего для них признака или действия, включающая объект сравнения (то, что сравнивают), предмет сравнения (с чем сравнивают), модуль сравнения (на основании чего сравнивают), показатель сравнительного отношения.

Обратимся подробнее к потенциалу корпуса. Во-первых, нужно отметить, что материал, включённый в корпус, является качественным про-



дуктом, т.е. содержит минимум ошибок, а значит, и данные научной работы будут точными, объективными. Как правило, тексты из Интернета изобилуют всевозможными ошибками (орфографическими, пунктуационными и др.), которые могут влиять на конечный результат исследования. В корпусе ошибки в текстах минимизированы, т.к. тексты вычитываются, ошибки и опечатки в текстах исправляются.

Во-вторых, использование корпуса экономит время. Основываясь на личном опыте, можем сказать, что поиск и сбор фактического материала занимает достаточно много времени. Нужно не только вычитать большой объём литературы, но и выписать каждый пример на отдельную карточку, чтобы потом можно было систематизировать эту информацию. Допустим, книгу в 100 страниц в среднем можно прочесть за 7–8 часов, и таких книг нужно прочесть не 2 и даже не 5. Налицо нерациональное использование времени. При этом в корпусе меньше чем за минуту можно найти весь (!) необходимый материал. Например, возьмём формант *кодъ*, введём его в строку поиска. В результате всего лишь за 10 секунд было найдено 32 020 примеров его употребления. Этого более чем достаточно. В текстах коми языка редко встречаются СК с формантом *-лань*, поэтому поиск данных конструкций был довольно проблематичен. В корпусе же за несколько секунд было найдено 23 820 примеров (результаты поиска формантов СК представлены в *табл. 1*).

Т а б л и ц а 1

Результатов поиска формантов СК

	кодъ	моз	бытьё	сяма	рёма	серти
Поиск по корпусу	32 020	41 556	52 565	2 369	4 191	24 715
Поиск по книгам	635	605	635	65	172	124

Конечно, из найденного материала тоже нужно будет делать выборку, но всё равно времени на поиск уходит гораздо меньше.

К тому же в корпусе существует возможность устанавливать условия поиска. Обеспечена возможность поиска по трём параметрам: все комбинации, точное совпадение и только как часть слова, которое делится на позиции в начале и в конце слова.

Приведём пример. В коми языке сравнение может образовываться благодаря суффиксам *-ён*, *-ысь*, *-лань*. Допустим, возьмём аффикс *-ён*, зададим условия поиска: только как часть слова, искомая часть в конце слов. По данному запросу было выявлено 701 887 примеров (см. *рис. 1*). Нуж-

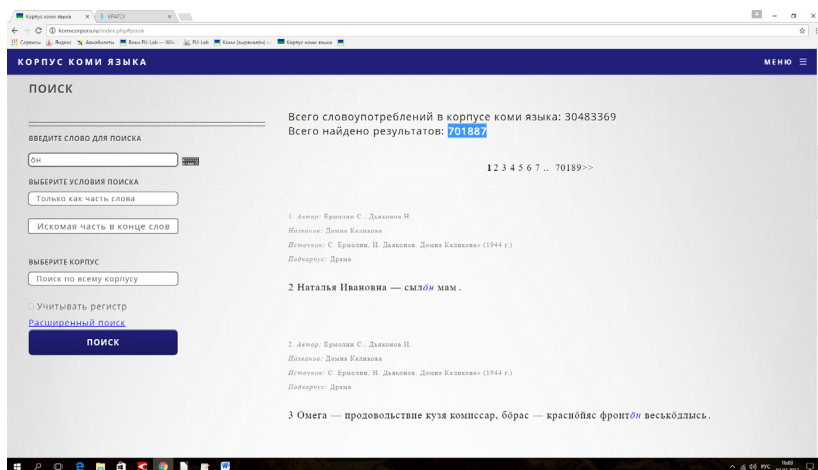


Рис. 1. Пример результатов поиска

но учитывать, что из найденных придётся делать выборку нужных вам примеров, т.к., помимо сравнительной, аффикс *-ӧн* имеет основную функцию — творительную.

Зачастую в научных работах немаловажным оказывается представить аналитическую картину изучаемого вопроса. Для лингвиста важным является количественное наличие того или иного форманта в тексте того или иного стиля; на основе этих данных можно вывести статистику. Среди всех материалов корпуса есть возможность найти примеры употребления слов в газетном, научном стиле, фольклоре и т.д. Корпус коми языка представляет всё многообразие жанров, нужно только задать необходимые условия поиска.

В качестве примера возьмём формант *бытьӧ*. По всему корпусу было найдено 52 565 примеров. Если мы сузим условия поиска, зададим, например, фольклорный текст, то мы увидим, что в количественном плане результат уменьшился (232 примера) (см. рис. 2).

В табл. 2 представлены примеры поиска формантов СК *бытьӧ* и *сяма*.

Конечно, нужно принимать во внимание наполняемость корпуса. На данный момент по объективным причинам в корпус включена не вся база текстов на коми языке, поэтому статистическая картина будет не совсем объективной. Но даже по имеющимся данным можно делать выводы.

Итак, корпус коми языка — это не только эффективный инструмент сохранения всего письменного наследия коми народа. Он является также

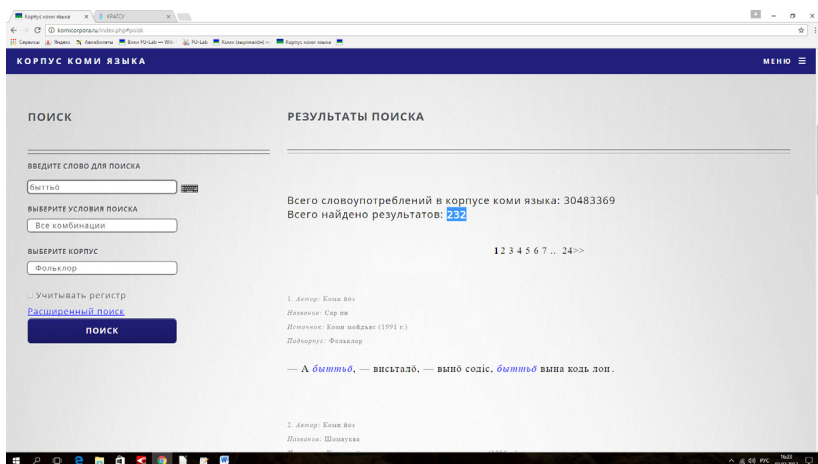


Рис. 2. Пример результатов поиска

Т а б л и ц а 2

Статистика поиска формантов по стилям

	Весь корпус	Худ. проза	Докум. проза	Драма	Газет. стиль	Учебно-науч. проза	Поэзия	Фольклор
бытьё	52 565	38 217	1 257	2 132	3 013	1 063	5 407	232
сяма	2 369	904	208	48	578	425	110	6

объективным собранием текстов коми языка, что позволяет опираться на точные данные при написании научной работы. Корпус языка за короткий промежуток времени позволяет найти большой объём нужной информации при помощи точно заданных условий поиска, а значит, экономит время. Всё это в совокупности помогает обеспечить и улучшить качество научной работы, а также повышает интерес студентов к тем или иным вопросам лингвистики.

* * *

1. Корпус коми языка: информационно-справочная система. – URL: <http://komicorpora.ru> (дата обращения: 04.02.2017).



Н.С. Коровина,

*канд. филол. наук, ст. научный
сотрудник сектора фольклора
Института языка,
литературы и истории
Коми НЦ УрО РАН
(г. Сыктывкар)*

Г.С. Савельева,

*канд. филол. наук, ст. научный
сотрудник Фольклорного
фонда Института языка,
литературы и истории
Коми НЦ УрО РАН
(г. Сыктывкар)*

ФОЛЬКЛОРНОЕ НАСЛЕДИЕ НАРОДА КОМИ В СОВРЕМЕННОЙ КУЛЬТУРЕ

Сегодня всё более очевидно, что происходит утрата старых этнобытовых связей фольклорных явлений и в качестве доминанты выдвигается художественная функция традиционного фольклора, фольклорная традиция воспринимается в первую очередь как эстетический феномен.

Изменение культурного контекста в немалой степени связано с интенсивным развитием средств массовой информации. Пропаганда фольклора при помощи радио и телевидения сделала его доступным широкой аудитории. Если же к этому добавить издания произведений фольклора, то поток фольклорной культуры, который предлагается современному читателю (слушателю, зрителю), представляется поистине безграничным.

В настоящее время насчитывается более 50 сборников, куда включены произведения устно-поэтического творчества народа коми, собранные в полевых условиях или извлечённые из архивов и личных коллекций собирателей.

Самой ранней публикацией произведений коми фольклора является сборник «Syrjäniläisiä häälauluja» (Зырянские свадебные песни), изданный в Хельсинки в 1878 году. Фольклорный материал для данной публикации был собран Маттисом Кастреном (1813–1852) – известным финно-угроведом [1]. С марта по ноябрь 1843 г. он изучал жизнь народов низовья Печоры, побывал в Усть-Цильме, Ижме, Колве. Собранные экспедицион-



ные материалы были изданы известным финно-угорским лингвистом Т.Г. Аминоффым. В сборник включено 7 текстов свадебных причитаний (общий объём текстов – 260 стихов), причём записи носят аутентичный характер, созданы на ижемском диалекте коми языка и являются первыми образцами причетов, представляющими ижемскую локальную традицию. Отсутствие паспортизации не позволяет говорить о месте записи и информантах, хотя, возможно, эта информация и содержится в архивных материалах. Как бы то ни было, собранные М. Кастреном тексты являются самыми первыми записями обрядовой поэзии ижемских коми.

Самый ранний коми фольклорный сборник, изданный в России, «Коми мойдкывъяс» (Коми сказки) был подготовлен в 1913 г. энтузиастом-собирателем, усть-сысольским учителем А.А. Цембером [2]. В 1914 г. им же был составлен ещё один сборник – «Коми мойдан и сыланкывъяс» (Коми сказки и песни) [3]. В два его издания вошла 31 сказка.

Публикации осуществлялись в виде: 1) сборников-антологий, включающих разные жанры фольклора; 2) сборников, представляющих конкретный фольклорный жанр; 3) сборников, отражающих локальную культуру; 4) сборников, посвящённых творчеству отдельного сказителя.

Типичным сборником первого типа является «Фольклорный сборник. Важ коми мойдъяс да сыланкывъяс» (Фольклорный сборник. Старинные коми сказки и песни), подготовленный в 1938 г. П.Г. Дорониным [4]. В данный сборник были включены материалы, собранные бригадой по сбору фольклора, которую возглавлял П.Г. Доронин. В сборник вошло 26 сказок, 36 песен, 14 причитаний. К сожалению, отсутствие паспортных данных и комментариев к текстам снижает достоинство данной публикации.

Один конкретный жанр является основой сборника П.А. Анисимова «Коми сыланкывъяс» (Коми песни, 1926) [5]. Материал был собран им во время совместной поездки на Вишеру в 1926 г. с известным коми писателем В.А. Савиным. Отчёт В. Савина о поездке, появившийся в журнале «Коми му» (Коми край), а также указанный музыкальный сборник «отразили специфику народной поэзии первых после революционных лет» [6].

В 1991 г. вышел сборник «Коми мойдъяс» (Коми сказки), подготовленный Ю.Г. Рочевым [7]. Из 51 текста, включённого в него, 6 – сказки о животных, 12 – бытовые сказки, 33 – волшебные. Сборник отличается от многих упомянутых публикаций хорошо подготовленным научным аппаратом: включённые в него сказки идентифицированы по Сравнительному указателю сюжетов, отмечены также их сюжетные особенности (наличие контаминации и т.д.).

Широкий охват жанров и локальных культур характерен для фольклорных сборников третьего типа. Такой принцип расположения материа-



ла положен в основу публикаций известных финно-угроведов: Ю. Вихмана [8], Д.Р. Фокош-Фукса [9], К. Редери [10].

Особое явление в собирательской и издательской деятельности представляет собой четырёхтомное издание «*Syrtjänische texte*» (Зырянские тексты), подготовленное и изданное в 1985–1995 гг. Паулой Кокконен [11]. Тома включают в себя 174 текста, представляющих как фольклорные жанры, так и рассказы автобиографического и этнокультурного характера, записанные Тойво Эмилем Утилой (1897–1947), финским языковедом в 1942–1943 гг., от коми военнопленных финской кампании. Первый том содержит коми-пермяцкий материал (1985); второй – записи вымской, ижемской, печорской традиций (1986). В третьем томе опубликованы материалы, представляющие удорскую и лузскую традиции (1989). Четвёртый том включает материалы, записанные от носителей вычегодско-сысольской традиции (1995).

Важным событием, «своего рода вехой» [12] в истории коми фольклористики XX в. стал выход в 1941 г. сборника И.А. Осипова «Висер во жса сьыланкывъяс да мойдкывъяс» (Песни и сказки Вишеры) [13]. Он составлен по географическому принципу и характеризует локальную традицию небольшой части Корткеросского района – поселений, расположенных по Вишере и Вычегде, включая сёла Богородск, Большелуг, Сторожевск, Нившеру. Сборник включает 31 песню, 17 свадебных причитаний, 2 похоронных плача и 20 сказок. Во вступительной статье проанализирована специфика коми народной поэзии в её связи с русской культурой. В приложении даны научные комментарии к текстам, сведения об исполнителях, словарь диалектизмов. И.А. Осиповым было сделано всё то, «что до него не удавалось ни одному коми фольклористу» [14]. Этот сборник остаётся одним из лучших и в настоящее время.

Творческий коллектив (А.К. Микушев, П.И. Чисталёв, Ю.Г. Рочев), подготовивший и выпустивший трёхтомную антологию «Коми народные песни» (1966–1971), продолжил традиции составителей фольклорных сборников третьего типа. Первый том антологии «Песни Вычегды и Сысолы» (1966) характеризует фольклорные традиции южных коми [15]; самобытным жанрам ижемских коми посвящён второй том – «Песни Ижмы и Печоры» (1968) [16]; третий том «Песни Вымы и Удоры» (1971) освещает состояние локальной фольклорной культуры западных коми [17]. Трёхтомный труд удостоен Государственной премии Коми АССР.

Сборник А.К. Микушева «Ипатьдорса фольклор» (Фольклор села Ипатьево, 1980) посвящён талантливой коми сказительнице А.А. Шуктовой [18]. Он является пока единственным коми фольклорным сборником четвёртого типа, где рассматривается творчество отдельного исполнителя. Во вступительной статье даётся творческий портрет А.А. Шуктовой.



Имеются двуязычные сборники с параллельным переводом коми текста на другие языки. В 1916 г. в Хельсинки вышел в свет сборник финского языковеда и фольклориста Юрьё Йоосеппи Вихмана (1868–1932) под заглавием «*Syrgjänische volksdichtung*» (Коми-зырянская народная поэзия) [19]. Материал был собран в Коми крае в 1901–1902 гг. во время научной командировки автора. В издание вошло 589 фольклорных текстов, представляющих нижневыхгодскую, средневыхгодскую, сысольскую, летскую, печорскую, лузскую традиции с переводом на немецкий язык.

В 1978 г. в Будапеште вышла книга австро-венгерского языковеда К. Редее «*Zyrgian Folklore Texts*» (Коми фольклорные тексты), материал для которой был собран им в 1964 г. во время научной поездки по территории Республики Коми [20]. Опубликованные здесь коми фольклорные тексты переведены на английский язык.

В сборнике 1913 г. «*Zürjén népköltészeti mutatóványok*» (Образцы коми народной поэзии) венгерский языковед и фольклорист Давид Рафаль Фокош-Фукс (1884–1977) издал фольклорные записи, сделанные им в ходе научных поездок в Коми край в 1911 и 1913 гг. с параллельным переводом на венгерский язык [21].

Среди фольклорных сборников выделяют также научные и научно-популярные издания для массового читателя. Единственным на сегодняшний день академическим изданием в коми фольклористике является сборник А.К. Микушева «Коми народный эпос», вышедший в издательстве «Наука» в 1987 г. [22]. Он подготовлен в лучших традициях эпосоведения, в его состав вошли наиболее оригинальные образцы коми народной поэзии. В общей сложности сборник включает 16 текстов ижмо-колвинского эпоса, 6 произведений коми-зырянского эпоса.

Начиная с 20-х гг. XIX в. стали издаваться первые научно-популярные фольклорные произведения для детей. Так, в сборники, составленные А.А. Сухановой, «Посни чой-воклы мойдан кывъяс» (Сказки для маленьких братьев и сестёр, 1921), «Мойдан кывъяс» (Сказки, 1922) вошло 11 [23] и 5 коми сказок [24] соответственно.

Опора главным образом на свой собственный фольклор при создании детской литературы была вызвана жизненной необходимостью: художественных произведений, написанных для детей, в тот период практически не было. В то же время произведения устного народного творчества, в частности сказки, были доступны детям. С точки зрения языка, формы, педагогической ценности это был благодатный материал для детского чтения. Сборники подобного типа можно считать первыми популярными изданиями для детей. Они издаются и до настоящего времени на основе особых исторически сложившихся принципов и приёмов, отвечающих требованиям педагогической науки.



Большой вклад в разработку принципов издания фольклорных текстов внесли языковеды, в том числе и зарубежные. Хотя лингвистические сборники не имеют научного фольклористического комментария к текстам, в них при этом есть паспортизация текстов, они в основном двуязычны, и, что самое важное, тексты, записанные лингвистами, аутентичны. Все вышеперечисленные фольклорные издания (кроме сборника «Коми народный эпос») претерпели литературную обработку составителей.

В настоящее время фольклористы всего мира обеспокоены сложной проблемой сохранения и поддержания в современной культуре аутентичного фольклора. Общеизвестны, например, выдающиеся заслуги братьев Гримм в открытии фольклорной сокровищницы немецкого народа. Как считает Н. Рошняну, «они старались “одеть” собранный материал в “платье”, очень близкое к народному. Тем не менее, очень трудно говорить об аутентичности сказок братьев Гримм в истинном смысле слова, поскольку данные ими “модели” в большой степени являются плодом их литературного таланта» [25].

Вопрос о степени стилизации непрост. Часто стилизация – единственный способ художественного воплощения фольклора. Однако нельзя не согласиться с тем, что обилие стилизаций приводит к тому, что публика начинает принимать их за истинный фольклор, что неизменно ведёт к её заблуждению и относительно фольклора. Так, современный исследователь коми языка Г.В. Пунегова подчёркивает, что «... крайне важно при первичной записи фольклорного произведения сохранять все особенности речи сказителя, не исправляя и не редактируя их, не выравнивая их под нормы какого-то другого говоря, поскольку подобный текст является источником исторической информации об утраченных ныне говорах и, возможно, ушедших лингвистических особенностях того времени» [26].

Путь, вероятно, может быть один: не отрицая талантливых профессионально выполненных стилизаций, давать дорогу и шире пропагандировать аутентичный фольклор, приучать к нему публику, развивая её вкус и умение отличить истинное от подделки.

Это может сделать только глубоко продуманная система образования и воспитания фольклором, в первую очередь школьников, студентов, предусматривающая на каждом возрастном этапе свои методы знакомства с повествовательным и песенным фольклором, свои формы участия в фольклорных мероприятиях разного уровня, наконец, соответствующий возрасту набор теоретических знаний. Пора перестать включать произведения фольклора в учебники по литературе, следует создавать школьные хрестоматии и учебники по коми фольклору.

Сборники, включающие в себя произведения устно-поэтического творчества коми народа, имеют большое значение не только для фолькло-



ристов, но и для лингвистов. В каждом из перечисленных выше сборников отражены особенности языка соответствующего периода, «раскрывающие отдельные стороны лингвистических единиц коми языка». Анализируя сказки из сборника А.А. Цембера «Коми мойдыквыяс» (Коми сказки), Г.В. Пунегова приходит к выводу: «Обобщение результатов лингвистического изучения сказки представляется нам важным, как для раскрытия собственно языковой специфики сказочного текста, характеристики системы его графических, орфографических, лексико-семантических и изобразительно-выразительных средств, так и для более углублённого исследования коми языка и стиля фольклора в целом» [27].

На фоне стремительного угасания и исчезновения национальных традиций, а также в связи с очевидной недостаточностью аутентичных фольклорных презентаций (будь то публикации материалов или деятельность фольклорных коллективов) особое значение приобретает обращение к архивным материалам, которые являются важнейшим источником не только для изучения, но и практического освоения различных явлений традиционной культуры. Опыт экспедиционных исследований последних лет показывает, что многие пробелы в составе устного поэтического творчества коми возможно восполнить лишь этими архивными записями. Поэтому актуальность обращения к ним связана не только с научными сферами, но и с деятельностью культурных и образовательных учреждений.

Основной массив материалов, собранных в период, начиная с 1939 г. по сегодняшний день, сосредоточен в трёх архивах: 1) Научный архив Коми научного центра Уральского отделения Российской академии наук (рукописные фольклорные материалы с 1939 г.); 2) Фольклорный фонд Института языка, литературы и истории (ИЯЛИ) Коми НЦ УрО РАН (аудио- и видеоматериалы фольклорных экспедиций с 1957 г.); 3) Фольклорный архив Сыктывкарского государственного университета (аудио- и видеоматериалы фольклорных экспедиций с 70-х гг. XX в.). Кроме того, имеются отдельные фонды в Национальном музее (рукописи 1930-х гг. XX в.) и Национальной библиотеке (аудиозаписи экспедиций П.И. Чисталёва) Республики Коми.

Фольклорный фонд ИЯЛИ Коми НЦ УрО РАН на сегодняшний день является самым крупным хранилищем аудио- и видеоматериалов по традиционной культуре коми. Плановое использование магнитофонных записей в полевой работе фольклористов началось с 1957 года. В конце 1970-х гг. было принято решение о создании фонотеки. В 2000 г. она была преобразована в Фольклорный фонд ИЯЛИ.

Территория экспедиционных исследований охватывает все районы республики и районы расселения коми за её пределами (Архангельская,



Кировская, Тюменская, Мурманская области, Пермский край). На сегодняшний день фонд включает 26 порайонных собраний. Записи конца 1950–1980-х гг. связаны с именами известных коми фольклористов – П.И. Чисталёва, А.К. Микушева, Ф.В. Плесовского, Ю.Г. Рочева. Отметим, что в фонде хранятся записи не только фольклорных, но и этнографических и диалектологических экспедиций.

Ранние аудиозаписи, вплоть до 70-х гг. XX в., представляют собой образцы, прежде всего, песенного фольклора, причитаний, жанров детского фольклора, сказок. Каждый из собирателей при фиксации материала на аудиоплётки ориентировался, прежде всего, на разрабатываемые ими темы научных исследований. Так, П.И. Чисталёвым собраны образцы инструментальной музыки (в последствие им была защищена диссертация по теме «Коми народные музыкальные инструменты»). Кроме того, им разрабатывалась тема русско-коми песенных заимствований, соответственно, в полевых материалах зафиксированы многочисленные образцы русскоязычных песен как в коми традициях, так и в русских поселениях на территории республики. А.К. Микушев выявлял национальные корни фольклорных традиций коми, в его записях представлен, прежде всего, песенный фольклор на коми языке. Ю.Г. Рочевым зафиксирован богатейший материал по сказочной прозе коми (небольшая часть представлена в научно-популярном сборнике «Легенды и предания народа коми» [28]), разнообразный материал по детскому фольклору (тема кандидатской диссертации). В записях Ф.В. Плесовского сделан акцент на фиксации образцов сказочной прозы (им написана монография по коми сказке, которая так и не была издана) и свадебного фольклора (монография «Свадьба народа коми» [29]). В какой-то мере записи коми песенного фольклора отражены в трёхтомнике «Коми народные песни».

Современный полевой материал во многом отличается от записей XX века. Прежде всего, в нём уже нет того огромного пласта музыкально-песенного фольклора в силу угасания или исчезновения певческих и инструментальных традиций. Кроме того, современные полевые исследования направлены на максимально широкий охват традиции: в идеале, здесь можно найти всё, что можно зафиксировать на видео- или аудиоплётку: от описаний различных этнографических объектов до современного фольклора.

За время деятельности фонда была осуществлена аналого-цифровая перезапись магнитофонных лент (собрания 1957–1991 гг.). Это необходимый этап в обеспечении доступности материалов фонда для потребителя. Одним из приоритетных направлений архивной деятельности всегда была и остаётся публикация архивных материалов. В последние годы это



не только сборники текстов, но и звуковые издания: как отдельные диски, так и сборники материалов с аудиоприложениями [30]. Актуален на сегодняшний день вопрос о создании электронного путеводителя по материалам Фольклорного фонда, размещённого в Интернете.

Таким образом, для современного функционирования фольклора и стимулирования активности языкового сообщества необходимо, на наш взгляд:

1. Создать на сайте ИЯЛИ Коми НЦ УрО РАН раздел по коми фольклору, разместив там:

а) аннотированный путеводитель по материалам Фольклорного фонда;

б) часть фонограммархива для свободного доступа в виде файлов в одном из стандартных форматов;

в) электронную библиотеку, включив туда все фольклорные и лингвистические сборники текстов, имеющиеся диски.

2. Издавать фольклорные сборники с аутентичными текстами.

Сегодня есть все основания прогнозировать возрастание роли всего фольклорного наследия в современной культуре коми народа.

* * *

1. *Castren M.A.* Syrjäniläisiä häälauluja. – Helsinki, 1878.

2. Коми мойдан кывъяс (Коми сказки) / сост. А.А. Цембер. – Устьысольск, 1913.

3. Коми мойдан и сыланкывъяс (Коми сказки и песни) / сост. А.А. Цембер. – Устьысольск, 1914.

4. Фольклорной сборник. Важ коми мойдъяс да сыланкывъяс (Фольклорный сборник. Старинные коми сказки и песни) / сост. П.Г. Доронин. – Сыктывкар, 1938.

5. *Анисимов П.А.* Коми сыланкывъяс (Коми песни). – М., 1926.

6. История коми литературы. – Сыктывкар, 1978. – Т. I. – С. 15.

7. Коми мойдъяс (Коми сказки) / сост. Ю.Г. Рочев. – Сыктывкар, 1991.

8. *Wichmann Y.J.* Syrjänische Volksdichtung. – Helsinki, 1916.

9. *Fokos-Fuchs D.R.* Volksdichtung der komi (Syrjänen). – Budapest, 1954.

10. *Redey K.* Zyrian Folklore Texts. – Budapest, 1978.

11. *Kokkonen P.* Syrjänische Texte. Gesammelt von T.E. Uotila. – Helsinki, 1985. – Band I; 1986. – Band II; 1989. – Band III; 1995. – Band IV; 2006. – Band V.



12. *Микушев А.К.* Роль русской науки в развитии коми фольклористики // *Фольклористика Российской Федерации.* – Л., 1975. – С.39.
13. Висер вожса сыланкывъяс да мойдкывъяс (Песни и сказки Вишеры) / сост. И.А. Осипов. – Сыктывкар, 1941.
14. *Микушев А.К.* Иван Осипов да сылӧн книга йылысь (Иван Осипов и его книга) // Висервожса сыланкывъяс да мойдкывъяс (Песни и сказки Вишеры) / сост. И.А. Осипов. – 2-е изд. – Сыктывкар, 1986. – С.233.
15. *Микушев А.К., Чисталев П.И.* Коми народные песни. – 2-е изд. – Сыктывкар, 1993. – Т.І.
16. *Микушев А.К., Чисталев П.И.* Коми народные песни. – 2-е изд. – Сыктывкар, 1994. – Т.ІІ.
17. *Микушев А.К., Чисталев П.И., Рочев Ю.Г.* Коми народные песни. – 2-е изд. – Сыктывкар, 1995. – Т.ІІІ.
18. Ипатьдорса фольклор (Фольклор села Ипатово) / сост. А.К. Микушев. – Сыктывкар, 1980.
19. *Wichmann Y.J.* Op. cit.
20. *Redey K.* Zyrian Folklore Texts. – Budapest, 1978.
21. *Fokos-Fuchs D.R.* Zürjén népköltészeti mutatóanyagok. – Budapest, 1913.
22. *Микушев А.К.* Коми народный эпос. – М., 1987.
23. Посни чой-воклы мойдан кывъяс» (Сказки для маленьких братьев и сестёр) / сост. А.А. Суханова. – Сыктывдін кар, 1921.
24. Мойдан кывъяс (Сказки) / сост. А.А. Суханова. – Усть-Сысольск, 1922.
25. *Рошину Н.* Традиционные формулы сказки. – М., 1974. – С.15.
26. *Пунегова Г.В.* Некоторые наблюдения над языком сказочной прозы начала XX века (по материалам коми сказок, записанных А.А. Цембером) // Актуальные проблемы собирания и публикации фольклорных материалов: сб. науч. трудов к 100-летию издания книги «Коми мойдан кывъяс». – Сыктывкар, 2014. – С.27.
27. *Пунегова Г.В.* Указ. соч. – С.30.
28. Коми легенды и предания / сост. Ю.Г. Рочев. – Сыктывкар, 1984.
29. *Плесовский Ф.В.* Свадьба народа коми. – Сыктывкар, 1968.
30. Памятники коми фольклора: Песенная и инструментальная традиции / сост. Г.С. Савельева, А.В. Панюков. – Сыктывкар, 2006 (CD-диск с приложением текстов); Песенная традиция Удоры: аудиоприложение / сост. Г.С. Савельева, Т.С. Канева, А.В. Панюков // Арт. – 2008. – № 1; Фольклор коми и ненцев Ненецкого автономного округа (в записях Фольклорного фонда ИЯЛИ Коми НЦ УрО РАН 1968–1973 гг.) / сост. А.В. Панюков. – Сыктывкар, 2009 (с аудиоприложением на компакт-диске).



УДК 378.147:371.123

В.С. Кунгурцева,

*канд. экон. наук, преподаватель
кафедры информационно-
коммуникационных технологий
в образовании ФГБОУ ВО
«Самарский государственный
социально-педагогический
университет»*

С.В. Тращев,

*студент кафедры информатики,
прикладной математики
и методики их преподавания
ФГБОУ ВО «Самарский
государственный социально-
педагогический университет»*

ОСОБЕННОСТИ ИНТЕГРАЦИИ СРЕДСТВ ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ В СОВРЕМЕННОЕ ОБРАЗОВАТЕЛЬНОЕ ПРОСТРАНСТВО

В настоящее время существует большое количество научных трудов, исследующих опыт использования информационно-коммуникационных технологий (далее – ИКТ). Однако нельзя сказать, что возможности применения ИКТ в современном образовательном пространстве реализуются в полной мере.

При этом существуют и положительные примеры внедрения средств ИКТ в образовательный процесс учащимися школ и университетов. В качестве такого примера можно рассмотреть работы выпускников программы «Обучение для будущего», целью которой является обучение школьных учителей методу проектов: http://www.iteach.ru/exp/learn_projects.php. Однако далеко не вся описываемая деятельность отвечает критериям учебно-го проекта в рамках концепции личностно-ориентированного образования.

ИКТ сегодня являются основным элементом системы инновационного модернизационного развития многих отраслей, и система образования здесь не является исключением. В историческом разрезе под ИКТ понимали различные процессы и средства. Так, Е.С. Полат под ИКТ предлагает понимать реализацию метода проектов, в которых дидактическая



цель завершается осязаемым и практическим результатом [1]. По мнению Т.А. Ильиной, система – это «выделенное на основе определённых признаков упорядоченное множество взаимосвязанных элементов, объединённых общей целью функционирования и единства управления, и выступающее во взаимодействии со средой как целостное явление» [2]. То есть система – это единое целое, состоящее из отдельных компонентов, объединённых внутренней связью.

Уход от традиционного понимания системы образования является задачей не только реформации, но и активного применения разработок ИКТ. Аналитические исследования *Ambient Insight* подтверждают повышение интереса к информационным технологиям в равной степени как у преподавателей, так и у студентов. В частности, их внимание привлекают интерактивные технологии: социальные сети, мобильные приложения, средства моделирования и др.

Сегодня применение ИКТ в образовании ограничено в основном финансовыми и организационными возможностями учебных заведений. Однако необходимо отметить общий позитивный тренд в адаптации образовательного процесса в высших учебных учреждениях к реалиям и потребностям современного информационного общества.

Информационные технологии понимаются различными авторами по-разному: М.Ю. Бухаркина под понятием ИКТ системы образования подразумевает систему методов и технологий [3]; В.В. Краевский, А.В. Хуторской считают, что метод обучения – это «конструируемая с целью реализации в конкретных формах учебной работы модель кооперированной деятельности по преподаванию и учению, представленной в нормативном плане и направленной на передачу обучающимся и усвоение ими определённой части содержания образования» [4]. Исходя из анализа научной литературы нами была составлена схема, наглядно иллюстрирующая роль ИКТ в системе обучения (см. *рис.*).

Итак, ИКТ в образовании подразумевают реализацию методов образовательных и технических процессов, основанную на эффективной реализации интеллектуального потенциала.

Ранее технологии ИКТ ограничивались организацией проектных предметов, открытых уроков, сегодня информационные средства представляют собой широкий пласт возможностей: от пользования электронными дневниками до возможности полноформатного виртуального присутствия учащегося на образовательном мероприятии.

По прогнозам зарубежных аналитиков, уже в 2020–2030 гг. традиционное образование будет уступать системам электронного удалённого обучения, а бумажные носители могут быть заменены исключительно теле-

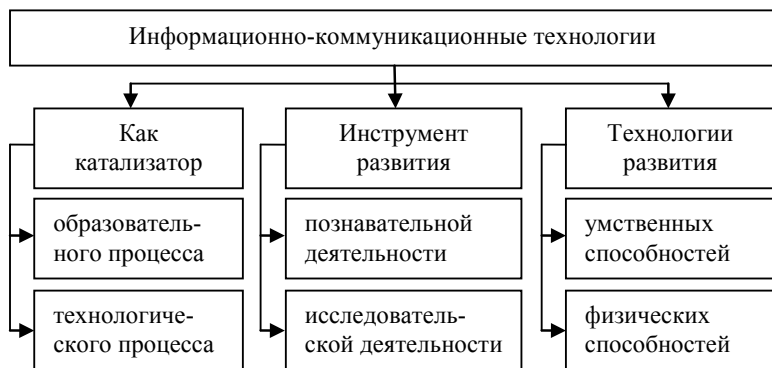


Рис. Роль ИКТ в системе обучения

коммуникационными контентами и электронными публикациями [5]. Однако к российскому образовательному пространству данные прогнозы в полной мере экстраполировать не представляется возможным, поскольку на пути к активному внедрению информационных технологий в систему российского образования есть немало препятствий. К основным из них можно отнести сложности при составлении методических материалов по дисциплинам, реализуемым по средствам ИКТ, необходимость затрачивания преподавателем дополнительных временных ресурсов на интеграцию средств ИКТ в образовательный процесс.

Другой проблемой внедрения ИКТ является низкий уровень компьютерной грамотности учащихся, проживающих в удалённых от центра Федерального округа населённых пунктах.

В результате приходится быть осторожным на пути преимущества информационных технологий в системе образования. И важно помнить, что ни одна автоматизированная система не сможет проверить порядок изложения сочинения школьника, заменить все свойства печатной книги и, конечно же, социальную коммуникацию.

Таким образом, мы можем сделать вывод о том, что все компоненты системы обучения образуют устойчивое единство, которое обладает интегративными свойствами и подчинено общим целям образования и воспитания. Предметная деятельность преподавания и учения является объединяющим фактором всех компонентов системы обучения. Решение проблемы использования ИКТ кроется в современных разработках и восполнении дефицита технологий образовательной среды.



* * *

1. *Полат Е.С., Бухаркина М.Ю., Моисеева М.В., Петров А.Е.* Новые педагогические и информационные технологии в системе образования: учеб. пособие для студентов пед. вузов и системы повыш. квалиф. пед. кадров / под ред. Е.С. Полат. – М.: Академия, 1999.
2. *Ильина Т.А.* Системно-структурный подход к организации обучения. – URL: <http://physics.herzen.spb.ru/teaching/materials/gosexam/b25.htm>
3. *Бухаркина М.Ю.* Методика использования современных информационно-коммуникационных технологий в проектной деятельности в школьном образовании. – URL: <http://cdn.scipeople.com/materials/8236>
4. *Краевский В.В., Хуторской А.В.* Основы обучения. Дидактика и методика: учеб. пособие для студ. вузов. – М.: Академия, 2007.
5. URL: <https://rg.ru/2013/10/09/obuchenie-site-anons.html>



УДК 004.912

Н.А. Леонтьев,

*канд. техн. наук, доцент кафедры
радиотехники и информационных
технологий Северо-Восточного
федерального университета
им. М.К. Аммосова (г. Якутск)*

Г.Г. Торотоев,

*канд. филол. наук, директор
Института языков и культуры
Народов Северо-Востока РФ
Северо-Восточного федерального
университета им. М.К. Аммосова
(г. Якутск)*

МНОГОПОЛЬЗОВАТЕЛЬСКАЯ МОРФОЛОГИЧЕСКАЯ РАЗМЕТКА КОРПУСА ЯКУТСКОГО ЯЗЫКА

Корпусные исследования языков народов России развиваются различными учёными, хорошо проработаны, например, чувашский [1], башкирский [2], тувинский [3], татарский и другие языки. В разработке находятся морфологические и синтаксические анализаторы, имеется большое количество прикладных программ.

Исследование корпусной лингвистики якутского языка прошло стадию накопления материала, были созданы системы для идентификации языка, статической обработки, идут работы по автоматическому морфологическому анализатору. В рамках исследования автоматической обработки якутского языка был создан газетный корпус якутского языка [4]. Объём корпуса составляет более 12 млн словоупотреблений.

Для многопользовательской разметки был создан программный код для веб-сервера с допуском для пользователя.

При выборе текста пользователем создаётся запись, где описываются тэги в определённом тексте. В тексте каждое слово автоматически размечается как текстовая единица через тэг [tN], где N – номер тэга. При добавлении или изменении тэга происходит изменение записи.

Преимущества данного выбора записи разметки в том, что каждый участник проекта будет знать и сохранять свой вариант разметки. При совпадении разметки тэгов автоматически можно будет считать правильным



сделанную разметку. При разных разметках данный тэг идёт на проверку модератору.

В числе проблем можно выделить следующие:

1. Необходимо чтобы один текст разметили минимум 3 раза случайные пользователи или же один модератор.
2. Возможны предумышленные действия пользователей.
3. Возможны непредумышленные ошибки пользователей.
4. Необходимо хорошее знание пользователем всей грамматики.

Разметить весь текст бывает сложно.

Пример расстановки тэгов в тексте:

[T1]= Кыым-85 [T2]Сонун, [T3]сэргэх [T4]буолан [T5]иһэр

Пример разметки документа:

[T1]= N

[T2]=ADJ

[T3] =ADJ

[T4]=CONV

[T5]=V

Т а б л и ц а

Части речи якутского языка

Сокращения	Расшифровка сокращений	Название категории
N	Noun	Имя существительное
POSS	Possessive	Имя притяжательное
PRO	Pronoun	Местоимение
NUM	Numeral	Имя числительное
ADJ	Adjective	Имя прилагательное
V	Verb	Глагол
PCP	Participle	Причастие
CONV	Converb	Деепричастие
ADV	Adverb	Наречие
MOD	Modal word	Модальное слово
INTJ	Interjection	Междометие
CONJ	Conjunction	Союз
PART	Particle	Частица
POST	Postposition	Послелог
IMIT	Imitative word	Звукоподражательное слово

В таблице приведены разметка и части речи якутского языка, согласно которым происходит обработка документа. Данная разметка разработана для якутского языка с учётом общетюркской морфологической разметки.



Разделение разметки и содержимого позволяет в дальнейшем произвести также синтаксическую разметку и лексико-семантическую доработку программы доступа к базе.

* * *

1. *Желтов П.В.* Создание национального корпуса чувашского языка: проблемы и перспективы // Современные проблемы науки и образования. – 2015. – № 1-1; URL: <http://www.science-education.ru/ru/article/view?id=19046> (дата обращения: 27.01.2017).

2. *Сиразитдинов З.А., Бускунбаева Л.А., Ишмухаметова А.Ш., Ибрагимова А.Д.* Информационные системы и базы данных башкирского языка. – Уфа, 2013.

3. *Салчак А.Я., Байыроол А.В.* Электронный корпус тувинского языка: состояние, проблемы // Мир науки, культуры, образования. – 2013. – № 6 (43). – С.408–409.

4. *Леонтьев Н.А.* Национальный корпус интернет-сайтов газет на якутском языке // Журнал научных и прикладных исследований. – 2014. – Т.4. – С.53–54; *Leontiev N.A.* The newspaper corpus of the yakut language // Proceeding of the International Conference «Turkic Languages Prosessing: TurkLang – 2015». – 2015. – P.233–235; *Леонтьев Н.А.* Вопросы автоматизированной лемматизации якутского языка // Современная наука: актуальные проблемы теории и практики. Серия: Познание. – 2015. – № 2 (41). – С.15–16.



А.Н. Лобыгин,

*канд. пед. наук, доцент кафедры общей
и физической географии Института
естественных наук Удмуртского
государственного университета
(г. Ижевск)*

Н.Н. Тимерханова,

*канд. филол. наук, доцент,
зав. кафедрой лингвистической
типологии и лингводидактики
Удмуртского государственного
университета (г. Ижевск)*

ЭЛЕКТРОННОЕ УЧЕБНОЕ ПОСОБИЕ ПО КРАЕВЕДЕНИЮ НА УДМУРТСКОМ ЯЗЫКЕ

На уроках по предмету «Окружающий мир», связанных с региональным содержанием, необходимо использовать учебные пособия, освещающие различные вопросы, имеющие отношение к конкретному региону. Однако на данный момент по предмету «Окружающий мир» нет учебных пособий с региональной географической тематикой для начального общего образования Удмуртской Республики, кроме учебного пособия А.Ю. Украинцевой и Н.Т. Козловой «Мой край – Удмуртия» [1] для 2 класса.

При этом изучение особенностей родной республики необходимо продолжать и в 3 и 4 классах. Учителям нужны учебные материалы для дальнейшего изучения родного края, в том числе и географических тем, поэтому необходимо создать условия, при которых учитель кроме своих материалов к уроку также сможет использовать и другие, проверенные с научной точки зрения и адаптированные к определённому возрасту обучающихся готовые учебные пособия с рекомендациями для учителя по использованию этих учебных материалов.

Поэтому нашим авторским коллективом в течение 2016–2017 гг. было подготовлено электронное учебное пособие к учебнику «Окружающий мир» (начальные классы) «По родным просторам. Удмуртия – край мой родниковый» с региональным содержанием для школ Удмуртской Республики. Пособие было предложено для апробации гимназии № 56 г. Ижевска.

Электронное учебное пособие является комплексным. Оно состоит из интерактивного анимированного электронного учебного пособия Н.Н. Ти-



мерхановой и А.Ф. Кудрявцева «По родным просторам. Удмуртия – край мой родниковый» (см. *рис. 1*), материал которого построен в форме игры-путешествия по родному краю – Удмуртской Республике. Методические рекомендации для учителя по использованию электронного ресурса подготовлены А.Н. Лобыгиным. В дальнейшем планируется издание CD-диска, в который также будут включены конспекты, рекомендации учителей, работающих на инновационной площадке по проекту «Разработка и апробация электронного учебного пособия “По родным просторам. Удмуртия – край мой родниковый” с региональным содержанием для школ Удмуртской Республики».

Материал пособия связан с темами учебников:

- 1) «Окружающий мир. 4 класс», авторами которого являются А.А. Плешаков и М.Ю. Новицкая [2];
- 2) «Окружающий мир. 4 класс», авторы – А.А. Плешаков и Е.А. Крючкова [3];
- 3) «Окружающий мир» (4 класс), авторы – О.Т. Поглазова, Н.И. Ворожейкина, В.Д. Шилин [4].

Возможно, подойдут и к другим учебникам.

Поскольку электронное учебное пособие «По родным просторам. Удмуртия – край мой родниковый» было создано для решения задачи совершенствования учебного процесса при изучении предмета «Окружающий мир» в начальной школе, оно снабжено комплексом различных учебных элементов. Диск состоит из 12 тематических блоков, а также содержит электронные игры (см. *рис. 2*), практические задания и аудиофайлы. Содержание пособия охватывает основные элементы географической составляющей программы, углубляет и закрепляет базовые понятия и представления с опорой на краеведческий материал.

При формировании содержания учебного диска использован методический приём привязки внимания ученика к сказочному персонажу – Крылоконю. Данный персонаж взят из древней удмуртской легенды, от его лица происходит знакомство с различными объектами и явлениями. В системе нескольких занятий у учеников формируется психологическая привязанность к сказочному существу. Упрощённые формулировки в обращении Крылоконя к другим представителям природы за разъяснениями помогают снимать психологическое напряжение при восприятии сложных понятий, научных объяснений, причинно-следственных связей.

На отдельных слайдах презентаций представлено сочетание текста, схем, фотоматериалов, элементов мультимедиа, что позволяет осуществлять одновременный процесс усвоения понятий и представлений, постоянно опирающихся друг на друга. Содержание выстраивается таким об-



разом, чтобы ученики к одним и тем же понятиям возвращались несколько раз с разных позиций. Практические задания, различающиеся на отдельных этапах уровнем и объёмом, формируют общую систему представлений о компонентах окружающего мира.

Учебный диск «По родным просторам. Удмуртия – край мой родниковый» выстроен как краеведческое пособие, раскрывающее географическую характеристику территории Удмуртской Республики. Здесь использован традиционный подход к реализации краеведческого принципа. Ученику сообщается краеведческая информация по тем направлениям, которые рассматриваются одновременно при изучении предмета «Окружающий мир». Психологическое восприятие краеведческого содержания младшими школьниками осуществляется более эмоционально, осознанно. На этом возрастном этапе слабо сформировано абстрактное мышление, и любой общенаучный материал требует конкретного подтверждения. Знакомство с тем, что окружает школьника непосредственно на территории его проживания, является хорошей конкретной базовой основой для формирования системы понятий и представлений по географии.

Ещё один важный приём, реализуемый в данном электронном учебном пособии, являющийся ведущим и традиционным в методике обучения географии – это постоянная привязка любого общенаучного содержания, любых общегеографических описаний к конкретным территориям. Данный приём становится опорным при выстраивании основного содержания пособия: путешествие по территории Удмуртии с последовательным посещением отдельных уголков родного края с обязательным рассмотрением картосхем, маршрутов путешествий (см. *рис. 3*).

Таким образом, пособие одновременно реализует комплекс учебных задач: закрепление и углубление программного материала предмета «Окружающий мир», его географического содержания; знакомство с особенностями природы родного края, Удмуртской Республики. Включение в содержание пособия легенд о родном крае, рассказов о путешествиях и приключениях сказочных героев воспринимаются детьми с повышенным интересом, на высоком познавательном уровне. Система игр и практических заданий позволяет реализовать в учебном процессе деятельностный подход, постоянно держит ученика в активном познавательном состоянии.

Необходимо отметить, что в пособии большое внимание уделено межпредметным связям. В него включены мифы (см. *рис. 4*), стихотворения, загадки, фразеологизмы, поговорки, афоризм, песня, задачи. Интегрированные уроки способствуют формированию целостной картины мира у детей, пониманию связей между явлениями в природе в целом, выявлению прямых связей, например произведений, изучаемых по литературе, с объектами, яв-

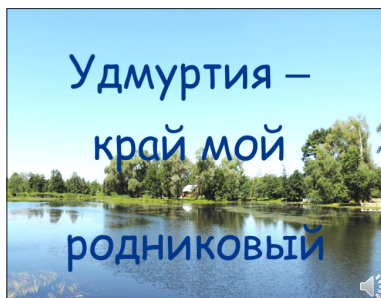


Рис. 1. Титульный экран пособия

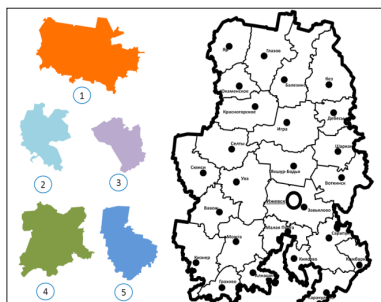


Рис. 2. Пример игры

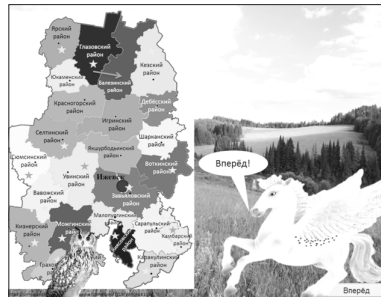


Рис. 3. Пример задания

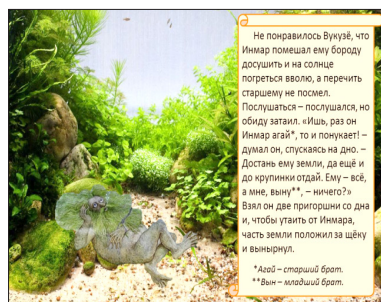


Рис. 4. Пример мифа, включённого в материал пособия



лениями, которые в них описываются (и наоборот, несовпадение с объективной действительностью, например в мифе о сотворении земли). Интеграция учебного материала способствует развитию творческого мышления учащихся. Межпредметные связи также позволяет учащимся применять полученные знания в реальных условиях, например знания для решения задачи на расчёт времени по скорости и расстоянию или на производительность.

Отдельные элементы электронного пособия были реализованы на уроках по предмету «Окружающий мир» в начальной школе в гимназии № 56 г. Ижевска. Отмечены высокая активность и повышенный интерес учащихся в работе на уроках с представленными материалами. Пособие хорошо вписывается в учебный процесс, позволяет менять объём используемых материалов в зависимости от наличия времени на уроке с учётом скорости учебной работы и степени усвоения содержания детьми в разных классах. Предполагается комплексный педагогический эксперимент включения пособия в учебный процесс в Удмуртской государственной национальной гимназии имени Кузебая Герда.

* * *

1. *Украинцева А.Ю., Козлова Н.Т.* Мой край – Удмуртия. 2 класс: учеб. пособие для общеобразовательных учреждений. – Ижевск: Удмуртия, 2013.

2. *Плешаков А.А., Новицкая М.Ю.* Окружающий мир. 4 класс: учеб. для общеобразовательных учреждений: в 2 ч. – 3-е изд. – М.: Просвещение, 2013. – Ч.1.

3. *Плешаков А.А., Крючкова Е.А.* Окружающий мир. 4 класс: учеб. для общеобразовательных учреждений: в 2 ч. – 8-е изд. – М.: Просвещение, 2010. – Ч.1.

4. *Поглазова О.Т., Ворожейкина Н.И., Шилин В.Д.* Окружающий мир: учеб. для 4 класса общеобразовательных организаций: в 2 ч. – 9-е изд., перераб., доп. – Смоленск: Ассоциация XXI век, 2015. – Ч.1.



Е. Миллер,

*студентка школы лингвистики
Национального исследовательского
университета «Высшая школа
экономики»*

АВТОМАТИЧЕСКОЕ ВЫРАВНИВАНИЕ СЛОВАРЕЙ ЛИТЕРАТУРНОГО УДМУРТСКОГО ЯЗЫКА И БЕСЕРМЯНСКОГО ДИАЛЕКТА

Создавая компьютерно-лингвистические инструменты для языков народов России, разработчики сталкиваются с отсутствием базовых ресурсов. Одним из видов таких ресурсов являются достаточно крупные корпуса и построенные на их основе *n*-граммные языковые модели [1], которые необходимы для создания широкого спектра инструментов от распознавания речи до спеллчекеров. Другим, ещё более редко встречающимся, ресурсом является корпус с ручной морфологической разметкой или со снятой омонимией.

В своей работе мы использовали материалы двух вариантов удмуртского языка: бесермянского диалекта и литературного удмуртского. Литературный удмуртский обладает корпусом объёмом около 7,3 млн словоупотреблений с морфологической разметкой без снятия омонимии, который подходит для создания языковой модели. Для бесермянского диалекта группой московских лингвистов собран корпус устных текстов объёмом около 70 тыс. словоупотреблений, расшифрованный и морфологически размеченный вручную. Целью работы является создание некоего автоматического преобразователя, который позволил бы использовать языковую модель литературного удмуртского при разработке инструментов для бесермянского диалекта и, наоборот, использовать бесермянский корпус в инструментах для литературного удмуртского. Результаты работы планируется использовать при разработке морфологического дизамбигуатора для литературного удмуртского и создании бесермянского распознавателя речи (аналогично тому, как это делалось в работе [2]).

Первый этап работы был посвящён созданию подобия бесермянско-удмуртского переводчика на уровне лексем. Для этого были использованы материалы бесермянского и удмуртского словарей и методы машинно-



го обучения. Для удобства удмуртские лексемы были транслитерированы из кириллицы в латиницу. Задача сводилась к обучению алгоритма по нескольким признакам определять, являются ли данные удмуртская и бесермянская лексемы родственными и близкими по значению. Были использованы следующие признаки:

- расстояние Левенштейна: определялось минимальное количество операций вставки, удаления и замены символов, необходимых для превращения удмуртской лексемы в бесермянскую. Вес операций определялись в зависимости от регулярности перехода (например, переход *ö* в *o* имел вес 0,2, в то время как переход *r* в *m* имел вес 1);

- семантическая близость переводов: слова из бесермянского перевода сравнивались со словами из удмуртского с помощью сервиса *RusVectōrēs*, который вычисляет косинусную близость между лексическими векторами (так, сходство переводов слов **ž'ek** ('стол') и **žer** ('рожь') равно 0,099, а сходство **ardok** ('промоина') и **ардо** ('польня') – 0,566);

- соответствие частей речи;
- длины удмуртской и бесермянской лексем;
- поле «*UdmurtOrtho*» в бесермянском словаре, частично заполненное вручную авторами словаря, – вариант лексемы в удмуртской орфографии.

После автоматической разметки лексем мы создали выборку из 200 пар слов, в которую включили 100 пар, где прослеживается соответствие бесермянской лексемы удмуртской, и 100 пар, где этого не происходит. При этом учитывались разные типы соответствий, с разной степенью проявления тех или иных признаков.

На основе обучающей выборки была создана модель (с использованием метода логистической регрессии и бинарной классификации), с помощью которой алгоритм учился определять, к какому из двух классов относится данная пара слов. Точность такой модели составила 93%; окончательно соответствия доводятся вручную.

В конечном итоге на основе автоматически выровненных удмуртского и бесермянского словарей и прописанных вручную соответствий грамматических категорий (морфем) планируется создать бесермянско-удмуртский и удмуртско-бесермянский переводчик. Такой преобразователь позволит использовать преимущества корпуса одного варианта удмуртского при разработке инструментов для другого и наоборот.



* * *

1. *Jurafsky D., James H. Martin.* Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. – Pearson Prentice Hall, 2009.

2. *Smit P., Leinonen J., Jokinen K., Kurimo M.* Automatic Speech Recognition for Northern Sámi with comparison to other Uralic Languages // Proceedings of the Second International Workshop on Computational Linguistics for Uralic Languages. – 2016. – December.



А.Г. Мусанов,
зав. сектором языка
Института языка, литературы
и истории Коми НЦ УрО РАН
(г. Сыктывкар)

ГЕОИНФОРМАЦИОННАЯ СИСТЕМА «ГЕОГРАФИЧЕСКИЕ НАЗВАНИЯ РЕСПУБЛИКИ КОМИ»

В современных интеграционных условиях этнические сообщества нуждаются в воспроизводстве своего этнического мира не столько в материальном плане, сколько в сфере информации, в создании единых национальных фондов, своеобразных этноархивов, моделирующих этническое пространство в информационных системах нового типа. Это в полной мере касается национально-региональной топонимии – составной части историко-культурного наследия народов, проживающих на определённой территории. Топонимическая система любого языка, являясь продуктом длительного исторического развития, содержит исключительно богатую и разностороннюю историко-этнографическую информацию. Чтобы стать доступной широкому кругу пользователей, эта информация должна быть максимально адаптирована к функционированию в современных информационных сетях, т.е. соответствующим образом обработана и представлена в виде автоматизированных информационно-поисковых баз данных, геоинформационных аналитических систем.

Геоинформационно-аналитическая система «Географические названия Республики Коми» – это электронная база, созданная для решения комплекса научно-исследовательских проблем, связанных с изучением истории коми языков (шире – пермских), исторических пластов в его лексическом составе, явлений исторической фонетики и грамматики, субстратных явлений в лингвогенезе и т.д. Эмпирические материалы, содержащиеся в базе данных, станут основой для решения проблем этногенеза и этнической истории народов данной территории, вопросов лингвогенеза и диалектообразования коми языка с учётом его контактирования с другими родственными и неродственными языками. Вместе с тем возможно использовать данную систему для решения практических задач функционирования коми языка в регионе, в том числе как одного из государственных языков



Республики Коми. Выполнение этих задач требует, прежде всего, разработки единых, последовательно применяемых принципов и правил ввода и обработки данных; создания нормализованного, канонически непротиворечивого свода географических названий на русском и коми языках, который позволил бы составить точные списки имён географических объектов на двух государственных языках республики, разработать правила употребления и адаптации. В свою очередь, это даст возможность нормализовать использование географических наименований на географических, топографических, административных и др. картографических материалах, устранить имеющиеся неточности и ошибки, унифицировать их написание в различных справочных изданиях, документах государственного и федерального значения, средствах массовой информации, в службах связи и транспорта, в современных навигационных системах GPS, ГЛОНАСС и т.д. Последовательная привязка электронных баз данных к картографическим объектам позволит проводить ареальный анализ топонимических материалов.

ГИС-технологии обладают большим аналитическим потенциалом при решении сугубо теоретических ономастических (функционирование топонимических систем в пространстве, формирование топонимических ареалов и т.д.), этнографических и этнокультурных проблем. Применение геоинформационных систем позволяет моделировать реальную ситуацию образования и функционирования топонимических систем региона.

В финно-угроведении имеется несколько крупных ономастических центров, в которых на основе собранных ранее архивных топофондов создаются электронные базы данных. Фонды финно-угорской топонимии имеются в Ономастическом архиве г. Уппсала (Швеция), в Ономастическом центре г. Осло (Норвегия), в Институте языков Финляндии (Хельсинки), в Институте языкознания Эстонской Академии наук (Таллин), в Институте венгерского языкознания (Дебрецен), в России – в Уральском государственном университете (Екатеринбург), Карельском научном центре (Петрозаводск) [1], Коми научном центре (Сыктывкар) [2], Мордовском государственном университете (Саранск) [3] и др.

Большая работа по созданию информационного компьютерного банка топонимических данных ведётся в государственном научно-исследовательском центре языков Финляндии (KOTUS), где уже с 1985 г. результаты полевых исследований по топонимии вносятся в компьютер. В базу данных топонимов включены как лингвистические сведения о топониме, так и экстралингвистические сведения об объекте, обозначаемом этим топонимом. Р. Мииккулайненом разработана структура описания банка финских топонимов, состоящая из 17 полей записей (*Miikkulainen R. Kotimaisten*



kielten tutkimuskeskuksen paikannimitieto kannan kirjoittaminen). В Институте языка, литературы и истории Карельского научного центра с 1991 г. авторская группа под руководством М.В. Горбаневского ведёт разработку принципов формализации «Многоцелевого компьютерного топонимического словаря». Каждая словарная статья состоит из 5 модулей – адресного, хронологического, этимологического, историко-культурного и нормативного. С 1995 г. работу в этом направлении ведёт авторская группа под руководством Г.М. Керта, разрабатывается «Компьютерный банк топонимии Европейского Севера России».

Классификация топонимов по структуре не вызывает особых трудностей, поскольку структура географических названий в известной степени отражает грамматику слов и словосочетаний на апеллятивном уровне. Сложность здесь состоит в многообразии и «нестандартности» описываемого материала, поскольку при вводе конкретного топонима всегда существует возможность его неоднозначной интерпретации. Поэтому оператор, осуществляющий ввод материала, должен достаточно хорошо владеть методикой лингвистического (морфемного, морфологического и словообразовательного) анализа.

Значительно сложнее характеризовать топонимический материал с точки зрения семантики компонентов. Очевидно, наиболее перспективной в этом плане может быть классификация, учитывающая ограниченное количество наиболее типичных мотивационных установок номинации. При первичной обработке материала учитывается: 1) вид описываемой реалии (гидроним, ороним, ойконим и т.д.); 2) отнесённость к одной из основных семантических групп: А – номинация по свойствам обозначаемого объекта, В – номинация по связи обозначаемого объекта с человеком, С – номинация от собственных имён (отантропонимические и оттопонимические имена), 3) соотносённость с конкретным типом номинации: А1 – характеристика по признаку предмета: А11 – по величине объекта, А12 – по цвету объекта, А13 – по звучанию, А14 – по геологическому строению объекта, А15 – по рельефу местности, А16 – по типу флоры, А17 – по типу фауны, А18 – по расположению, А19 – метафорические топонимы и др.; В1 – номинация по хозяйственным занятиям человека, В2 – номинация по религиозно-мифологическим верованиям, В3 – номинация по объектам хозяйственного назначения, средствам сообщения, В4 – номинация по социальным отношениям, В5 – этэтнонимные топонимы, В6 – номинация по национальным отношениям, В7 – окказиональные топонимы; С1 – патронимы, С2 – личные имена, С3 – отчества и фамилии, С4 – прозвища, С5 – оттопонимические наименования и др. Предполагается дальнейшая разработка семантического классификатора.



Проблема ввода топонимических единиц по семантическим признакам заключается прежде всего в том, что семантические связи и отношения в топонимообразовании весьма разнообразны, поэтому список лексических групп остаётся открытым. Кроме того, топоним может иметь несколько мотивационных установок, или эта мотивация утрачена. В топонимической базе большой пласт составят наименования, значение которых никак не объясняется средствами современного коми или русского языков.

В основном это субстратная топонимия и деэтимологизированная лексика коми языка. При вводе этой лексики в базу данных топоним характеризуется только с точки зрения вида описываемой им реалии (гидроним, ороним, ойконим и т.д.). При желании, пользователь может найти информацию о происхождении топонима и возможной его мотивации в этимологическом словаре, предусмотренном системой.

Важной частью автоматизированной информационной системы является представление средствами геоинформационных технологий единой цифрованной многофункциональной карты Республики Коми и сопредельных областей (в том числе Коми-Пермяцкого автономного округа). Она позволит не только наглядно представить местоположение конкретного топонима в системе географических координат, оперативно вывести на экран любые фрагменты карты, при необходимости уменьшая или увеличивая масштаб и высвечивая слои крупных, средних, мелких топонимов и микротопонимов, но и с помощью специально разработанной системы поиска и гипертекстовой поддержки выявить нужные пласты топонимической лексики в зависимости от потребностей пользователя.

* * *

1. *Керт Г.М.* Применение компьютерных технологий в исследовании топонимии (прибалтийско-финская, русская). – Петрозаводск, 2002; *Керт Г., Вдовицын В., Веретин А.* Компьютерный банк топонимии Европейского Севера России. – Петрозаводск, 1998.

2. *Мусанов А.Г.* Перспективы и возможности применения ЭВМ для обработки топонимических данных // XIV Коми республиканская молодёжная науч. конф. – Сыктывкар, 2000. – Т.1. – С.17–18; *Мусанов А.Г.* Проблемы создания компьютерного банка коми топонимических данных // Развитие коми языка на современном этапе: проблемы и перспективы. – Сыктывкар, 2000. – С.130–132; *Мусанов А.Г.* Технические приёмы в исследовании географических названий Республики Коми // Коми-зырянская культура и финно-угорский мир. – Сыктывкар, 2002; *Мусанов А.Г., Федюнева Г.В.* Проблемы создания коми топонимического фонда Республики



Коми: лингвистический аспект // Научные доклады / Коми научный центр УрО РАН. – Сыктывкар, 2002. – Вып. 450.

3. *Рябов И.Н.* Структура базы данных компьютерного топонимического фонда Республики Мордовия // Лексика мордовских языков: проблемы и перспективы развития: материалы Международной науч.-практ. конф., посвящённой юбилею Хейкки Паасонена / отв. ред. М.В. Мосин. – 2010. – С.85–88.



УДК 811.351.22

Р.О. Муталов,

д-р филол. наук, профессор,

гл. научный сотрудник

отдела кавказских языков

Института языкознания РАН

(г. Москва)

М.Р. Муталов,

начальник отдела web-технологий

информационно-вычислительного

центра Дагестанского

государственного университета

(г. Махачкала)

МИКРОСТРУКТУРА ЭЛЕКТРОННОГО СЛОВАРЯ ДАРГИНСКОГО ЯЗЫКА*

Даргинский язык относится к нахско-дагестанской ветви северокавказских языков; число говорящих – более 600 тыс. человек. По данному языку пока отсутствуют изданные даргинско-русский, толковый, этимологический и другие словари; задача создания различных словарей остаётся актуальной. В последние годы информационные технологии стали широко применяться в лингвистической науке, как и в других областях науки и жизни. Особенно быстро стала развиваться компьютерная лексикография, позволяющая лингвистам намного облегчить процесс подготовки электронных словарей, которые дают возможность пользователю быстро найти необходимую информацию. Остаётся актуальной и проблема сбора и сохранения материала языков малых народов, поскольку из-за глобализации очевидным стал процесс их интенсивного исчезновения.

Группой лингвистов и программистов из Москвы и Махачкалы был подготовлен электронный словарь даргинского языка. Следует отметить, что в даргинском языке функционирует большое число диалектов, сильно отличающихся друг от друга, вследствие чего в последние годы они рассматриваются в качестве самостоятельных языков [1]. Электронный словарь подготовлен на базе литературного языка (основа – акушинский диалект). Основой словаря стал «Большой даргинско-русский словарь» (грант

© Муталов Р.О., Муталов М.Р., 2017

* Исследование выполнено при финансовой поддержке РГНФ (проект № 15-04-12013 в «Информационно-поисковая система “Электронный словарь даргинского языка”»).



РГНФ, проект № 10-04-00353а). Важное значение при подготовке данного ресурса придавалось принципам построения словарной статьи электронного словаря.

Электронный словарь представляет собой особый лексикографический объект, в котором могут быть реализованы и введены в обращение многие продуктивные идеи, не востребованные по разным причинам в бумажных словарях [2]. По сравнению с бумажной версией, электронные словари обладают рядом преимуществ. В такой словарь можно вносить новые слова, дополнять и изменять уже имеющуюся информацию; он даёт возможность также быстро получить необходимую информацию о слове. Электронный словарь может включать в себя не только тексты словарных статей, но и различные иллюстрации, анимацию, видео, звук.

Словарь, как известно, состоит из макро- и микроструктуры. Под макроструктурой словаря понимается организация корпуса словаря или словника; она определяет порядок расположения словарных статей. В электронном словаре даргинского языка принят алфавитный порядок расположения заглавных слов.

Микроструктура словаря представляет собой организацию совокупности данных, заключённых в пределах словарной статьи. Она определяется задачами словаря и включает в себя заголовочную единицу и текст, разъясняющий, описывающий её основные характеристики и представленный различными зонами. По мнению В.В. Дубичинского, микроструктура словаря представляет собой «построение словарной статьи как особой разноплановой системы внутри всего словаря как сложной самостоятельной системы» [3].

Основными компонентами структуры словарной статьи являются заглавное слово, зона лексического значения, информация о грамматических свойствах лексемы, лексикографические пометы, иллюстративный материал, зона фразеологизмов и паремиологических единиц.

Словарная статья в электронном словаре начинается с заглавного слова, которое пишется прописными буквами и выделяется полужирным шрифтом. Именные части речи даргинского языка даются в форме именительного падежа единственного числа, исходной формой глаголов служит инфинитив. В ходе подготовительных работ был разработан модуль для каталогизации исходных лексем. Данный модуль внедрён в электронный словарь, что позволяет отображать в виде каталога в алфавитном порядке все исходные лексемы даргинского языка.

Одним из основных параметров электронного словаря является озвучивание исходных лексем. С этой целью был разработан модуль озвучивания опорных слов даргинского языка. Данный модуль был подключён и



проверен в системе словаря. Для ввода в электронный словарь были озвучены опорные лексемы даргинского языка, после чего они были размещены в веб-системе. В дальнейшем работу по озвучиванию других исходных лексем планируется продолжить.

Толкование лексического значения предполагает установление количества значений слова и определение каждого значения в отдельности. Объяснение значений даргинских лексем в электронном словаре дано на русском языке. В словаре указаны самые тонкие оттенки значений, которые могут быть выражены данными словами. В тексте словарной статьи различные значения слов пронумерованы арабскими цифрами. Более употребительное значение при этом предшествует менее употребительному. Большинство значений многозначных слов проиллюстрированы примерами.

В даргинском языке довольно часто встречается явление омонимии. В таких случаях для каждого значения даётся отдельное описание; за исходной лексемой ставится висячая арабская цифра. Для разграничения омонимичных слов в электронном словаре дан специальный столбец «Омоним».

В век глобализации в малых языках исчезают целые пласты лексики, связанные с традиционной деятельностью народа. В частности, они исчезают вместе с вымиранием той или иной профессии. В словаре читатель сможет найти различные профессионализмы, которые в современном быте не используются. В художественных произведениях читатели также часто сталкиваются с большим количеством малоупотребительных слов и архаизмов, значения которых для многих уже стали непонятными. Словарь позволит читателю быстро находить значения подобных слов.

Важно снабдить слово не только точным переводом, но и соответствующими грамматическими сведениями и иллюстративным материалом [4]. В электронном словаре дан специальный столбец, указывающий на то, к какой части речи относится то или иное слово. Помимо этого, частеречная принадлежность заглавных слов обозначена также в самом тексте словарной статьи посредством специальных помет. После каждой исходной лексемы в электронном словаре даются показатели грамматических категорий, которые могут функционировать в данном слове. Для именных частей речи и наречий указываются аффиксы числа, падежа и грамматического класса. Для глагольных лексем даются маркеры класса, времени, наклонения, а также переходности глагола. В словаре указаны также суффиксы причастий, деепричастий, масдаров и каузативов, которые могут присоединяться к исходным глаголам. В даргинском языке при словоизменении возможны также различные фонетические явления (ассимиляция, чередование и выпадение звуков, наращение формата); они также нашли своё отражение в словаре.



В словаре широко используются лексикографические пометы, представляющие собой применяемый в лексикографии способ краткой грамматической, стилистической или иной характеристики слова, выраженный принятым сокращением соответствующего термина. К отдельным словам в целях терминологической стилистической и грамматической характеристики перед переводами даются следующие пометы: *мед.*, *бот.*, *бран.*, *дет.*, *анат.*, *иутл.*, *рел.*, *уст.*, *нареч.*, *числ.*, *понуд.*, *многокр.*, и т.д. Если слово или его значение употребляется во многих областях или его отраслевая принадлежность не может быть точно определена, то ставится помета обобщающего характера *спец.* Если слово или выражение в прямом и переносном значении переводится на русский язык одинаково, то оно сопровождается пометами (*прям.* и *перен.*) и общим переводом.

В качестве иллюстративного материала в электронном словаре привлекается значительное количество примеров из произведений даргинских писателей и поэтов. В пределах одного значения иллюстративные примеры в словаре размещены в следующем порядке: на первом месте даны примеры, имеющие прямое значение, затем – переносное или более отдалённое от основного значения слова.

Большое количество словарных статей проиллюстрированы фразеологизмами. Фразеологические выражения и фольклорный материал даются в конце статьи за знаком «ромбик». Пользователь может легко найти в словаре также пословицы и поговорки, используемые в даргинском языке. В словаре данные паремиологические единицы снабжаются пометами *посл.*, *погов.*, затем приводится русская пословица или поговорка такого же или близкого значения. В ряде случаев для раскрытия образно-семантической структуры даргинского оригинала даётся и буквальный перевод. В словаре приведены также благопожелания и проклятия даргинцев, снабжённые, соответственно, пометами *благопож.*, *прокл.*

Пользователь может выбрать словарную статью из каталога либо найти необходимое слово, используя поиск. Для того чтобы найти информацию о слове, необходимо ввести это слово в строку поиска. При этом на экране всплывает список исходных лексем, содержащих данное слово; из списка нужно выбрать необходимую лексему. В нижней части экрана появляется текст словарной статьи, где указаны лексико-семантические значения, грамматические признаки слова и иллюстративный материал.

Таким образом, пользователь может получить исчерпывающие сведения о нужном слове: словарная статья содержит всю необходимую лексико-грамматическую информацию о слове и является своего рода его лингвистическим «паспортом».



* * *

1. *Mutalov R.* The linguistic situation in Central Daghestan // Language in Daghestan: материалы конференции «Язык в Дагестане» (Университет Чикаго (США), 14–15 мая 2015 г.). – Издательство: Франкийский гуманитарный институт, 2015. – С.17–18.

2. *Селегей В.П.* Электронные словари и компьютерная лексикография. – URL: http://www.lingvoda.ru/transforum/articles/selegey_a1.asp

3. *Дубичинский В.В.* Теоретическая и практическая лексикография. – Wien-Charkov: Wiener Slawistischer Almanach, 1998.

4. *Юсупов Х.А., Муталов Р.О.* Национально-русский словарь: словник, переводы, статьи (на примере даргинско-русского словаря) // Кавказские языки. Генетико-типологические общности и ареальные связи: тезисы докладов 2-й Международной науч. конф. – Махачкала, 2010. – С.229–231.



С.В. Нагурная,

канд. филол. наук, учёный секретарь

Института языка, литературы

и истории Карельского НЦ РАН

(Республика Карелия, г. Петрозаводск)

НОВОПИСЬМЕННЫЙ КАРЕЛЬСКИЙ ЯЗЫК В СОЦИАЛЬНЫХ СЕТЯХ

Процессы ревитализации карельского языка, начавшиеся в конце 1980-х гг. на волне этнокультурного ренессанса народов России, затронули основные сферы институционального дискурса – образование, СМИ, науку, культуру. С 1989 г. началось преподавание карельского языка в школах, с 1990 г. – в Петрозаводском государственном университете. В том же году вышел первый номер карелоязычной газеты «*Ota tuua*», впоследствии здесь печатались ливвиковские материалы, а у носителей собственно карельского наречия появился свой печатный орган – «*Vienan Karjala*». С 2014 г. обе газеты объединились в одну с прежним названием; в обновлённом еженедельном издании публикуются материалы на ливвиковском, собственно карельском наречиях, а также тверских говорах и людиковском наречии карельского языка. Получила развитие карелоязычная литература, отличающаяся многообразием жанров. «Созданные карельскими писателями произведения прозы, поэзии, драматургии стали живыми и непосредственными свидетельствами современников о процессе национально-культурного возрождения, становления карельской литературной традиции, укрепления национального самосознания прибалтийско-финских народов Карелии» [1]. Необходимость кодификации и нормирования карельского языка в процессе создания письменности дала толчок более глубокому изучению языковой структуры, активизировалась работа исследователей по систематизации грамматики, лексики, фонетики и фонологии и т.д.

Ревитализационные процессы карельского языка проходят на фоне постоянного сокращения числа как самих карелов, так и владеющих языком. В 2010 г. доля карельского населения в Республике Карелия составляла 7,4% (в 2002 г. – 9,2%), владевших карельским языком было 36,8% из всех указавших владение языками карелов (в 2002 г. – 48,1%) [2]. Большинство владеющих карельским языком – представители старшего поколения. Но и пожилые карелы жалуются, что общаться на родном языке не с кем,



дети, а тем более внуки, как правило, языка не знают. Практически исчезла естественная языковая среда, в которой карельский язык мог бы использоваться, семейное общение на карельском как важнейшее условие передачи языка от поколения к поколению наблюдается всё реже и реже.

Современная ситуация: новые подходы. В сложившейся неблагоприятной этноязыковой ситуации для привлечения внимания к карельскому языку, роста его привлекательности и престижа в последние годы принимаются всё новые меры. Язык становится объектом формирования брендовости, что стало отличительной чертой современной языковой ситуации. С 2012 г. в конкурс «Учитель года Карелии» включена номинация «Учитель родного языка». Преподаватели карельского языка становятся призёрами конкурса, что, несомненно, повышает престиж языка и привлекает к нему внимание. Школьники участвуют в языковых олимпиадах, результаты которых широко освещаются в СМИ. Одним из последних мероприятий стало проведение диктанта для знатоков карельского языка. С целью привлечения будущих студентов к изучению карельского, вепсского и финского языков на кафедре прибалтийско-финской филологии Петрозаводского государственного университета в 2015 и 2016 гг. для школьников 8–10 классов был организован этнокультурный лагерь «Встреча юных на карельской земле».

Создан портал «Этнокультурное образование в Республике Карелия», освещающий новости национальной образовательной политики (<http://edu-rk.ru/index.php/edu/info>). Карельский и вепсский языки стало возможно изучать посредством Интернета. Для прохождения онлайн-курсов желающим достаточно зарегистрироваться на сайте электронных образовательных ресурсов Петрозаводского государственного университета.

Долгожданным событием явилось начало преподавания в кружках для школьников людиковского наречия карельского языка: с 2013 г. – в с. Михайловское Олонецкого района, с 2014 г. – в сёлах Кончезеро Кондопожского района и Святозеро Пряжинского района [3].

В Петрозаводске в 2014 г. Союзом карельского народа впервые был реализован проект «Языковое гнездо выходного дня», в рамках которого дети из двуязычных семей приобщались к языку и культуре вместе с родителями. Курсы карельского языка, ежегодно организуемые в Центре национальных культур и народного творчества в Петрозаводске, привлекают всё большее число слушателей, в 2015/2016 году их окончили более 100 человек. В 2016 г. было подано более 100 заявлений, причём многие из курсантов приходят во второй, третий, четвёртый и даже пятый раз.

Традиционными стали встречи для любителей карельского языка в разговорном клубе «*Paginkanzu*». Анонсы и тематика встреч, а также



другая полезная информация о карельском языке и обо всём, что с ним связано, размещаются на страничке сообщества «*Paginkanzu*» в социальной сети «ВКонтакте» (<https://vk.com/paginkanzu>).

С 2015 г. Информационным агентством «Республика» в Интернете реализуется проект «Уроки карельского» (<http://rk.karelia.ru/special-projects/uroki-karelskogo/uroki-karelskogo-karjalua-tundies>), на страницах сайта можно найти информацию об истории, быте, вере, современной жизни карелов. Заголовки статей, а также часть материалов даны с переводом на карельский язык. Статистика сайта показала, что в среднем каждый выпуск «Уроков карельского» собирает около 2 000 просмотров в первый месяц, т.е. проект имеет свою постоянную аудиторию.

В числе мероприятий по брендированию карельского языка – запуск с 2016 г. компанией «Славмо» линейки карельских молочных продуктов с надписями на карельском языке. В своё время оформление на карельском и вепском языках дорожных указателей и фасадных вывесок стало своего рода способом привлечения внимания карелов и вепсов к нормированной, письменной форме этих языков (за 2006–2015 гг. были изготовлены 101 фасадная вывеска и 88 дорожных указателей). Сегодня выпуск «карельских продуктов» – это способ привлечь внимание к языку не только карелов, но и всего населения.

Культурные инициативы последних лет также направлены на популяризацию карельского языка, в их числе – возрождение национальной карельской игры «*Kuukkä*» (карельских городков), ставшей популярной в последние годы и даже вошедшей в региональный комплекс ГТО. С 2013 г. в соцсети «ВКонтакте» открыта группа, на страницах которой представлены материалы на карельском и русском языках об истории карельских городков, правилах игры, анонсы соревнований (<https://vk.com/kyykkakarjala>).

Культурным событием в республике стало открытие Дома карельского языка (далее – ДКЯ). В январе 2015 г. пресса сообщила, что карельский язык получил прописку в с. Ведлозеро Пряжинского района, где силами волонтеров построен Дом карельского языка – этнокультурный центр, в котором реализуются практические проекты по возрождению и сохранению карельского языка. В 2016 г. ДКЯ выступил в качестве своеобразного фонда, объявившего конкурс малых грантов «*Karjalan kieldy tihieh. Карельского – видимо-невидимо!*» Весной текущего года в ДКЯ планируется открытие языкового гнезда для детей дошкольного возраста. Информация о деятельности ДКЯ обновляется на страничке группы «*Karjalan kielen kodi / Дом карельского языка*» в соцсети «ВКонтакте» (<https://vk.com/public46199459>).



Жизнь языка в сети. Популяризации языка способствует создание национального контента в Интернете. Шире использовать возможности современного медиапространства в целях сохранения малых языков призывали активисты молодёжного финно-угорского движения на XII Конгрессе (2015 г.) Молодёжной ассоциации финно-угорских народов (далее – МАФУН). В резолюции конгресса подчёркивалось, что в условиях современности медиа выступает ключевым каналом трансляции идентичности и культуры, создания единой повестки для каждого народа в отдельности в условиях географической дисперсности расселения и нехватки официальных институций [4]. Медиа создаёт среду функционирования языка для малоиспользуемых финно-угорских языков. МАФУН призывает стимулировать развитие блогосферы и социальных медиа на национальных языках, создавать языковую инфраструктуру и «языковой рынок» для носителей уральских языков – современный контент на уральских языках, приложения и электронные словари для ежедневного использования.

Наряду с родственными языками карельский всё активнее осваивает интернет-пространство. В 2017 г. на странице интернет-радио Пенсионного фонда Карелии появился новый раздел «Слушаем на карельском». С материалами «пенсионных» новостей жители Карелии теперь имеют возможность ознакомиться по-карельски (<http://pfkarelia.ru/2017/02/02/penzien-ballat-da-ruadovuvvet-2>).

Удобной с точки зрения результативности и широты охвата аудитории площадкой для размещения материалов на карельском и о карельском языке стала социальная сеть «ВКонтакте» (некоторые ссылки на карело-язычные сообщества сети сделаны выше). Так, с 2012 г. программы национального вещания ГТРК «Карелия» «*OMIN SILMIN*» («СВОЙ ВЗГЛЯД») транслируются на карельском языке с субтитрами на русском в Глобальной сети (<https://vk.com/club23517911>). С новостями республики телеканала «Россия 1» на карельском языке можно познакомиться также на одной из страничек «ВКонтакте» – «*VIESTIT – KARJALA*» («ВЕСТИ – КАРЕЛИЯ», <https://vk.com/viestitkarjala>).

Одной из наиболее популярных групп соцсети «ВКонтакте» является существующая уже 4 года группа «*Карельский язык / Pagizemmo Karjalakse*» (<https://vk.com/karjalanlanguage>), насчитывающая более 2 тыс. подписчиков. Администратор группы Н. Гилюева, выпускница Петрозаводского госуниверситета, в настоящее время – исследователь карельского языка университета Восточной Финляндии, рассказывает, что созданию сообщества предшествовала публикация постов на карельском языке в «*Facebook*» [5]. На страницах группы представлены разнообразные образовательные, новостные материалы, тематические подборки «Карель-



ского словарика», информация культурной направленности, видеоматериалы (в том числе мультипликационные фильмы) на карельском языке.

Не менее привлекательна как для взрослых, так и для детей группа «*Luvemmo yhtes / Читаем вместе*» (<https://vk.com/club49692726>). Появление группы, по словам одной из её создательниц и автора текстов Т. Барановой, было обусловлено дефицитом литературы для детей на карельском языке [6]. Первоначально в группе публиковались короткие детские рассказы на карельском языке, сопровождаемые забавными иллюстрациями. Впоследствии подбор материалов значительно расширился, сейчас это полноценный образовательный ресурс, где представлена лексика различной тематики, стихотворные произведения и т.д. Но наиболее удачным, на наш взгляд, решением стало включение в материалы сообщества афоризмов, высказываний, философских цитат, текстовых статусов в переводе на карельский язык. Благодаря им язык зазвучал по-новому, демонстрируя свои богатые выразительные способности. Подписчиков не оставят равнодушными переводы на карельский язык таких цитат, как: «Если по одному отвергать людей за любой грех, нам скоро окажется не с кем жить. Жить среди людей – означает прощать... (Р. Фрост)» – «*Ruvennemmo yksitellen hylgämäh kaikkii joga riähkäs, meil terväh ei roi kenenke eliä. Eliä rahvahan keskes, se merkiččöy – prostie...*» (https://vk.com/club49692726?z=photo-49692726_456239100%2Falbum-49692726_00%2Frev). Подобные материалы подтверждают, что возможности карельского языка не ограничиваются бытовым пространством, язык становится своеобразным посредником между читателем и миром философии.

В последние годы социальная сеть буквально пестрит разнообразием групп на карельском языке (или использующих карельский), удовлетворяющих самые разные интересы подписчиков. Идея создания песенной группы «*Pajatammo karjalakse / Поём по-карельски*» (https://vk.com/pajatammo_karjalakse) принадлежит Н. Гиловой, автор большинства переводов – Т. Баранова. В сообществе представлены караоке-версии известных песен на карельском языке. Рэп-альбом на карельском языке можно прослушать в группе «*Ondrei. Karjalankielen räppäri*» (<https://vk.com/karjalanrappi>). Подписчикам сообщества предлагается «... национальный язык и национальные мотивы под классические биты, смешение старины и современности, живые музыкальные инструменты и много другое».

Культурно-просветительскую миссию выполняют сообщества «*Карельские пословицы / Karjalan sananpolvet*» (<https://vk.com/karjalanprivutkat>) (группа дублируется в Твиттере – <https://twitter.com/sananpolvet>), «*Karjalazien uskomukset / Вераования карел*» (<https://vk.com/karjalazienuskomukset>), «*Карельская литература в цитатах*» (<https://vk.com/club840-23440>). Образовательными ресурсами служат сообщества «Учим карель-



ский язык» (https://vk.com/karjalan_kieli) и «*Karjalan kielilaskoile / Карельский для лентяев*» (<https://vk.com/karjalankieli>). Злободневные проблемы современной языковой ситуации обсуждаются на страницах сообществ «*Карельскому языку – официальный статус в Карелии*» (<https://vk.com/club79741361>), «*Onko karjala siun oma kiel / Носители карельского языка*» (<https://vk.com/club2562309>), «*Karjalan rahvahan liitto / Союз карельского народа*» (<https://vk.com/karjalanrahvahanliitto>).

За годы возрождения карельского языка выросло целое поколение молодой карельской интеллигенции, активно участвующей в процессах сохранения, популяризации и развития карельского языка. С сентября 2016 г. активисты карельского движения приступили к переводу интерфейса социальной сети «ВКонтакте» на карельский язык. Перевод осуществляется на собственно карельское наречие карельского языка, на котором говорят преимущественно в северных районах Карелии. В интервью Радио Карелии (запись от 01.02.2017) участница проекта, председатель правления Союза карельского народа Н. Воробей сообщала: «Современный мир диктует свои условия. Традиционная среда уходит, уходит лексика, связанная с ней. На смену приходит новая лексика; карельский язык – живой, он развивается. Молодёжь убеждается, что карельский – это не только язык бабушек и дедушек, он годится не только на рыбалке, но способен стать языком современного молодёжного контента» [7]. Координатор проекта М. Кундозерова рассказала о работе переводчиков: «Переводчики работают в специальном приложении сайта “ВКонтакте”», где по группам распределены ключи, подлежащие переводу. Ключ представляет собой отдельное слово, фразу, предложение, либо же небольшой текст. На данный момент над переводом трудится четыре человека. За 5 месяцев переведено 8 800 ключей, что составляет примерно треть от общего количества ключей. В процессе работы активистам приходится иметь дело с большим количеством современной лексики, связанной с компьютерами, программированием, языком разметки, программным обеспечением всевозможных гаджетов. Современные словари карельского языка подобной лексики практически не содержат, поэтому многие слова приходится создавать, заимствуя из других языков (финского, английского, русского). Данный процесс способствует развитию карельского языка, показывает его жизнеспособность, применимость во всех областях жизни. Появление карельской версии интерфейса популярной социальной сети позволит многим пользователям поупражняться в языке, освоить новую лексику, погрузиться в карельский мир “без отрыва от производства”, ведь именно в соцсетях многие из нас проводят большое количество времени. А чем чаще перед глазами карельский язык, тем быстрее он становится закономерной частью нашей жизни» [8]. За время



работы по переводу интерфейса участницы проекта составили словарь, включающий на сегодняшний день более 500 лексем. Интернет-терминология весьма не проста для перевода. Частично выручает финский язык, часть терминов калькируется, для обозначения некоторых употребляются английские заимствования, но немало и собственных элементов языка, которые используются в переводах: «Долго думали, например, как перевести глагол «подписаться (*follow*)» – *tarkkual'l'a* (*tarkkual'e*); «подписаться на автора» – *tarkkual'e kirjuttajua*; «подписаться (*subscribe*)» – *tilata* (*tilua*); «подписаться на SMS» – *tilua SMS-uutisie*. В финском языке употребляется в первом случае «*seurata*», а у нас такого слова нет, поэтому попробовали взять карельский глагол *tarkkual'l'a*» [9]. М. Кундозерова уверена в том, что «... карельский язык – это язык, на котором можно говорить везде и обо всём» [10].

Использование карельского языка в сети создаёт возможности не только для его сохранения, но и развития, совершенствования, создания новых терминологических пластов. Кроме того, социальная сеть выполняет и не менее важную экстралингвистическую функцию: она стимулирует интерес к языку, формирует мотивацию к его изучению, расширяет функциональное пространство.

* * *

1. Чикина Н.В. Современное состояние литературы на карельском языке. – Петрозаводск: КарНЦ РАН, 2011.
2. Национальный состав и владение языками, гражданство. Итоги Всероссийской переписи населения 2010 года. – Петрозаводск: Карелия-стат., 2013. – Т.5.
3. Родионова А.П., Нагурная С.В., Чикина Н.В. Людики: вопросы сохранения языка и культуры. – Петрозаводск: КарНЦ РАН, 2017.
4. Резолюция XII Конгресса МАФУН. – URL: <http://uralistica.com/profiles/blogs/rezolucija-12-kongressa-mafun>
5. Из интервью с Н. Гилоевой (интервью записано 13.02.2017).
6. Из интервью с Т. Барановой (интервью записано 13.02.2017).
7. Запись радиопередачи «Доброе утро, Карелия» от 01.02.2017 на канале «Радио России – Карелия 102,2 FM». – URL: <https://vk.com/radio-karjala>
8. Из интервью с М. Кундозеровой (интервью записано 13.02.2017).
9. Из интервью с М. Кундозеровой (интервью записано 13.02.2017).
10. Запись радиопередачи «Доброе утро, Карелия» от 01.02.2017 на канале «Радио России – Карелия 102,2 FM». – URL: <https://vk.com/radio-karjala>



Н.А. Нестерова,

*канд. ист. наук, доцент, Председатель
Комитета Государственного Совета
Республики Коми по законодательству
и местному самоуправлению
(г. Сыктывкар)*

ЛАБОРАТОРИЯ БУДУЩЕГО

Каждое событие нашей жизни имеет историю вопроса. Конференция «Электронная письменность народов Российской Федерации: опыт, проблемы, перспективы» не проходила бы сегодня в Республике Коми, если бы в своё время не родилась идея об объединении усилий людей из самых разных финно-угорских регионов России, думающих о будущем родных языков и делающих всё для его приближения. Информационные технологии – тот современный актуальный и востребованный инструмент, который позволит приблизить это будущее. Объединение информационных технологий и филологии создаёт именно те механизмы и инструменты, которые позволяют включить родные языки, языки малочисленных народов в современное информационное пространство наравне с самыми развитыми языками мира.

Первые объединённые проекты были созданы в рамках деятельности Финно-угорского центра при международном отделе Сыктывкарского государственного университета в 2008–2009 годах. Создание Центра инициировала, а затем возглавила его канд. филол. наук М.С. Федина. Её стратегической идеей уже тогда было включение коми языка в глобальное цифровое информационное пространство. Эта идея родилась из сотрудничества с коллегами – Трондом Тростерундом (Норвегия), А. Чемышевым (Марий Эл), В. Степановым (Пермский край), которые уже тогда считались авторитетными специалистами в профессиональном сообществе. Однако их деятельность не находила поддержки, они опережали своё время, – всё, что они делали, было делом будущего.

Лишь в 2011 г. А. Чемышев и В. Степанов были приглашены на работу в республику и при Коми республиканской академии государственной службы и управления (далее – Академия) с декабря 2011 г. начала работу Межрегиональная лаборатория информационной поддержки функционирования финно-угорских языков (далее – Лаборатория). При объединении



Лаборатории с созданным чуть позже, в феврале 2012 г., Бюро официального перевода (далее – Бюро), был создан **Центр инновационных языковых технологий**. С этого времени вся деятельность Центра была направлена на обеспечение функционирования языка коренного народа республики – коми языка – в современных условиях и реализацию закона Республики Коми «О государственных языках Республики Коми». В то же время параллельно формировались продукты для марийского и коми-пермяцкого языков.

Создание первых информационных продуктов сопровождалось активным продвижением деятельности Лаборатории, разъяснением её значимости, актуальности и перспектив, востребованности в будущем. Сторонников было немного, правительство не торопилось с финансированием, поэтому особое значение имела поддержка деятельности лаборатории в это непростое время со стороны коллектива Академии – она полностью финансировалась за счёт средств Академии.

Первым продуктом Лаборатории стала единая коми раскладка компьютерной клавиатуры в формате ЮНИКОД, которая сразу же доказала свою значимость при проведении единого дня голосования на выборах – при формировании бюллетеней для голосования на государственных языках. Впоследствии формат ЮНИКОД коми раскладки клавиатуры был утверждён Распоряжением Главы Республики Коми.

Со временем **Межрегиональная лаборатория информационной поддержки функционирования финно-угорских языков** (<http://fu-lab.ru>) стала настоящей межрегиональной и международной площадкой, единственным в России системным механизмом для создания информационных продуктов, ресурсов, инструментария, которые способствуют расширению функционирования языков, сохранению и пропаганде этнокультурного наследия коми и других финно-угорских народов, развитию международных коммуникаций, программ дистанционного образования, обеспечению доступа к информационным ресурсам этнокультурной направленности.

Впервые государственное финансирование Лаборатория получила лишь в 2014 году.

Деятельность Лаборатории строится на системной основе, включающей проведение научно-исследовательских работ и на их основе – разработку и апробацию электронных продуктов, обучение использованию этих продуктов и др.

С начала работы Лаборатории и по сегодняшний день были реализованы следующие проекты.

Разработка и апробация электронных продуктов:

1. Электронные онлайн- и офлайн-коми-русские и другие словари представлены на <http://dict.fu-lab.ru> и <http://fu-lab.ru/elslovari>



2. Создана компьютерная терминология на коми языке. Словарь «Базовая терминология для коми локализации компьютерных программ» размещён на <http://komikyv.ru/blog/dictionary/375.html>

3. Разработаны и апробированы онлайн- и оффлайн-системы проверки правописания коми и других языков (коми спеллер для *Mozilla*, проверка правописания коми и удмуртского языков в *Mac OS*).

4. Адаптирована для коми текстов программа корпус-менеджера *AntConc*, которая позволяет автоматически анализировать коми тексты.

5. Создан онлайн-конвертер с нестандартной кодировки, конвертер с нестандартной кодировки для ОС *Windows* на <http://fu-lab.ru/converter>

6. Создана и систематизирована полная аннотированная база изданий (книг и журналов) на коми языке с 1920 г. по настоящее время, размещённая на http://wiki.fu-lab.ru/index.php/Материалы_по_коми_языку

7. Создана автоматическая система проверки правописания коми языка: онлайн на <http://webspell.fu-lab.ru> и оффлайн на <http://fu-lab.ru/pravopisanie>

8. Создана автоматическая система переноса слов коми языка со строки на http://wiki.fu-lab.ru/index.php/Автоматическая_расстановка_переносов

Особо выдающимся результатом деятельности Лаборатории стал проект создания Национального корпуса коми языка (<http://komicorpora.ru>).

Важнейшим направлением работы Лаборатории на данном и последующем этапах является **методическое сопровождение национального педагогического образования**, в том числе организация и проведение обучающихся семинаров, занятий для пользователей, использующих коми и другие финно-угорские языки в Интернете и компьютерной технике и т.д. По данному направлению, кроме различных обучающих семинаров, Лабораторией создан специальный раздел сайта «Коми школа». На площадке Лаборатории успешно реализована первая часть международного проекта по созданию электронных учебных пособий на родном языке по предметам школьного курса («Логика», «Коми литература», «Окружающий мир» и «История») как дополнение к имеющимся учебникам.

Сегодня Лаборатория стала общепризнанным международным авторитетом, её деятельность отмечалась и была известна практически всем участникам VII Всемирного конгресса финно-угорских народов, проходившего в июне 2016 г. в г. Лахти (Финляндия).

Сегодня к специалистам Лаборатории за консультациями и поддержкой обращаются специалисты из разных регионов России – Республик Удмуртии, Марий Эл, Башкортостана, Тывы, Алтайского края и др. В сотрудничестве созданы электронные словари для марийского, удмуртского язы-



ков; орфосервисы (онлайн-система проверки правописания) для удмуртского, марийского, чувашского языков; раскладки клавиатуры для удмуртского, марийского, алтайского, тувинского языков; параллельно с коми языком создаются продукты для коми-пермяцкого языка. В последнее время в Лабораторию обращаются представители регионов РФ с просьбой о проведении тематических обучающих семинаров по созданию определённых электронных продуктов.

Сотрудники Лаборатории активно участвуют в международных проектах и мероприятиях, сотрудничают с зарубежными партнёрами (центр саамских языковых технологий университета г. Тромсе (Норвегия), университет г. Фрайбург (Германия), университет г. Уппсала (Швеция) и т.д.). В резолюции VII Всемирного Конгресса финно-угорских народов рекомендовано создание Международной ассоциации по поддержке электронной письменности уральских языков. В настоящее время функционирует Специальная группа по компьютерной лингвистике для уральских языков в рамках Международной ассоциации компьютерной лингвистики, членом которой является и Лаборатория.

У Лаборатории на сегодняшний день много планов: дальнейшая работа по пополнению базы Национального корпуса коми языка; продолжение работы над электронной библиотекой на коми языке и коми-пермяцком языке (<http://komikyv.org>); работа по расширению онлайн-медиотеки коми устного корпуса диалектных текстов (коми медиатека); создание русско-коми переводчика, параллельного русско-коми корпуса текстов, электронного коми-русского разговорника с аудиосопровождением, словаря коми официальной терминологии, аудиоверсии коми-русского электронного словаря и многое другое.

Сегодня мечта Лаборатории и её руководителя – создание федеральной сети учреждений и специалистов по поддержке и развитию электронной письменности уральских языков, официальному переводу с русского на государственные финно-угорские языки, и Лаборатория здесь готова выступить координатором. Возможно, на данной конференции начнётся воплощение этой мечты в жизнь.



Н.М. Новикова,
руководитель АУ РК
«Редакция журнала «Арт»»

О РЕАЛИЗАЦИИ ПРОЕКТА АУДИОБИБЛИОТЕКИ НА КОМИ ЯЗЫКЕ «ЛОВЬЯ КЫВ» (Живоекомислово.рф)

Развитие коми языка и литературы во многом зависит от самих людей, их заинтересованности говорить, читать и писать по-коми.

Одну из главных ролей в продвижении слова играет Книга. Особенно она важна для детской аудитории, т.к. чтение помогает познать окружающий мир, воспитывает дух подрастающего поколения, рассказывает, что такое «хорошо», а что такое «плохо». Однако изданные в Республике Коми на сегодняшний день книги по объективным причинам не могут полностью представить «золотой фонд» детской и юношеской литературы на коми языке, который был накоплен писателями и культурологами.

Большую ценность в этой ситуации приобретают электронные информационные ресурсы, предоставляющие возможность широкой аудитории познакомиться с литературными произведениями на коми языке. Портал «Ловья кыв» (Живое слово) – один из них.

Портал с аудиозаписями произведений для детей на коми языке «Ловья кыв» был запущен в декабре 2015 года.

Мысль о создании портала с аудиокнигами на родном языке возникла на очередном празднике «Коми книга», который состоялся в июне 2015 г. в Удорском районе. Там же было решено реализовать проект «Ловья кыв» совместно с редакцией журнала «Арт».

В качестве эксперта проекта выступила З. Остапова – канд. пед. наук, зав. лабораторией национальных проблем дошкольного образования Коми республиканского института развития образования (КРИРО).

Позже к проекту подключились специалисты Коми народного радио, которые за последние несколько лет озвучили большое количество произведений на коми языке.

Реализация идеи по созданию портала стала возможна благодаря финансовой поддержке Агентства Республики Коми по печати и массовым коммуникациям. Проект «Ловья кыв» в середине 2015 г. выиграл кон-



курс на выделение субсидии (грантов) для реализации социально-значимых проектов в сфере СМИ.

Сегодня «Ловья кыв» содержит более 400 аудиокниг на коми языке для детей дошкольного возраста. Помимо этого, здесь представлена информация о культурных событиях, происходящих в Республике Коми. Также на сайте открыт раздел «Вместо слов», где планируется размещать видеоматериалы коми спектаклей, телепередач и выставок известных художников.

Портал «Ловья кыв» является уникальным информационным ресурсом, пропагандирующим коми язык и литературу. Сегодня любой желающий может ознакомиться с золотым фондом произведений на коми языке для детей дошкольного возраста. Однако «Ловья кыв» охватывает только определённую часть коми литературы. На портале не представлены произведения для учащихся начальных и средних учебных заведений, произведения современных авторов, рассказы и повести для взрослых.

Это означает, что портал лишь частично решает проблему доступности для населения литературных произведений на коми языке. Юношеская проза и поэзия, художественная литература на коми языке до сих пор не представлены массовой аудитории в полном объёме. Технические возможности портала «Ловья кыв» позволяют заполнять его новой информацией, в том числе и другими, даже большими по объёму, аудиокнигами. Именно этим обусловлена необходимость развития проекта в 2017 году.

Цель развития портала «Ловья кыв – 2017» – популяризация разных жанров художественной литературы и эпоса на коми языке, оказание методической помощи начальным и средним образовательным учреждениям в расширении репертуара детского чтения, а также распространение детского и юношеского художественного материала среди молодых родителей, заботящихся о воспитании своих детей на родном языке.

В результате реализации проекта «Ловья кыв – 2017» будет решён ряд важнейших сопутствующих задач:

- литературные произведения на коми языке станут доступны для людей с ограниченными возможностями здоровья по зрению (слепые и слабовидящие);
- пожилые плохо видящие люди смогут слушать коми произведения вместо затруднительного чтения;
- учащиеся начальных и средних учебных заведений смогут без труда находить литературу на коми языке, рекомендованную к изучению Министерством образования, науки и молодёжной политики Республики Коми.



Н.В. Остапов,

*канд. ист. наук, преподаватель
истории ГПОУ «Сыктывкарский
колледж сервиса и связи»*

УЧЕБНЫЙ МУЛЬТИМЕДИЙНЫЙ ПРОЕКТ «АСЛАМ ЧУЖАНИНӨ ТУЙ» / «ДОРОГА ДОМОЙ»: ОТ ПРОЕКТИРОВАНИЯ ДО РЕАЛИЗАЦИИ

Современные студенты, как правило, стараются избегать мировоззренческой рефлексии, подменяя размышления над «вечными» вопросами житейским прагматизмом и отказываясь от традиционного выбора между ценностными позициями в пользу потребления «здесь и сейчас». Учебные дисциплины социально-гуманитарного цикла в системе среднего профессионального образования, на наш взгляд, не могут в полной мере помочь обучающемуся выработать смысловые и жизненные ориентиры, однако способны подтолкнуть студента к определению и аргументации некоторых ориентиров собственной жизни и своей будущей профессиональной деятельности.

Практика преподавания социально-гуманитарных дисциплин показывает, что преподаватели испытывают трудности с формированием устойчивого интереса к дисциплинам из-за нестыковки между предлагаемыми для размышлений проблемными полями, с одной стороны, и ценностно-смысловыми доминантами юного субъекта общественных отношений, с другой. Безусловно, большое значение в преподавании приобретает погружение на учебных занятиях в ситуации неопределённости с целью стимулирования у студентов творческой активности и поисков нестандартных решений на основе личного опыта. Это могли бы быть проблемы разного масштаба:

1) ситуативного типа, связанные с личным выбором человека; масштаб глобальный (например, выбор Сократа между недостойной жизнью и достойной смертью, с которым мог бы столкнуться в жизни каждый и решить для себя независимо от реальных обстоятельств выбора древнегреческого философа в конкретно-исторических условиях);

2) ситуативного типа; масштаб локальный, ограниченный конкретно-историческими обстоятельствами (например, я больше симпатизирую взглядам западников, потому что в России XIX века...);



3) ситуативного типа; масштаб отдельной личности при решении конкретной задачи в актуальной ситуации (например, покупка в магазине телефона, не соответствующего заявленным характеристикам, и последовательные шаги выхода из сложившейся ситуации).

Проблемы масштаба отдельной личности, как правило, затрагивают глубинные мотивы, чувства, смыслы личности и выходят за пределы учебных занятий. Они настолько индивидуальны, что могут основываться преимущественно на опыте личного переживания ситуации, способны реализоваться в неожиданных итоговых формах и не всегда могут быть однозначно оценены. Одной из подобных работ в Сыктывкарском колледже сервиса и связи стал студенческий мультимедийный проект «Аслам чужанінö туй» / «Дорога домой» по произведению коми поэта А.Е. Ванеева [1]. Этапы реализации проекта представлены в виде *таблицы*.

Т а б л и ц а

Этапы реализации проекта

Название этапа	Задачи этапа	Принципы деятельности	Виды деятельности
I. Этап мотивации	Условия: многие студенты колледжа приехали из сёл и испытывают ностальгию по родным местам, своей семье, часть из них скучает по родному языку. Задача: пробудить желание выразить свои чувства и эмоции, которые появляются в момент возвращения домой и встречи с родными и близкими	Приоритетность мотивации проектной деятельности	Выявление студентов, способных и желающих заняться подобной деятельностью. Критерии отбора: 1. Тяга к творчеству (рисование, чтение стихов и т.д.). 2. Знание коми языка. 3. Уверенный пользователь компьютера
II. Этап выбора формата проекта	1. Составить перечень художественных произведений, которые можно было бы использовать в качестве посылки к творчеству. 2. Определить формат конечного продукта проекта «Дорога домой». 3. Очертить этапы реализации проекта	Свобода выбора; новизна	1. Знакомство со сборником стихов Альберта Ванеева «Ловья би» и выбор одного произведения для чтения. 2. Определение замысла и общей композиции рисунков. 3. Знакомство с оборудованием и программами, позволяющими работать со звуком и изображением



Продолжение таблицы

III. Этап реализации и оформления проекта	1. Нарисовать художественные композиции и перевести их в цифровой вариант (всего было сделано 5 рисунков). 2. Записать чтение автором проекта стихотворения Альберта Ванеева «Ме локта гортõ». 3. Подобрать музыкальный фон для видеоролика. 4. Соединить изображения и звуки в видеоролик с помощью программы Macromedia <i>Flash MX8</i> и опубликовать в формате <i>swf</i>	Отсутствие временных границ; кооперация; индивидуализация	1. Рисование. 2. Перевод рисунков в формат <i>jpg</i> . 3. Чтение стихов и их запись в формате <i>mp3</i> . 4. Компоновка всего материала в видеоролике формата <i>swf</i> . 5. Редактирование материалов в видеоклипе
IV. Этап подведения итогов	Представить работу перед аудиторией	Практическая значимость	Демонстрация проектной работы на конкурсах, конференциях разного уровня

При сегодняшней загруженности студента и преподавателя краткосрочные проекты являются важным условием самореализации личности. В частности, уникальный опыт подобной проектной деятельности накоплен научно-образовательным центром «Инновации в национальном гуманитарном образовании» Сыктывкарского государственного университета имени Питирима Сорокина под руководством Е.В. Остаповой. Результаты работы в данном направлении опубликованы в ряде статей [2]. При реализации такого рода проектов решаются многие актуальные задачи современного образования: студент становится действующим субъектом, активным участником образовательного процесса; повышается степень сотрудничества между преподавателем и обучающимся; создаются условия для реализации творческого потенциала с использованием современных информационных технологий.

* * *

1. Ванеев А.Е. Ловья би: кывбурьяс да поэмаяс. – Сыктывкар: Коми книжное изд-во, 1983.



2. *Остапова Е.В.* Мультимедийные литературно-образовательные ресурсы на коми языке // Родной язык в современных условиях двуязычия: сб. статей / под общ. ред. Е.В. Остаповой, С.Н. Терентьевой. – Сыктывкар: Кола, 2014. – С.267–274.



Е.В. Остапова,

*канд. филол. наук, доцент кафедры финно-
угорской филологии и национального
образования Сыктывкарского
государственного университета
имени Питирима Сорокина*

О ВНЕДРЕНИИ В ОБРАЗОВАТЕЛЬНЫЙ ПРОЦЕСС ЭЛЕКТРОННЫХ РЕСУРСОВ ПО КОМИ ЛИТЕРАТУРЕ

В данной статье мы осветим вопросы внедрения в литературно-образовательный процесс созданных нами электронных ресурсов на коми языке. В образовательном пространстве Республики Коми активно используются созданные на базе научно-образовательного центра «Инновации в национальном гуманитарном образовании» Сыктывкарского государственного университета имени Питирима Сорокина мультимедийные учебные пособия по коми литературе на коми языке с грифом Министерства образования, науки и молодёжной политики Республики Коми, а также ряд других интерактивных образовательных ресурсов. По виду деятельности это интерактивные тестирующие задания и аудиохрестоматия («Небыд, мича, гора» / «Мягкий, красивый, звучный» (2007)), демонстрационная программа («Шудторйой, кодзулйой» / «Счастье моё, звёздочка» (2009)), аудиохрестоматия для 5 класса («Мый меддонайс» / «Что дороже всего» (2009)), мультимедийный альманах на пермских языках «Выль чужом» / «Новое лицо» (2011)), интерактивная игра «Чуня-чалля» с прослушиванием текста, собиранием пазлов и раскраской для младших школьников (2012). Данные мультимедийные издания на коми языке для учащихся средних общеобразовательных школ успешно внедряются в школьную практику, используются в тесном взаимодействии с функционирующими учебниками коми языка и литературы. Авторы-разработчики проектов в течение ряда лет выступают с презентациями изданий на конференциях регионального, всероссийского и международного уровней, принимают участие в теле- и радиопередачах, семинарах и курсах повышения квалификации учителей в Сыктывкарском государственном университете имени Питирима Сорокина, Коми республиканском институте развития образования и переподготовки кадров (КРИОиПК), проводят индивидуальные консультации по применению этих и других учебных проектов в образовательном процессе.



«Знание и понимание насущных потребностей модернизации национального гуманитарного образования, интеграция филологии, педагогики, информатики с такими видами искусства, как художественная литература, музыка, изобразительное искусство, позволяют оценивать подобного рода продукцию как варианты инновационной модели учебных пособий по коми литературе, способствуют модернизации образовательной среды, развивают перспективное направление в национальном гуманитарном образовании» [1].

По нашим наблюдениям, из вышеперечисленных ресурсов наиболее используемы учителями демонстрационные пособия и аудиоматериалы, т.к. они более адаптированы к традиционным методам преподавания. Так, российской школой преподавания литературы накоплен большой опыт работы с фонохрестоматией – предшественницей аудиохрестоматии. Дикторская речь на пластинках и сейчас вспоминается сформировавшимся в советское время поколением учителей как правильная, ясная, образцовая. В то же время в современных аудиоматериалах мы слышим авторское, субъективное прочтение художественного текста, исполненное артистами Национального музыкально-драматического театра Республики Коми, студентами Сыктывкарского государственного университета имени Питири-ма Сорокина, самими авторами произведений. Также в некоторых школах своими силами создаются ученические аудиокниги с использованием различных гаджетов. Демонстрационный метод с использованием наглядного материала был одним из основных в преподавании почти всех дисциплин, поэтому мультимедийное учебное пособие «Шудторйбй, кодзулбй», представляющее собой своеобразное слайд-шоу о двенадцати коми писателях и поэтах, эффективно используется на различных этапах уроков и в самостоятельной работе учащихся. При этом слайдовый материал используется в разном направлении: в оформлении стендов о творчестве коми писателей и поэтов, для составления тестовых заданий, конспектирования, как дополнительный материал к учебнику для подготовки докладов, выполнения творческих заданий, создания нового мультимедийного ресурса.

С известной долей осторожности на уроках коми литературы выполняются интерактивные задания, представленные в мультимедийных пособиях в виде тестовых заданий теоретического характера с включением музыкально-художественных материалов. По-видимому, это объясняется определённым неприятием гуманитариями в целом тестовой формы обучения как инструмента, предназначенного для измерения обученности учащегося, имеющего свои преимущества и недостатки. Составление и выполнение тестовых заданий требуют перестройки мышления и освоения нового приёма деятельности проверочно-закрепительного и оценочного характера, иного взгляда на изучаемый текст.



В связи с вышесказанным подчеркнём, что экранное чтение как один из приёмов формирования компетентностного читателя действительно требует особой подготовки прежде всего от учителя, особого настроя всей аудитории. В 1997 г. автором многочисленных научных трудов и книг по юзабилити интернет-сайтов Якобом Нильсеном было проведено исследование, по итогам которого выяснилось, что «визуальному восприятию информации с экрана компьютера присуща своя схема. Зона на мониторе, которую “выхватывает” глаз человека, очень напоминает по форме латинскую букву “F”. Причём геометрия этой зоны практически не меняется даже при смене шаблона представленной информации» [2]. Сканирование читателем всего представленного информационного наполнения происходит на уровне рефлексии, затем на уровне анализа объектов на основе трёх принципов: принципа пяти строк, принципа притягивания или подталкивания взгляда, принципа управления восприятием. Необходимо учитывать опыт современных исследований проблем чтения и восприятия изучаемого текста и при создании электронных образовательных ресурсов, особенно, подчеркнём, на коми языке, т.к. мы наблюдаем ослабление навыков всех видов чтения на родном языке. В связи с этим напомним, что почти сразу же после принятия в 1992 г. закона о государственных языках появилась тенденция массового перехода средних общеобразовательных школ Республики Коми к обучению по программе «Коми язык как государственный», подразумевающей изучение коми языка с опорой на методику изучения иностранных языков и исключаяющей изучение родной литературы. Таким образом, на сегодняшний день подавляющее большинство учеников средних общеобразовательных школ, разговаривающих на родном коми языке, лишены возможности изучения письменной культуры коми, коми художественной литературы, насчитывающей несколько веков развития. В настоящее время данная тенденция обрела катастрофический размах: вырастает поколение детей, пока ещё владеющих родной коми речью и в лучшем случае лишь слышавших имена коми писателей и поэтов, литературных героев, детей, не читающих художественных литературных произведений на родном языке.

Итак, можно констатировать, что рассматриваемые мультимедийные учебные пособия нашли своё достойное место в этнокультурном образовательном пространстве. Однако дальнейшие электронные разработки и включение их в учебный процесс в настоящее время требуют более пристального внимания и профессионального подхода. Создание электронного образовательного ресурса (далее – ЭОР), издание его на диске или размещение на сайте ещё не даёт гарантии, что данный продукт будет должным образом использоваться в учреждениях образования и культуры. Так, в 2015 г. в рамках проекта «Создание электронных учебных пособий по предметам



школьного курса на родном языке», инициируемого венгерским Институтом языковой стратегии, подготовлены мультимедийные учебные пособия на коми языке по ряду школьных предметов, в том числе интерактивное пособие по коми литературе для 5–6 классов «Гулюсянь юёр». Данные ресурсы помещены на сайте <http://fu-lab.ru>. Авторами новых пособий было проведено несколько выступлений на международных и всероссийских конференциях, опубликованы статьи, в частности, в сборнике «Управленческие аспекты развития северных территорий России» [3]. Но, как мы видим, этого на сегодняшний день явно недостаточно. Необходимо работать адресно, и не только с целью ознакомления с новым продуктом, но и выходить в учительскую аудиторию с методическими рекомендациями.

На сегодняшний день можно выделить 3 основные проблемы, на наш взгляд, тормозящие успешное внедрение в образовательные учреждения информационно-коммуникационных ресурсов на коми языке, в частности, по коми литературе: техническая, методическая, психологическая.

Отсутствие или неравномерное аудиторное распределение соответствующего качественного оборудования в некоторых средних общеобразовательных, профессиональных учреждениях и высших учебных заведениях сводит к минимуму саму возможность использования ЭОР. Зачастую работа с электронными ресурсами носит фрагментарный характер и вряд ли может привести к развитию продуктивной творческой деятельности, формированию необходимых компетенций у обучающихся. У учителя, преподавателя, обучающегося может выработаться мнение о необязательности, факультативности требования использования ЭОР на родном языке в образовательном процессе. Имеются нормативные документы, регулирующие условия организации современного образовательного процесса. Исходя из требований, наиболее правильным является такой урок, на котором равномерно чередуются различные формы работы: работа с тетрадью и учебником чередуется с работой за компьютером, работой на интерактивной доске.

Не секрет, что наиболее уязвимым местом в этнокультурной образовательной среде многие годы является учебно-методический комплекс (далее – УМК) предметов этнокультурного цикла, в данном случае – в преподавании коми языка (родного) и коми литературы. УМК по литературе включает: рабочие программы, учебник, фонохрестоматию (аудиохрестоматию) с методическими рекомендациями, рабочие тетради, хрестоматии художественных произведений, дидактические материалы, поурочные разработки. К сожалению, завершённой линии УМК по коми литературе нет ни к одному из классов. В настоящее время принята новая программа по коми литературе для общеобразовательных школ в соответствии с Федеральным государственным образовательным стандартом, начат выпуск



новой линии учебников и хрестоматий, содержащих поисковые задания с использованием Интернета, мультимедийных технологий. Налицо большие изменения и в образовательной среде по отношению к ЭОР. Так, при разработке первых инновационных электронных продуктов мы сами учились новому, иногда не чётко представляя итог. На учебных курсах большое внимание уделялось знакомству как с работой на компьютере, так и с принципиально новым литературным учебным ресурсом на коми языке. В настоящее время, думается, мы можем говорить об освоении учителями и учащимися нового образовательного информационного пространства, хотя и в разной степени. Поэтому с выходом нового поколения учебников коми языка и литературы необходимо разрабатывать ЭОР с включением методических рекомендаций, как того требует образовательный процесс. В настоящее время готовятся методические рекомендации по работе с мультимедийным интерактивным учебным пособием «Гулюсянь юёр», которое состоит из 4 разделов и представляет собой синтез отсутствовавших до настоящего времени элементов учебно-методического комплекса для 6 класса, таких как рабочая тетрадь и аудиохрестоматия. В качестве упражнений учащимся предложены следующие задания: решить кроссворды по актуальным темам коми фольклора и литературы; собрать пазлы из фрагментов обложек книг, рекомендуемых для чтения, к каждой из которых имеется аннотация, адаптированная к детскому чтению; отдельный блок посвящён заданию выбрать из нескольких вариантов пропущенные выражения в фольклорных текстах (песнях-сказках, пословицах, поговорках); во фрагменте аудиохрестоматии представлено чтение стихотворений В.Т. Чисталёва и В.Г. Лодыгина выпускниками кафедры финно-угорской филологии и национального образования Сыктывкарский государственный университет имени П.А. Сорокина.

Предполагаем, что в процессе решения вышерассмотренных проблем будет меняться к лучшему психологический климат в педагогическом коллективе. Не секрет, что преподаватели и учителя родного языка и литературы испытывают большой дискомфорт, возникший вследствие как интенсификации изменений в образовательном процессе, так и процесса сужения самого образовательного пространства на коми языке. Проблемы преподавания родного языка и литературы требуют самого пристального внимания, грамотного подхода к внедрению современных образовательных технологий, в том числе передовых информационных технологий.

Опыт создания ЭОР и высокая востребованность подобной продукции позволяют ставить вопрос об исследовательском проекте по изучению процесса внедрения компьютерных технологий в этнокультурное образование. Краткость, точность и уместность использования электронных об-



разовательных ресурсов и их возможностей позволяют повысить качество обучения, донести до ученика нужное знание, простимулировать самостоятельную творческую работу, закрепить пройденный материал, создать новый ресурс. При этом использование ЭОР в образовательном процессе не должно быть целью преподавателя. Но, как и традиционные методы, информационные технологии, ЭОР являются средствами, с помощью которых педагог формирует картину мира у подрастающего поколения.

* * *

1. *Остапова Е.В.* Мультимедийные литературно-образовательные ресурсы на коми языке // Родной язык в современных условиях двуязычия: сб. статей / под общ. ред. Е.В. Остаповой, С.Н. Терентьевой. – Сыктывкар: Кола, 2014. – С.273.

2. См.: *Александрова Е.В.* Экранное чтение как один из приёмов формирования компетентностного читателя // Литература в школе. – 2013. – № 10. – С.23–25.

3. *Остапова Е.В.* Об интерактивном мультимедийном учебном пособии по коми литературе «Гулюсянь юёр» (Весть от голубя) // Управленческие аспекты развития северных территорий России: материалы Всероссийской науч. конф. (20–23 октября 2015 г., Сыктывкар): в 4 ч. – Сыктывкар: ГОУ ВО КРАГСиУ, 2015. – Ч.4. – С.31–34.



УДК 371.315:004

З.В. Остапова,

*канд. пед. наук, зав. лабораторией
национальных проблем дошкольного
образования ГОУ ДПО «Коми
республиканский институт
развития образования»
(г. Сыктывкар)*

Л.Г. Кулышева,

*методист лаборатории национальных
проблем дошкольного образования
ГОУ ДПО «Коми республиканский
институт развития образования»
(г. Сыктывкар)*

М.В. Рудецкая,

*методист лаборатории национальных
проблем дошкольного образования
ГОУ ДПО «Коми республиканский
институт развития образования»
(г. Сыктывкар)*

ЭЛЕКТРОННЫЕ ОБРАЗОВАТЕЛЬНЫЕ РЕСУРСЫ ЭТНОКУЛЬТУРНОЙ НАПРАВЛЕННОСТИ В ДОШКОЛЬНОМ ОБРАЗОВАНИИ РЕСПУБЛИКИ КОМИ

В системе дошкольного образования в настоящее время происходят значительные перемены, которые связаны с введением дошкольного образования в систему общего образования в соответствии с Законом «Об образовании в Российской Федерации» [1]. Важным условием модернизации дошкольного образования является использование новых информационных технологий, в первую очередь компьютеров. В соответствии с Концепцией внедрения новых информационных технологий в дошкольное образование компьютер должен стать в детском саду ядром развивающей образовательной среды. Он рассматривается не как отдельное обучающее игровое устройство, а как информационная система, способная применяться в различных направлениях образовательного процесса и изменить развивающую среду детского сада в целом [2]. Именно поэтому можно утверждать, что проблема компьютеризации образования, в частности дошкольного, является одной из актуальных проблем модернизации российского образования.



Компьютеризация образования в нашей стране уже имеет свою небольшую историю. Однако применение компьютеров в системе дидактических средств дошкольного образования только начинается. Необходимо отметить, что использование компьютера в детском саду предусматривает не обучение детей школьным основам информатики и вычислительной техники, а преобразование предметно-развивающей среды ребёнка, создание новых, научно обоснованных средств его развития. Благодаря внедрению компьютерной техники в дошкольные учреждения у педагогов появилась возможность использовать в педагогическом процессе электронные образовательные ресурсы (далее – ЭОР) с целью повышения качества образовательного процесса. ЭОР используют в дошкольном образовании с целью обновления форм и методов работы с детьми. Установлено, что при соответствующем подходе многие направления, задачи и содержание образовательного процесса с детьми могут быть обеспечены развивающими компьютерными программами [3]. Но нужно учитывать, что в детском саду могут использоваться только такие компьютерные программы и игры, которые соответствуют психолого-педагогическим и санитарно-гигиеническим требованиям [4].

В Республике Коми особую актуальность имеют ЭОР этнокультурной направленности. Так, согласно Концепции развития этнокультурного образования в Республике Коми на 2016–2020 гг. одной из важных задач является создание системы информационно-методических условий развития этнокультурного образования, включающей в себя [5]:

- возможность широкого использования информационных технологий, технологий дистанционного (мобильного) характера и обеспечения образовательного процесса соответствующими *IT*-технологиями;
- комплектность обеспечения образовательного процесса учебниками, учебными пособиями, учебно-методической, дополнительной литературой, ЭОР с учётом специфики этнокультурного содержания основных образовательных программ;
- создание единого республиканского электронного этнокультурного образовательного пространства.

Лаборатория национальных проблем дошкольного образования (далее – Лаборатория) на протяжении многих лет создаёт электронные издания разнообразных видов. Детским садам, где нет музыкального руководителя, большую помощь оказала **«Фонохрестоматия “Музыка в коми детском саду”»** (набор из 7 дисков; составители Р.С. Малыгина, С.В. Головина; Сыктывкар, 2001). Данная фонохрестоматия содержит музыкальные произведения на коми языке для слушания, пения, танцев. Особенно актуальной она будет для тех детских садов, где нет музыкального руково-



дителя или музыкального инструмента. Фонохрестоматия включает в себя также диски **«Финно-угорская музыка для детей»** и **«Коми народные музыкальные инструменты»**.

Также в Лаборатории подготовлена фонохрестоматия «Художественная литература на коми языке для дошкольников». Из фонда этих аудиоматериалов была создана аудиокнига **«Олісны-вылісны...»** («Жили-были») (составитель З.В. Остапова). В содержание диска включены 14 коми народных и авторских сказок, составляющих «золотой фонд» коми детской литературы. Сказки озвучены в форме литературных композиций и радиоспектаклей мастерами художественного слова, артистами.

В Республике Коми с 2015 г. запущен новый проект сайта с произведениями на коми языке «Ловъя кыв» (<http://живоекомислово.рф>). Наполнение сайта было начато с произведений детской литературы. На сегодняшний день портал содержит более 200 аудиопроизведений на коми языке. Из фондов Лаборатории портала «Ловъя кыв» предоставлена коми детская литература.

В 2006 г. была создана компьютерная игра «Кикуруллюкӧд ворсӧмъяс» («Игры с Кикуруллю») (авторы З.В. Остапова, И.А. Пасынкова). Она стала абсолютно новым видом учебной продукции по дошкольному и начальному образованию и заняла достойное место среди педагогических технологий по этнокультурному образованию детей. Игра направлена на воспитание интереса к традиционной культуре коми народа и активизации коми языка как средства общения и игровой деятельности. Игра состоит из 6 частей – мини-игр: «Отгадай загадку», «Найди пару», «Давай споём», «Отгадай, на чём играют», «Отгадай сказку», «Одень куклу». Игры построены по методике опережающего обучения, что позволяет ребёнку свободно переходить от одной игры к другой, при желании не проигрывая её до конца, и постепенно учиться всему и понемногу. В ходе игры текст звучит на коми языке, а письменные указания, в целях расширения аудитории играющих, даются на русском языке. Игра предназначена для детей от 5 до 10 лет и может быть использована в образовательных учреждениях, учреждениях культуры, а также в семье.

В Республике Коми есть малокомплектные детские сады, где дети воспитываются в разновозрастных группах. Для обеспечения более качественного дошкольного образования детей старшего дошкольного возраста созданы аудиопособия для реализации разных образовательных областей. В то время, когда педагог занимается с младшими детьми, старшие дети могут самостоятельно прослушать занятие, а задания выполнить вместе с педагогом. К каждому аудиопособию имеется текстовое приложение со сценариями этих занятий:



1. Аудиопособие **«Тõдмõдам дзюлюкъясõс музыкаõн»** («Знакомим детей с музыкой») (составитель Н.Н. Морозова). В аудиопособии представлены музыкальные занятия для детей дошкольного возраста. В ходе занятий дети знакомятся с известными коми и русскими композиторами, их произведениями, а также с различными видами музыкальных инструментов. Данное аудиопособие может быть использовано во всех формах организации музыкальной деятельности: на музыкальных занятиях, в повседневной жизни детского сада, а также в семье.

2. Аудиопособие **«Чужан му историяõн да коми йõзлõн оласногõн тõдмõдõм»** («Знакомство с историей и традициями коми народа») (составитель В.А. Муравьёва). В аудиопособии представлены занятия для детей дошкольного возраста по ознакомлению с историей и бытом коми народа для дошкольников. Тематика занятий включает в себя ознакомление с традиционным коми народным костюмом, коми игрушкой, историей уюга, культурой гостеприимства и т.д.

3. Аудиопособие **«Коми йõзкостса календарõн тõдмõдõм»** («Знакомство с коми народным календарём») (составитель В.А. Муравьёва). В аудиопособии представлены занятия по ознакомлению с коми народным календарём для детей дошкольного возраста.

4. Аудиопособие **«Тõдмõдам дзюлюкъясõс пемõсьяскõд»** («Знакомим детей с животными») (автор-составитель В.А. Муравьёва). Представлены аудиозанятия на коми языке: «Шыр йылысь» (Про мышку), «Гут-гаг йылысь» (Насекомые), «Лягуша йылысь» (Лягушка), «Черияс йылысь» (Рыбы), «Чипан да петук йылысь» (Курица и петух), «Кыдзи тõвйõны вõрса пемõсьяс» (Как зимуют лесные звери), «Ёж йылысь» (Про ёжика), «Гортса пемõсьяс йылысь» (Домашние животные).

5. Аудиопособие **«Тõдмõдам дзюлюкъясõс финн-йõгра войтыркõд»** («Знакомим детей с финно-угорскими народами») (автор-составитель В.А. Муравьёва). В пособии представлены аудиозанятия на коми языке: «Мадьяр (венгрыяс)», «Ханты да манси», «Финьяс», «Вепсьяс», «Коми», «Мордва», «Мари», «Эстьяс».

6. Аудиопособие **«Съылыштам, дзюлюк!»** (Споём, малыш!) (составитель Н.Н. Морозова). Представлены 17 детских песен и фонограммы. Начинается аудиосборник с гимна «Дзюлюк» («Малышок»).

7. Аудио-видеопособие **«Учимся петь и танцевать по-финно-угорски»** (составители Н.Н. Морозова, И.Н. Леонова). На CD-диске «Музыка финно-угорских народов» собраны музыкальные произведения финно-угорских народов. На DVD-диске «Учимся танцевать по-финно-угорски» демонстрируются танцы финно-угорских народов. Кроме того,



показываются и объясняются основные движения танцев, положения рук и ног (всоеобразный самоучитель финно-угорским танцам).

8. Аудиопособие **«Тодмёдам дзюлюкъясёс музыкаён»**: «Дзюлюк» уджтас серти кывзём вылё диск / Лёсьбдіс Н.Н. Морозова. – Сыктывкар, 2012. На аудиодиске представлены музыкальные занятия для детей дошкольного возраста. В ходе занятий дети знакомятся с известными коми и русскими композиторами, их произведениями, а также с коми народными песнями.

Особое место в перечне ЭОР занимают видеопособия, которые позволяют в доступной форме, ярко, образно, преподнести дошкольникам материал, что соответствует наглядно-образному мышлению детей дошкольного возраста:

1. Видеопособие **«Вогёгёрся гажъяс»** («Праздник круглый год») (автор-составитель З.В. Остапова). В данном пособии представлены видеозанятия: «Вогёгёрся йёзкостса гажъяс» (Коми народные праздники), «Йёввыв» (Масленица), «Ыджыд лун» (Пасха), «Пёкрёв», «Победа лун», «Вильвоасём» (Новогодняя).

2. Видеопособие **«Ворсам театрысь. 1 петас»**: «Дзюлюк» уджтас серти видзёдём вылё диск = Играем в театр. 1 выпуск / Лёсьбдіс З.В. Остапова. – Сыктывкар, 2012. Диск подготовлен на основе материалов детской познавательной телепередачи «Чолём, дзюлюк!» («Привет, малыш!»).

3. Видеопособие **«Ворсам театрысь. 2 петас»** = Играем в театр. 2 выпуск / З.В. Остапова. – Издатель ООО «Телесемь Коми». Диск создан на основе детской телепередачи «Чолём, дзюлюк!», автор разработки З.В. Остапова, режиссёр М. Кузьчуктова. На следующие записи диск вошли: А.С. Пушкин «Сказка о царе Салтане» – ДОО № 8 с. Визинга Сысольского района; В. Лыткин «Дзизгысь-дзизгысь гут» – ДОО № 45 Эжвинского района г. Сыктывкара; Мойд туй, Сыктывдинский район, «Ыбица» парк; Г. Юшков «Куимпорсьпи» – ДОО № 8 с. Визинга Сысольского района.

4. Видеопособие **«Ворсам театрысь бать-мамкёд тшётш»** (Играем в театр вместе с мамой и папой) / автор-составитель З.В. Остапова, режиссёр М. Кузьчуктова, звукорежиссёр А. Шабалин, монтаж С. Шмаков, актёры А. Софронова, В. Пивкина. – Издатель ООО «Телесемь Коми». На диск вошли фестивали и спектакли с участием детско-родительских театров детских садов Республики Коми: с. Пажга Сыктывдинского район, с. Сторожевск Корткеросского района, с. Куратово Сысольского район, детских садов Усть-Куломского района.

5. Видеопособие **«Ворсыштам, дзюлюк!»** = Поиграем, малыш! / автор-составитель З.В. Остапова, режиссёр М. Кузьчуктова, звукорежиссёр А. Шабалин, монтаж С. Шмаков, актёры А. Софронова, В. Пивкина. –



Издатель ООО «Телесемь Коми». Диск создан в помощь воспитателям дошкольных образовательных учреждений и родителям. На диск вошли коми народные игры, музыкальные игры, подвижные игры, игры на развитие речи, театрализованные игры.

6. Видеопособие **«Татчö волыны ме рад!»**: «Дзюлюк» уджтассер-тивидзöдöмвылö диск = Сюда я рад приходить! / Лöсьöдис З.В. Остапова. – Сыктывкар, 2012. Диск подготовлен на основе материалов детской познавательной телепередачи «Чолöм, дзюлюк!» («Привет, малыш!»). Автор – З. Остапова, режиссёр – М. Кузьчукмова, артисты: А. Софронова, В. Пивкина.

7. Видеопособие **«Чолöм, дзюлюк!»** (Привет, малыш!). На диске собраны 17 познавательных и обучающих передач для детей дошкольного возраста на коми языке. Передачи предназначены для детской аудитории в возрасте 3–7 лет. В названии передачи «Чолöм, дзюлюк!» заложена идея дружеского отношения к ребёнку её персонажей, предлагающих ему полезные и интересные занятия – от знакомства с окружающим миром и разучивания чисел и букв до решения логических и этических задач. Важное место в телепрограмме занимает освещение жизни детей в разных районах Республики Коми, что способствует созданию единого образовательного пространства в детской среде. Основные персонажи передачи – забавный петушок Кикурулло и рассудительная собачка Ваткыль – ведущие и собеседники для детей.

Так или иначе, дидактические материалы в электронном виде начинают занимать свою нишу в образовательном пространстве дошкольных образовательных организаций, позволяя сделать привлекательным процесс обучения коми языку. Сегодня их использование позволяет:

- показать информацию на экране в игровой форме, что вызывает у детей огромный интерес, т.к. это отвечает основному виду деятельности дошкольника – игре;
- ярко, образно преподнести дошкольникам материал, что соответствует наглядно-образному мышлению детей дошкольного возраста; привлечь внимание детей движением, звуком, мультипликацией, но не перегружать ими материал;
- способствовать развитию у дошкольников исследовательских способностей, познавательной активности, навыков и талантов;
- поощрять детей при решении проблемных задач и преодолении трудностей.

Использование дидактических материалов в электронном виде в дошкольном образовании даёт возможность расширить творческие возможности самого педагога, что оказывает положительное влияние на развитие дошкольников.



Сегодня в Лаборатории создаются мультимедийные электронные издания с материалами из опыта работы педагогов детских садов республики. Это позволяет представить необходимый практический набор наглядного материала для детей, интерактивные игры, кроссворды, видеоматериал. К таким изданиям относятся:

1. **Диск «Юргö коми кыв» («Звучит коми язык»):** материалы I Республиканского фестиваля творческих идей этнокультурной направленности в дошкольном образовании (авторы-составители З.В. Остапова, Л.Г. Кулышева). Диск создан по итогам Фестиваля творческих идей этнокультурной направленности «Юргö коми кыв» («Звучит коми язык») в целях трансляции инновационного опыта коми республиканских ресурсных центров, опорно-методических площадок, базовых национальных детских садов Республики Коми и поддержки дошкольных образовательных организаций, активно внедряющих в дошкольную среду национальные культурные традиции.

2. **Диск «Детям о Республике Коми: дидактический материал для обеспечения реализации этнокультурного образования»** (авторы-составители З.В. Остапова, Т.И. Чудова, Л.Г. Кулышева, М.В. Рудецкая). На диск вошли разнообразные дидактические материалы по приобщению дошкольников к культуре, традициям, фольклору, языку и искусству коми народа, формированию первоначальных представлений о Республике Коми, воспитанию любви к своей Родине.

Диск «Детям о Республике Коми» является одной из составляющих учебно-методического комплекта, разработанного сотрудниками лаборатории национальных проблем дошкольного образования под руководством канд. пед. наук З.В. Остаповой. Представленный на диске иллюстративный материал поможет обогатить образовательную деятельность с детьми дошкольного возраста, для которых главным является наглядность. Разглядывание фотографий, картин, просмотр мультфильмов и видеофильмов оставит яркие впечатления от рассказа педагога об особенностях традиционной культуры коми народа, пробудит интерес узнавать всё больше о нашей Республике Коми. В дальнейшем, мы надеемся, дети смогут увидеть многое в музеях, а также во время путешествий по республике, и тогда они сами станут интересными собеседниками на разные темы, связанные с историей, традициями и особенностями культуры коми народа. На диске представлены уникальные полевые фотографии и фотографии из собраний Национального музея Республики Коми и музея археологии и этнографии Сыктывкарского государственного университета имени Питирима Сорочкина; картины художников Республики Коми из фондов Национальной галереи Республики Коми; телевизионные фильмы КРТК «Юрган»; мультфильмы, подготовленные сотрудниками Национального музея Республики Коми,



а также практические материалы сотрудников лаборатории национальных проблем дошкольного образования Коми республиканского института развития образования (КРИО) и педагогов-победителей республиканских конкурсов этнокультурной направленности. Диск адресован педагогическим работникам дошкольных организаций, реализующим этнокультурное содержание в образовательной деятельности в условиях Федерального государственного образовательного стандарта дошкольного образования, а также будет полезен и для семейного воспитания.

Умение пользоваться Интернетом позволяет быть в курсе событий, происходящих в педагогических сообществах, отслеживать анонсы мероприятий (проведение конкурсов, семинаров), получать консультации по возникшим проблемам, размещать свои работы на сайтах и знакомиться с разработками мероприятий коллег. Общаясь на форумах с коллегами всей России, можно заявить о себе и своей деятельности педагогическому сообществу.

В 2014 г. Лаборатория создала свою **группу в контакте «Дзюлюк-Малышок»** (<https://vk.com/club69286686>), где размещены такие рубрики, как: полоса новостей, полезные ссылки, анонс новинок литературы этнокультурного содержания, детской телепередачи «Чолъм, дзюлюк!».

С 2013 г. Лаборатория является участником Проекта «Финно-угорские языки и культуры в дошкольном образовании», в котором участвуют все финно-угорские регионы России: Карелия, Коми, Марий Эл, Мордовия, Удмуртия, Пермский край, Ханты-Мансийский, Таймырский Долгано-Ненецкий автономные округа, частично Ненецкий автономный округ, Архангельская, Ленинградская, Вологодская, Тверская области. Сегодня активно действует виртуальная сеть – информационный портал FINNOUGR-DOU.org. Здесь размещаются различные материалы образовательных семинаров, полезные ссылки для всех, кто интересуется двуязычным образованием. В разделе «Блоги» публикуются материалы финно-угорских регионов России.

Сегодня можно говорить о том, что введение компьютера в систему дидактических средств дошкольных образовательных организаций является мощным фактором обогащения интеллектуального, нравственного, эстетического развития ребёнка, а также приобщения его к миру информационной культуры. Однако, несмотря на значительные возможности ЭОР в работе с детьми дошкольного возраста, актуальным является создание интерактивных материалов по коми языку и краеведению для детей дошкольного возраста в Интернете, что будет способствовать повышению интереса к коми языку и расширению возможностей его функционирования.



* * *

1. Об образовании в Российской Федерации: федеральный закон от 29.12.2012 № 273-ФЗ // Собр. законодательства РФ. – 2012. – № 53 (ч.1). – Ст.7598.

2. *Могилева В.Н.* Психофизиологические особенности дошкольника и их учёт в работе с компьютером: учеб. пособие для студентов образовательных учреждений сред. проф. образования. – М.: Академия, 2007.

3. *Трофимова В.В.* Формирование информационно-технологической культуры младших школьников // Информатика и образование. – 2007. – № 5. – С.125–126.

4. Рекомендации по использованию компьютеров в детском саду: письмо Министерства образования РФ // Информатика и образование. – 2006. – № 6. – С.34–40.

5. Концепция развития этнокультурного образования в Республике Коми на 2016–2020 гг.: приказ Министерства образования Республики Коми от 23.11.2015 № 255. – URL: <http://www.komiedu.ru/communities/dopobr/docs/detail.php?ID=7300>



К. Пишлётгер,

*докторант Венского Университета,
стажёр Марийского государственного
университета по программе Эрасмус
(координатор Университет Турку)*

УДМУРТСКИЙ ЯЗЫК В СОЦИАЛЬНОЙ СЕТИ «ВКОНТАКТЕ»: КВАНТИТАВНЫЕ И (ВОЗМОЖНЫЕ) КВАЛИТАТИВНЫЕ ИССЛЕДОВАНИЯ*

Среди финно-угорских языков России удмуртский язык является одним из самых востребованных в социальных сетях Интернета, особенно в соцсети «ВКонтакте», которая была создана в 2006 году. Уже в ноябре 2007 г. была создана группа «Удмуртлык» [1], которая до сих пор остаётся самым значительным сообществом удмуртского Интернета, а в марте 2008 г. появилась группа ЮМШАН57** [2] в соцсети «ВКонтакте». Обе группы играют важную роль в истории удмуртского онлайн-активизма. Самым большим достижением удмуртского Интернета на российском и даже на международном уровне стал успех группы «Бурановские бабушки» из удмуртского с. Бураново, победившей в российском этапе и занявшей 2-е место в финале конкурса «Евровидение» в Баку [3]. В 2008 г. началась работа над удмуртским интерфейсом, который вышел в 2011 году как результат коллективной работы [4]. Деятельность удмуртских пользователей Интернета стала предметом и западных исследований [5], недавно значительная роль удмуртского языка в соцсети «ВКонтакте» была подтверждена и количественными данными проекта «Языки России» [6] школы лингвистики Национального исследовательского университета «Высшая школа экономики» (далее – НИУ ВШЭ)***.

© Пишлётгер К., 2017

* Благодарю Б.В. Орехова и исследовательскую группу «Языки России» НИУ ВШЭ за создание нужного и актуального для исследователей корпуса, и лично Б.В. Орехова за уточнения и ответы на мои вопросы в ходе подготовки данной статьи.

** Группа ЮМШАН57 сейчас имеет другое название, потому что администратор решил не вести её дальше, и пока дальнейшая судьба группы не ясна, но все прежние посты здесь сохраняются. Мы упоминаем о данной группе потому, что она занимает значимое место в истории удмуртского Интернета, речь о ней пойдёт далее в статье.

*** См. вторую таблицу «Количество сообществ в ВКонтакте, приходящихся на малый язык», где удмуртский язык занимает 1-е место, опережая даже якутский, башкирский и татарский (см.: <http://web-corpora.net/wsgi3/minorlangs/view>).



Сколько удмуртских групп в соцсети «ВКонтакте»?

Подсчёт удмуртских групп «ВКонтакте» в декабре 2013 г., произведённый автором путём ручного поиска, выявил примерно 90 сообществ. Среднее число участников и подписчиков составляло 696 человек, имелась 481 запись. Медиана составляла 275 участников и подписчиков и 113 записей [7]. Автоматический подсчёт, сделанный исследовательской группой «Языки России» НИУ ВШЭ (данные за май 2016 г.), выявил 857 удмуртских URL в соцсети «ВКонтакте». Однако эти результаты не совпадают с реальным числом удмуртских групп и страниц «ВКонтакте», потому что там собраны не только страницы и группы, но и мероприятия, обсуждения, личные профили и т.п. (список URL можно найти на странице проекта) [8]. В итоге после вычистки (исследовательская группа, например, не рассматривала личные профили [9]) – тексты 288 групп, страниц и мероприятий (данные за май 2016 г.) были поставлены на сайте, откуда их можно скачать [10].

Для того чтобы сравнить эти данные с ручным подсчётом (декабрь 2013 г.), автором были удалены все мероприятия, ненайденные группы и страницы, неумуртские группы и страницы (об этом пойдёт речь ниже). Таким образом, было получено 196 групп/страниц. Среднее число участников/подписчиков составляет 791 человек и 835 записей. Как и при ручном подсчёте, проведённом автором тремя годами ранее (декабрь 2013 г.), медиана значительно ниже – 242 участника/подписчика, 149 записей (см. табл.). Существенное различие между средним числом и медианой можно объяснить, по крайней мере частично, сравнительно высокой активностью некоторых удмуртских пользователей и удмуртских групп [11].

Т а б л и ц а

*Развитие некоторых показателей удмуртских групп
в соцсети «ВКонтакте»*

	Ручной подсчёт (декабрь 2013 г.)	Автоматический подсчёт на основе данных проекта «Языки России» (май 2016 г.) с ручной обработкой автора
Число групп	90	196
Среднее число	696 участников 481 запись	791 участник 835 записей
Медиана	275 участников 113 записей	242 участника 149 записей



Удмуртский язык в проекте «Языки России»

Насколько нам известно, в настоящее время наряду с проектом «Языки России» существует ещё 3 проекта по созданию корпуса удмуртского языка на основе его представления в социальных сетях и в Интернете – это финский проект SUKI [12] (*Suomalais-ugrilaiset kielet ja internet*, в переводе на русский – «Финно-угорские языки и Интернет» [13]) венгерский проект FinUgRevita «Компьютерные средства для ревитализации финно-угорских языков под угрозой исчезновения» [14] и ещё 1 проект НИУ ВШЭ – «Корпус удмуртского языка» [15]. Последний проект охватывает, кроме других жанров, и удмуртские тексты в соцсетях, но не в соцсети «ВКонтакте», а в первую очередь удмуртские блоги и статьи из Удмуртской Википедии.

Поскольку только проекту «Языки России» на сегодняшний день удалось выложить в свободном доступе результаты своей работы с соцсетью «ВКонтакте» в форме скачанных текстов, то ниже речь пойдёт именно об этом проекте. Цель проекта, реализованного в рамках магистерской программы «Компьютерная лингвистика» школы лингвистики НИУ ВШЭ, состояла в поиске и сохранении интернет-ресурсов на миноритарных языках Российской Федерации, т.е. не только уральских. В целом были найдены 96 «малых» языков Российской Федерации [16]. Проект отличается также особенным вниманием к соцсети «ВКонтакте», и как отмечалось, это единственный корпус удмуртских текстов «ВКонтакте», который даёт хорошую основу для дальнейших исследований. Отметим некоторые сложности, возникшие при работе с этой частью корпуса.

Как было сказано выше, сегодня не все страницы/группы доступны: они «исчезают», становятся закрытыми, меняют администраторов, названия и/или тематику, что естественно для соцсетей. Значительным препятствием при анализе групп является то, что в список попали группы и страницы, которые не рассчитаны на удмуртских пользователей, а адресованы, в частности, якутской [17], бурятской [18], горномарийской [19], или общей аудитории [20]. Это, к сожалению, искажает и количественный анализ, т.к. некоторые группы являются достаточно большими группами. Например, группа «Stream by Kudes» [21] имеет 36 тыс. участников, «Платформа сеть спортивных клубов» [22] – примерно 29 тыс. участников (данные за январь 2017 г.). То, что в список удмуртских групп/страниц попали не удмуртские группы/страницы, может быть связано с неверно выбранными лексическими маркерами. Вероятно, этого можно избежать, уточнив их выбор.

Следующая проблема состоит в том, что, к сожалению, не удалось собрать все посты. Особенно это заметно на примере группы «ЮМШАН57», первый пост которой был написан 22 марта 2008 г. [23]. Однако в качестве



первого в проекте представлен лишь 916-й пост, который был создан 29 октября 2010 г. [24]. Таким образом, в группе «ЮМШАН57», по данным проекта, отсутствуют первые 915 постингов, или примерно первые 1,5 года. Спорным является и решение исследовательской группы игнорировать удмуртские тексты, написанные не кириллицей [25], хотя удмурты используют и латиницу. В качестве пример можно привести группу из соцсети «Фейсбук» «Udmurt sjamen» [26], где не только название, но и некоторые посты на удмуртском написаны латиницей.

(Возможные) качественные анализы

При качественном анализе удмуртских групп соцсети «ВКонтакте» мы обнаружили некоторые проблемы. Например, файлы здесь располагаются по внутренней логике JSON (*.json*), т.е. не в (хроно)логическом порядке постов и комментариев к ним. Метаданные [27], которые имели важную функцию при количественном анализе, мешают при качественном анализе, т.к. они никак не отделены от текста постов. Исходя из целей исследовательской группы, главной задачей было собрать письменные тексты на миноритетных языках России, поэтому члены групп не старались отражать интер-/трансмедиальность, т.е. картинки, видео-, аудиоматериалы и их взаимодействие, хотя для некоторых исследований это необходимо. Особенно ярким примером является самая большая удмуртская группа в «ВКонтакте» «Эктоника» [28], где содержится очень много языкового материала, представленного именно в аудио и визуальном виде, что не отражается в собранных текстах.

О каких именно качественных анализах идёт речь, и как должен выглядеть корпус, чтобы удовлетворять этим целям? Корпус можно использовать, например, при дискурс-анализе (напр. дискурс-лингвистика, критический дискурс-анализ (CDA), социосемиотика [29], и также корпус-лингвистический дискурс-анализ [30]) или качественном контент-анализе [31]. Также используются корпуса при изучении билингвизма, в том числе, и переключения кодов [32]. Для этого можно применять различные софтверы, которые позволяют даже не специалистам по компьютерной лингвистике проводить анализ. Их можно использовать так просто как удмуртский онлайн-корпус [33], но они предоставляют и иные возможности для работы (преимущество удмуртского онлайн-корпуса состоит в наличии аннотированных текстов и морфологического анализатора).

Можно привести в качестве примера такие софтверы, как анализаторы коллокаций (англ. «concordancer, concordance tools»), например AntConc для офлайн-использования [34], SketchEngine для онлайн-использования [35], которые помогают при дискурс-анализе, или MAXQDA [36] – при качественном контент-анализе. Эти софтверы работают с простыми



форматами файлов (например, *.doc*, *.docx*, *.htm*, *.html* или *.txt*) и частично даже могут отражать и трансмедийность, как, например, MAXQDA.

Тексты в корпусе могут также не иметь аннотирования. Например, 2 таких известных корпуса английского языка, как Lancaster-Oslo-Bergen Corpus (LOB) с текстами на британском английском [37] или Brown Corpus с текстами на американском английском [38], существуют в аннотированных и неаннотированных версиях. При этом даже неаннотированные тексты дают основу для исследований и возможность получить интересные результаты. Поэтому хотелось бы рядом с файлами в формате JSON (*.json*) видеть и такие версии скачанных полных текстов со всеми постами и комментариями к ним в хронологическом порядке и без метаданных, которые можно использовать вышеупомянутыми софтверами и без специальной обработки. Таким образом, результаты проекта «Языки России» стали бы доступными более широкому кругу исследователей.

Другой вопрос, который однако не входит в тематику данной статьи, это вопрос референциального корпуса, например, при создании списка ключевых слов. Какие именно тексты должны войти в референциальный корпус? Будут ли это тексты всех групп, найденные при помощи ручного или автоматического подсчёта, как описано выше? Ответить на эти вопросы затруднительно вследствие ряда причин. Так, тексты в удмуртских группах «ВКонтакте» характеризуются языковой разнородностью: здесь встречаются тексты на «чистом» удмуртском (удм. *чылыкыт удмурт*), диалектном удмуртском, с переключением кодов (удм. *сур-пожо*), на «олбанском языке» («падонкафский йезыг») [39], на литературном русском языке. Часто они написаны кириллицей, но частично и латиницей, содержат ошибки, типичные для быстрого и невнимательного написания текстов и сообщений в соцсетях, не говоря уже о разных вариантах передачи «удмуртских» графем, не существующих в русском алфавите (*Ж, Ъ, Ъ, Ъ, Ъ, Ж, ъ, ѡ, ѡ, ѡ*), несмотря на доступность технических средств, позволяющих грамотно писать на компьютерах и мобильных устройствах.

Заключение

Среди уральских и даже среди всех миноритетных языков России удмуртский язык является значительно востребованным в соцсетях Интернета. Существуют разные количественные и качественные исследования применения удмуртского языка [40] или удмуртского языка одновременно с другими языками [41] в Интернете. Из существующих проектов, насколько нам известно, пока только проекту «Языки России» удалось собрать тексты на удмуртском языке в соцсетях, в частности «ВКонтакте», с целью количественного анализа, однако их можно использовать и в качестве основы для разных качественных исследований, например дискурс-анализа, качественного контент-анализа и т.д.



Корпус имеет некоторые объективные недостатки (не все собранные группы/страницы являются удмуртскими, включены группы/страницы для другой аудитории, не удалось собрать все посты и т.п.), которые можно исправить в последующих анализах. Наблюдаются и технические проблемы, связанные с расположением отдельных постов и комментариев в файлах в формате JSON. Они не позволяют лингвистам, антропологам, социологам и другим исследователям, интересующимся удмуртским или другими миноритетными языками в соцсетях, но не являющимся специалистами в компьютерной лингвистике, использовать их без обработки третьими лицами. Для вышеупомянутых исследований есть отдельные софтверные решения (concordancer, QDA software и т.д.), которым для первых шагов качественного анализа достаточно текстов без таких специальных требований, как аннотирование. Представление таких текстов рядом с уже существующими увеличит круг исследователей, которые могут работать с ними, и расширит возможности их изучения.

* * *

1. УДМУРТЛЫК ■*■ (удмуртъёс, удмурты). – URL: <https://vk.com/udmurt> (дата обращения: 31.01.2017).

2. Князь Поздей. Экс-Юмшан57. Шулдыр информ-агентство. – URL: <https://vk.com/club1966810> (дата обращения: 29.01.2017).

3. *Pischlöger C.* Udmurt on Social Network Sites: A Comparison with the Welsh Case // *Linguistic Genocide Or Superdiversity?: New and Old Language Diversities.* – 2016. – Т.14. – С.115.

4. *Pischlöger C.* Udmurtness in Web 2.0: Urban Udmurts resisting Language Shift // *Finnisch-Ugrische Mitteilungen.* – 2014. – Т.38. – С.149–150.

5. *Пишлӧгер К.* «Бурановские бабушки» на Youtube: роль социальных сетей в сохранении удмуртской идентичности и языка в условиях глобализации // *Проблемы сохранения нематериального культурного наследия в условиях глобализации.* – Ижевск: УИИЯЛ УрО РАН, 2013. – С.161–165; *Пишлӧгер К.* Запис(к)и из Муржол Underground: Super удмурты в Cyberspace // *Творчество Флора Васильева и вопросы языка, литературы, образования в глобализирующемся мире: материалы IV Международной науч.-практ. конф. «Флоровские чтения», посвящённой 80-летию удмуртского поэта Флора Ивановича Васильева.* – Глазов: Глазов. гос пед. ин-т, 2014. – С.56–59; *Cagnoli S.* Les langues finno-ougriennes dans la révolution médiatique du «Web 2.0» // *Études finno-ougriennes.* – 2012. – № 44. – С.18; *Casen M.* La langue oudmourte en Oudmourtie entre 1990 et



2013: état des lieux et perspectives // *Études finno-ougriennes*. – 2013. – № 45; *Casen M.* Contemporary Udmurt Ethnic Activity in the context of Udmurt Identity Issues // *Journal of Ethnology and Folkloristics*. – 2014. – Т.7. – № 1. – С.91–110; *Pischlöger C.* Udmurtness in Web 2.0: Urban Udmurts resisting Language Shift...; *Pischlöger C.* Udmurt on Social Network Sites: A Comparison with the Welsh Case...

6. Языки России. – URL: <http://web-corpora.net/wsgi3/minorlangs> (дата обращения: 24.01.2017).

7. *Pischlöger C.* Udmurtness in Web 2.0: Urban Udmurts resisting Language Shift... – С.115.

8. URL: http://web-corpora.net/wsgi3/minorlangs/static/files/url_lists/udm_url_lists.zip

9. *Krylova I. et al.* Languages of Russia: Using Social Networks to Collect Texts // *Information Retrieval: 9th Russian Summer School, RuSSIR 2015, Saint Petersburg, Russia, August 24–28, 2015, Revised Selected Papers* / ed. by P. Braslavski и et al. – Cham: Springer International Publishing, 2016. – P.182.

10. URL: http://web-corpora.net/wsgi3/minorlangs/static/files/vk_corpus/udm_vk_corpus.zip

11. *Pischlöger C.* Udmurtness in Web 2.0: Urban Udmurts resisting Language Shift... – С.151; *Pischlöger C.* Udmurt on Social Network Sites: A Comparison with the Welsh Case... – С.116.

12. *Jauhiainen H., Jauhiainen T., Lindén K.* The Finno-Ugric languages and the Internet project // *Septentrio Conference Series*. – 2015. – С.87; *Vincze V. et al.* FinUgRevita: Developing language technology tools for Udmurt and Mansi // *Septentrio Conference Series*. – 2015. – P.110.

13. SUKI (Finno-Ugric Languages and the Internet). – URL: <http://suki.ling.helsinki.fi/eng/project.html> (дата обращения: 24.01.2017).

14. Компьютерные средства для ревитализации финно-угорских языков, находящихся под угрозой исчезновения (FinUgRevita). – URL: http://www.ieas-szeged.hu/finugrevita/project_ru.html (дата обращения: 24.01.2017).

15. Корпус удмуртского языка. – URL: http://web-corpora.net/UdmurtCorpus/search/?interface_language=ru (дата обращения: 06.12.2016).

16. *Orekhov B. et al.* Russian Minority Languages on the Web: Descriptive Statistics // *Proceedings of the Annual International Conference «Dialogue» (2016) – Computational Linguistics and Intellectual Technologies*. – Moscow: RSUH, 2016. – P.501.

17. We ♥ dormitory 7/2. – URL: <https://vk.com/club79725008> (дата обращения: 29.01.2017); Заочные Студенты 2012 АГИИК «Менеджмент». – URL: <https://vk.com/club40986644> (дата обращения: 29.01.2017).



18. =====Клан «БАТУЕВЫХ» Бурятия-Улан-Удэ=====
- == – URL: <https://vk.com/club18757275> (дата обращения: 29.01.2017).
19. Горномарийцы в Йошкар-Оле ~ Кырык марынвла Йошка. – URL: <https://vk.com/club10286319> (дата обращения: 29.01.2017).
20. См., напр.: Stream by Kudes. – URL: <https://vk.com/kudestv> (дата обращения: 29.01.2017); Платформа сеть спортивных клубов. – URL: <https://vk.com/platformasport> (дата обращения: 29.01.2017).
21. Stream by Kudes...
22. Платформа сеть спортивных клубов...
23. URL: https://vk.com/club1966810?w=wall-1966810_1
24. URL: https://vk.com/club1966810?w=wall-1966810_915_r916
25. *Orekhov B. et al.* Op. cit. – P.502.
26. Udmurt sjamen. – URL: <https://www.facebook.com/groups/121020634578219> (дата обращения: 29.01.2017).
27. *Krylova I. et al.* Op. cit. – P.182.
28. Эктоника. – URL: <https://vk.com/ektonika> (дата обращения: 29.01.2017).
29. *Bendel Larcher S.* Linguistische Diskursanalyse: ein Lehr- und Arbeitsbuch. – Tübingen: Narr Francke Attempto, 2015.
30. *Bubenhof N., Scharloth J.* Korpuslinguistische Diskursanalyse: Der Nutzen empirisch-quantitativer Verfahren // Diskurslinguistik im Spannungsfeld von Deskription und Kritik / под ред. Warnke Ingo, Meinhof Ulrike, Reisigl Martin. – Berlin: Akademie-Verlag, 2013. – С.147–168.
31. *Kuckartz U.* Einführung in die computergestützte Analyse qualitativer Daten. – Wiesbaden: VS, Verl. für Sozialwiss, 2007. – Вып.2, aktualisierte und erw. Aufl; *Kuckartz U.* Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung. – Weinheim Basel: Beltz Juventa, 2016. – Вып. 3. – überarbeitete Auflage.
32. *Backus A.* Data Banks and Corpora // The Blackwell guide to research methods in bilingualism and multilingualism / ed. by L. Wei, M.G. Moyer. – Malden, MA: Blackwell Pub, 2008. – P.232–248.
33. Корпус удмуртского языка...
34. Laurence Anthony's AntConc. – URL: <http://www.laurenceanthony.net/software/antconc> (дата обращения: 29.01.2017).
35. Sketch Engine. – URL: <https://the.sketchengine.co.uk/login> (дата обращения: 29.01.2017).
36. MAXQDA: Qualitative Data Analysis Software | Windows & Mac – MAXQDA – The Art of Data Analysis. – URL: <http://www.maxqda.com> (дата обращения: 29.01.2017).



37. Oxford Text Archive B.L. Lancaster-Oslo-Bergen corpus of modern English [Electronic resource]: [tagged, horizontal format] / Stig Johansson. – URL: <http://ota.ox.ac.uk/desc/0167> (дата обращения: 29.01.2017).

38. Brown Corpus Manual. – URL: <http://www.hit.uib.no/icame/brown/bcm.html> (дата обращения: 29.01.2017).

39. Кронгауз М.А. Самоучитель олбанского. – М.: AST, 2013; *Pischlöger C.* Udmurt on Social Network Sites: A Comparison with the Welsh Case... – С.117.

40. Пишлөгер К. «Бурановские бабушки» на Youtube: роль социальных сетей в сохранении удмуртской идентичности и языка в условиях глобализации.; *Пишлөгер К.* Запис(к)и из Муржол Underground: Super удмурты в Cyberspace.; *Casen M.* La langue oudmourte en Oudmourtie entre 1990 et 2013: état des lieux et perspectives.; *Casen M.* Contemporary Udmurt Ethnic Activity in the context of Udmurt Identity Issues.; *Pischlöger C.* Udmurtness in Web 2.0: Urban Udmurts resisting Language Shift...

41. Зайдельман Л.Я., Крылова И.В., Орехов Б.В. Технология поиска и сбора в Интернете текстов на малых языках России. – Ларнака, Кипр; ЦарьГрад, Моск. обл., 2015. – С.179–181; *Cagnoli S.* Op. cit; *Krylova I. et al.* Op. cit; *Orekhov B. et al.* Op. cit; *Pischlöger C.* Udmurt on Social Network Sites: A Comparison with the Welsh Case...



УДК 378.147:004

И.А. Попова,

*учитель русского языка
и литературы Центра
дистанционного образования
детей-инвалидов при ОГАОУ
«Белгородский инженерный
юношеский лицей-интернат»
(г. Белгород)*

РЕСУРСЫ ОБЛАЧНЫХ ИНТЕРНЕТ-ТЕХНОЛОГИЙ В СИСТЕМЕ ДИСТАНЦИОННОГО ПРЕПОДАВАНИЯ КУРСОВ РУССКОГО ЯЗЫКА И ЛИТЕРАТУРЫ В ОСНОВНОЙ ШКОЛЕ

Стремительное развитие информационно-коммуникационных технологий (далее – ИКТ) повлияло и на жизнь общества в целом, и на образовательную систему в частности. Так, в значительной мере изменились пути получения знаний (от классно-урочной системы до мобильного образования), изменился классификационный перечень типов уроков в основной общеобразовательной школе (от урока с применением ИКТ до мультимедийных уроков). Появляется и новая форма получения образования – дистанционная – со своими особенностями и требованиями к уроку.

Дистанционное преподавание русского языка и литературы – одно из важнейших в системе основного и среднего общего образования, адресованного детям с ограниченными возможностями здоровья. В Белгородской области проект организации дистанционного образования реализуется с 2010 г., и учителя, осуществляющие предметную подготовку по различным направлениям, находятся в постоянном поиске оптимальных инструментов реализации целей и задач, определённых образовательными стандартами учебных дисциплин. Кроме того, на выбор инструментов влияет индивидуальность ребёнка, его психофизиологическое состояние, особые возможности и потребности. Система работы дистанционного учителя определяется следующими характеристиками:

- 1) особенности дистанционной формы преподавания;
- 2) индивидуальные особенности ребёнка.

Поэтому, моделируя урок, учитель руководствуется требованиями подбора эффективного «визуального ряда», т.к. основной канал получе-



ния информации – визуальный, определяет формы, методы, программно-информационные инструменты, адекватные индивидуальности обучающегося (есть так называемые «пишущие» и «непишущие» дети, у которых затруднена или невозможна функция письма вследствие заболеваний опорно-двигательного аппарата). Некоторые аспекты дистанционного взаимодействия учителя и ученика на уроке суммированы А.А. Муссоровой [1], однако с течением времени создание электронных образовательных ресурсов значительно упрощается, и уже учитель (не программист!) создаёт и использует их различные виды на дистанционном уроке, в том числе на дистанционных уроках русского языка и литературы.

Электронные образовательные ресурсы на сегодняшний день чрезвычайно многообразны, однако одной из тенденций «нового» времени является их классификация, упорядочение и оценка соответствия существующим задачам, функциональности. Кроме существующих «площадок» федерального значения (к примеру, «Единая коллекция цифровых образовательных ресурсов»), учителя русского языка и литературы Центра дистанционного образования детей-инвалидов при ОГАОУ «Белгородский инженерный юношеский лицей-интернат» (далее – ЦЕНТР) используют ресурсы облачных интернет-технологий, в частности «*Google-Disk*», т.к. они позволяют строить системную работу с детьми с сохранённой функцией письма и без неё. «Применение облачных сервисов является неотъемлемой составляющей современного дистанционного образования», – справедливо считают авторы учебного пособия по применению облачных технологий в системе работы образовательного учреждения [2]. Создав учётные записи, учителя используют их в качестве «доски» для объяснения нового материала, закрепления изученного, проверки знаний (разного рода контрольных работ и работ по развитию речи) в режиме онлайн с «пишущими» детьми, которые работают, в том числе, и в рабочих тетрадях. «Непишущие» дети, кроме того, имеют возможность накапливать учебно-справочный материал, по которому затем могут готовиться к контрольным работам без постоянной пересылки файлов от учителя ученику и обратно, т.е. выполняется функция «тетради». Данный ресурс в Центре получил название «Электронная тетрадь».

Таким образом, «Электронная тетрадь» представляет собой текстовый электронный документ, созданный на основе облачных интернет-технологий «*Google-Disk*». В данном документе могут одновременно работать и учитель, и ученик удалённо в режиме реального времени, зная лишь адрес доступа к нему. Систематическая работа в подобного рода документе и накопление в нём учебного материала позволяют ему играть роль «до-



ски», «тетради» и служить эффективным инструментом на любом уроке русского языка и литературы, в том числе при создании и заполнении таблиц, рисунков, подготовке и проведении таких творческих работ, как сочинение, изложение и др.

* * *

1. *Муссорова А.А.* Особенности организации дистанционных уроков // Научный поиск. – 2014. – № 2.1. – С.19–20.

2. *Кузьмина Н.В., Пивоварова Т.С., Чупраков Н.И.* Облачные технологии для дистанционного и медиаобразования. – Киров: Изд-во КОГОКУ ДПО (ПК) «Институт развития образования Кировской области», 2013.



К.В. Рыбка,

студентка факультета

журналистики Белорусского

государственного университета

(Республика Беларусь, г. Минск)

ЗАИМСТВОВАННАЯ ЛЕКСИКА В СФЕРЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ КАК СТИЛИСТИЧЕСКОЕ СРЕДСТВО

На протяжении своего исторического развития русский язык пополнялся и обогащался иноязычными источниками. Некоторые заимствования многолетние, некоторые – относительно новые. Под заимствованной лексикой мы понимаем такой «лингвистический процесс, при котором слова одного языка переходят в другой на различных этапах развития последнего» [1].

Существует множество классификаций заимствований, которые приводятся различными филологами и лингвистами. Но среди всего разнообразия можно выделить основные признаки, по которым составляется классификация: это источник, время заимствования и тематика.

Так, по источникам выделяют славянские и неславянские заимствования, по времени – заимствования периода индоевропейской общности, общеславянского единства, древнерусского и собственно русского периода.

Наиболее интересная и разнообразная классификация – по тематике [2]: общественно-политическая лексика («демократия», «президиум», «электорат»); финансово-экономическая лексика («аренда», «бизнес», «инвестициях», «спонсор»); административно-правовая лексика («нотариус», «промоутеры», «шериф»); спортивная лексика («апперкот», «гейм», «скайсерфинг»); рекламная («маркетинг», «копирайтер»); техническая (компьютерная) лексика («роутер», «ксерокопия», «компьютер»); художественная и культурная лексика («фольклор», «ритм-энд-блюз», «джаз»); кулинарная лексика («фондю», «спагетти», «бутерброд»); лексика моды («хенд-мейд», «парфюм», «педикюр»).

Всю иноязычную лексику можно разделить на 2 группы [3]: освоенные заимствования – это заимствования, пришедшие в язык очень давно, которые уже не воспринимаются как чужие («рисуюнок», «сундук», «галстук», «серьга»); неосвоенные заимствования – заимствования, не освоенные русским языком, которые резко выделяются на фоне русской лексики.



К ним относятся экзотизмы («слова, которые изменяются только графически, их применение обусловлено необходимостью обозначить предметы и явления чужого народа или страны» («саке», «вигвам», «шатер») [4]), иноязычные вкрапления (чаще всего не меняют своего графического облика, их применение обусловлено «степенью знакомства с иностранным языком, некоторыми стилистическими или жанровыми особенностями речи» («фас», «хеппи-энд») [5]), интернациональные термины – слова, которые совпадают в разных языках по своей внешней форме с полностью или частично совпадающим смыслом («компьютер», «робот», «комбайн») [6]).

Проанализировав работы Е.И. Дибровой «Современный русский язык: Теория. Анализ языковых единиц», А.И. Дьякова «Причины интенсивного заимствования англицизмов в современном русском языке» и В.В. Виноградова «Вопрос об исторической словаре русского литературного языка XVIII–XX вв.», «Избранные труды. Лексикология и лексикография», мы пришли к выводу о том, что причинами заимствований могут быть: потребность в наименовании нового понятия, предмета, явления («ноутбук», «маркетинг», «копирайтинг»); отсутствие точного аналога в русском языке («топ-модель», «спонсор», «бизнес-ланч»); новое слово является наиболее удобным обозначением того, что до сих пор называлось словосочетанием («мерчендайзер» – человек, который занимается распространением продукции; «менеджер» – наёмный работник, занятый управлением на определённом участке производства или сферы услуг; «пуловер» – свитер без воротника и застёжек с треугольным вырезом); обогащение языка выразительными средствами («рандомный» – случайный, «шопинг» – покупка товаров); мнение, что заимствованные слова звучат более научно, престижно («симультанность» – синхронность; «бойфренд» – молодой человек, который находится в романтических отношениях; «менеджер» – наёмный работник, занятый управлением на производстве или в сфере услуг); необходимость в более узком значении слова («дизайнер»: «гейм-дизайнер» – дизайнер видеоигр; «ландшафт-дизайнер» – дизайнер садовых участков); необходимость указания на изменение социальной роли предмета в обществе («полиция» – милиция, «валидатор» – компостер); мировая тенденция к интернационализации лексического фонда.

Для полного перехода слов из одного языка в другой необходим определённый фундамент в виде психологической открытости общества к принятию заимствованной лексики, социальные, экономические, политические или культурные условия, которые позволяют (христианизация Руси) или заставляют (татаро-монгольское иго) принять заимствованные слова.

Но важнейшее условие перехода – это двуязычие. «Двуязычие – соприкосновение языков, контакт их носителей. В настоящее время контакт



носителей языка проявляется не только в устном и виртуальном общении, при непосредственном контакте, но и при опосредованном контакте: при чтении иностранных газет, книг и других текстов, при просмотре иностранных передач, изучении иностранных языков и др.» [7].

Таким образом, можно утверждать, что основные причины заимствования иностранных слов – это отсутствие эквивалентов в русском языке и языковая мода, тенденции, которой диктуют, как правило, Европа и Соединённые Штаты Америки, будучи наиболее развитыми странами.

В периоды социальных, культурных и научно-технических преобразований происходил наибольший наплыв иностранных слов в русскую лексику. Так было во времена татаро-монгольского ига (наименования предметов повседневности: «сарафан», «храп», «серьга»), период принятия христианства (богослужебная лексика и собственные имена: «ангел», «икона», «Фёдор»), эпоха петровских преобразований (военная и бытовая лексика: «лагерь», «вексель», «галстук»), период наполеоновских войн (политическая и бытовая лексика: «революция», «абажур», «батон»), революции начала XX в. (политическая и бытовая лексика: «буржуазия», «митинг», «тоннель»).

Обогащение лексики продолжается и в наше время. Сегодня одна из основных причин заимствования иностранных слов – развитие информационных технологий. В компьютерной сфере происходит множество изменений и открытий, поэтому именно в ней процесс заимствования новых слов идёт исключительно быстрыми темпами. Основное назначение лексем компьютерной лексики – наименование понятий компьютерной сферы.

Компьютерная лексика отличается стилистической неоднородностью. В её состав входят:

1) **литературная лексика** – представлена терминологией («графический интерфейс пользователей» – присутствие на экране изображений, которые позволяют пользователю легко взаимодействовать с компьютером) [8];

2) **нелитературная компьютерная лексика** – используется для придания языку разговорности. Применяется такая лексика в частных устных или письменных беседах. В её состав входят компьютерные профессионализмы и жаргон (сленг). Компьютерные профессионализмы относятся скорее к научному стилю. Кроме наименования понятий, в их функции входит создание удобной номинации. Например, к особенностям профессиональной компьютерной лексики относится тенденция сокращения английских выражений («WWW» – от англ. *World Wide Web*; «PC» – от англ. *Personal computer*, «SMM» – *Social Media Marketing*) [9]. Компьютерный жаргон придаёт речи экспрессивность, выполняет эмоционально-оценочную функцию. В компьютерном жаргоне часто присутствуют слова с де-



монстративно ошибочном написанием, часть этих слов стали уже традиционными («превед», «килита», «деффчонки», «извените») [10].

Первым пользователем Интернета было международное сообщество программистов, которые использовали англоязычные термины в качестве универсального средства коммуникации [11]. На сегодняшний день тенденция заимствования слов из английского языка не исчезла. Поэтому одна с особенностей компьютерной (как литературной, так и нелитературной) лексики – это большое количество англицизмов.

Англицизмы – лингвистические слова или выражения, заимствованные из английского языка (например, «имидж», «сканер») [12].

Определяют следующие типы англицизмов:

1) *прямые заимствования* – слова, которые встречаются в русском языке приблизительно в том же виде и в том же значении, что и в языке-доноре («смартфон», «чизкейк», «френдзона», «чайлдфри»);

2) *гибриды* – слова, образованные присоединением к иностранному корню русских аффикса, префикса или окончания («креативные» – от **creative** – (творческий); «зафрендить» – от **friend** (друг); «флудить» – от **flood** (поток (слов)), «флешка» – от **flash** (мгновенно));

3) *калька* – слова иноязычного происхождения, употребляемые с полным сохранением их фонетического и графического облика («Интернет» – от **Internet**, «диск» – от **disk**, «маркет» – от **market**, «клуб» – от **club**) или частичным – полукалька («чипсы» – от **chips**);

4) *экзотизмы* – слова, которые характеризуют специфические национальные обычаи других народов и употребляются при описании нерусской действительности. Отличительной особенностью данных слов является то, что они не имеют русских синонимов («чипсы» – от **chips**, «хот-дог» – от **hot-dog**, «шейк» – от **shake**, «бейсбол» – от **baseball**);

5) *иноязычные вкрапления* – данные слова обычно имеют лексические эквиваленты, но стилистически от них отличаются и закрепляются в той или иной сфере общения как выразительное средство, придающее речи особую экспрессию («о'кей» – от **OK** (хорошо); «баксы» – от **bucks** (деньги); «гудбай» – от **Goodbye** (до свидания));

6) *компози́ты* – слова, состоящие из двух английских слов («хэнд-мэйд» – от **hand made** (сделанный своими руками); «прайм-тайм» – от **prime time** (время, когда собирается большая аудитория); «арт-директор» – от **art-director** (художественный руководитель); «топ-менеджер» – от **top-manadger** (сотрудник фирмы, занятый продвижением));

7) *жаргонизмы* – слова, появившиеся благодаря намеренному искажению каких-либо звуков («ламер» – от **lame** (хромой); «мыло» – от **e-mail** (электронная почта); «клёвый» – от **clever** (умный); «писюха» – от **PC** (**Personal computer**) (персональный компьютер)).



Большая часть англицизмов относится к неосвоенной лексике. Поскольку мы живём в эпоху информационных технологий, которые быстро развиваются, то очень важно понимать значение новых терминов, уметь ориентироваться в них и правильно пользоваться ими.

Для анализа заимствованной лексики в материалах на ИТ-тематику, публикуемых в СМИ, мы использовали белорусский сайт «*Onliner.by*» и международный журнал «Компьютерная газета». К анализу привлекались англицизмы как наиболее популярные и недостаточно изученный на данный момент слой лексики. В работе мы использовали следующие словари: Л.М. Баш «Современный словарь иноязычных слов: толкование, словоупотребление, словообразование, этимология», Л.П. Крысин «Толковый словарь иноязычных слов», Н.Г. Комлев «Словарь новых иноязычных слов».

«*Onliner.by*» – один из наиболее популярных сайтов в Байнете. Количество просмотров его страниц достигает 10 млн в сутки. Тематика сайта – цифровые технологии. Следует обратить внимание, в том числе, на доменное имя сайта – «*Onliner*». «*Onliner*» – это относительно новое слово, и у сайта с таким названием есть свои преимущества. Во-первых, это интернациональность, которая даёт возможность посещать сайт не только жителям Беларуси, но и гражданам других стран. Во-вторых, название происходит от английского «*online*» – быть в Интернете, «*onliner*» – человек, который всегда в онлайне. Таким образом, сайт даёт посетителям гарантию наличия информационных материалов в течение 24 часов в сутки. Для точности анализа мы взяли по одному материалу из каждой рубрики сайта «*Onliner.by*».

«Компьютерная газета» – международное еженедельное специализированное издание, основная тематика которого – компьютерные технологии. Тираж в Беларуси составляет около 10 тыс. экземпляров. Для анализа мы также взяли несколько статей из разных рубрик.

Результаты анализа двух изданий показали, что:

- 1) большинство англицизмов используются для обозначения новых реалий, при этом сохраняется их первоначальное звучание;
- 2) наблюдается стремление к заимствованию слов ради престижа, лучшего звучания, что отчётливо видно в материалах «*Onliner.by*»;
- 3) в обзорных материалах на компьютерную тематику преимущественно применяется литературная компьютерная лексика, что заметно в обоих источниках;
- 4) в «*Onliner.by*» чаще всего встречаются жаргонизмы и профессиональная компьютерная лексика, причина этому – популярность сайта, ориентация на массового потребителя.



Подводя итог, можно сказать, что социально-экономические и политические условия определяют приоритетность того или иного мирового языка. Английский язык на современном этапе оказывает значительное влияние на развитие лексики множества языков. Можно сказать, что в конце XX столетия английский стал языком глобальной коммуникации. Поэтому на современном этапе очень важно ориентироваться в заимствованных словах, в частности в англицизмах.

Компьютерные технологии будут развиваться и дальше, а это значит, что словарный состав будет с каждым годом пополняться. На вопрос «Стоит ли опасаться интенсивного заимствования слов?» Х. Вальтер в своей книге «Словарь: Заимствования в русском субстандарте. Англицизм» отвечает так: «Теперь, когда мощный поток европеизмов и англицизмов хлынул на русский язык, многие лингвисты отказываются от директивных элиминации таких слов и выражений и апеллируют к “языковому вкусу эпохи” (А.А. Земская, В.В. Колесников, В.Г. Костомаров, Л.П. Крысин и др.). И действительно, бояться за судьбу русского языка из-за притока новых, пусть и “чужих”, слов не следует, так как языковая система, как показывает многовековой опыт, сама отбирает путь освобождения от лишнего, а полезное – сохраняет и развивает, делая “своим”» [13].

Таким образом, можно сделать вывод о том, что заимствованная лексика нуждается в подробном и регулярном исследовании, наличии толковых словарей современного типа, ежегодном их обновлении. Всё это нужно для того, чтобы иметь правильное представление о новых заимствованных терминах, которые являются неотъемлемой частью жизни человека в эпоху развития информационных технологий.

* * *

1. *Гринев С.В.* Терминологические заимствования (Краткий обзор современного положения вопроса) // Лотте Д.С. Вопросы заимствования иноязычных терминов и терминологических элементов. – М.: Наука, 1982. – С.201.

2. *Гринев С.В.* Указ. соч. – С.33.

3. *Брейтер М.А.* Англицизмы в русском языке: история и перспективы: пособие для иностранных студентов-руссистов. – Владивосток, 1997. – С.9.

4. *Брейтер М.А.* Указ. соч. – С.60.

5. *Брейтер М.А.* Указ. соч. – С.60.

6. *Брейтер М.А.* Указ. соч. – С.59.

7. *Брейтер М.А.* Указ. соч. – С.64.



8. *Галичкина Е.Н.* Специфика компьютерного дискурса в английском и русском языках: автореф. дис. ... канд. филол. наук. – Волгоград, 2001. – С.3.

9. *Колмыков А.А.* Интернет-журналистика. – М., 2005. – С.227.

10. *Колмыков А.А.* Указ. соч. – С.227.

11. *Колмыков А.А.* Указ. соч. – С.223.

12. *Диброва Е.И., Касаткин Л.Л.* Современный русский язык: Теория. Анализ языковых единиц: учеб. пособие. – Ростов н/Д: Феникс, 1997. – С.24.

13. *Вальтер Х.* Словарь: Заимствования в русском субстандарте. Англицизмы. – М., 2004. – С.5.



С.Д. Саая,
*научный сотрудник
сектора тувинского языка
Тувинского института
гуманитарных и социально-
экономических исследований
(Республика Тыва, г. Кызыл)*

СПОСОБЫ СОЗДАНИЯ КОМПЬЮТЕРНОГО ИНТЕРФЕЙСА НА ТУВИНСКОМ ЯЗЫКЕ (на основе социальной сети «ВКонтакте»)

Область применения тувинского языка в Интернете нельзя назвать развитой. Состояние Интернета в Тувинской Республике и услуг, связанных с ним, очень скромное: отсутствуют раскладки тувинских шрифтов для создания электронных ресурсов, необходимое программное обеспечение, единый компьютерный интерфейс на тувинском языке. В результате сужается сфера применения языка, падает уровень владения его литературными нормами. Тувинский язык не успевает за развитием информационных технологий.

В связи с этим появляется потребность в создании удобного и понятного компьютерного интерфейса. Понятие «интерфейс» можно трактовать как способ (стандарт) взаимодействия между объектами, а точнее между компьютерной программой и пользователем этой программы [1].

Для создания терминологии интерфейса была выбрана социальная сеть «ВКонтакте». В результате проведённого исследования было выявлено, что жители Республики Тыва больше ознакомлены с этим сайтом и регистрируются здесь чаще чем на других сайтах и в социальных сетях. Данная работа имеет большое значение, т.к. создание интерфейса позволит тувинскому языку расширить своё функциональное значение и обогатить лексику. Это связано с тем, что сегодня информационные технологии развиваются очень быстро, поэтому задача тувинского языка – успевать за новыми явлениями времени, не теряя своей области применения. К тому же важно не допускать «загрязнения языка» подобными конструкциями: «*настройкаларны редакторлаар*» – редактировать настройки, «*чуруктарны удаалгытаар*» – удалять фотографии и т.д. Отсюда нам видно, что при всей взаимозаменяемости допускаются такие «погрешности», которые на-



прямою переходят в разговорный язык, что уже сейчас наблюдается среди носителей тувинского языка.

Для создания компьютерного интерфейса на тувинском языке применяются следующие способы создания терминов: лексический, аффиксальный и синтагматический.

Лексический способ. Многозначные слова, попадая в систему знаний, вполне могут употребляться в функции терминов тувинского интерфейса. Метафора – результат метафорического переноса сигнификативного или денотативного содержания одного предмета на другой. Метафора, связанная с понятием, является важным средством для образования научных терминов [2]. С развитием научного мышления совершается перенос житейского опыта на специальную сферу деятельности. В этом плане метафорические образования открывают хорошие перспективы для терминологизации обыденных слов [3]. Усилия должны быть направлены на дальнейшее расширение этого способа. Таким образом, принадлежность к определенному терминологическому полю является самым существенным признаком, отличающим слова-термины от обычных слов. Для примера возьмём слово *арын* ('лицо'), которое в грамматике становится термином: *1 лицо, 2 лицо, 3 лицо* [4]. Точность термина достигается наличием дефиниции. Например, слово *арын* в данном случае в системе понятий грамматики обозначает категорию, показывающую отношения действия к говорящему и выражающуюся изменением глаголов в соответствующих формах спряжения. Если вывести данное слово из системы грамматических понятий, то оно потеряет своё терминологическое значение и превратится в обыденное слово с понятийным содержанием «передняя часть головы человека».

Метафорические образования являются важным средством для образования научных терминов. В этом плане метафорические образования открывают хорошие перспективы для терминологизации обыденных слов. Усилия должны быть направлены на дальнейшее расширение этого способа. Перенос значения по внешнему сходству, т.е. какой-либо сходный признак денотата, становится достаточным для обозначения другого предмета. Например, слово *тана* 1) перламутр // перламутровый; 2) кнопка (напр., *электрического звонка*) [5]. На основе внешнего сходства может образоваться новое терминологическое значение – *клавиша*. Здесь для переноса достаточно, чтобы какой-либо признак одного предмета совпадал или имел некоторое сходство с признаком другого, или один предмет внешне чем-то напоминал другой.

Перенос значения по аналогии понятий возникает из ассоциативной связи с каким-либо значением или оттенком значения в системе знаний, вследствие чего появляется новое терминологическое обозначение. Из сло-



ва *шол* ('прозвище, псевдоним, кличка') [6], на основании понятийной ассоциативной связи со значением '*псевдоним*' в компьютерной области возник термин *ник-нейм*, или *логин*, т.е. другое имя, используемое пользователем в Интернете: *Шолаңар кириңер* [7] – '*Введите ваш логин*'.

Значение по близости обозначаемой функции также может быть перспективным. Слово *балаар* ('вычёркивать, стирать') [8] в системе нового интерфейса на тувинском языке может функционировать по близости выполняемой функции как термин *удалить* (удалить текст или сообщение): *Медегелди балаар* [9] – '*Удалить сообщение*'. Здесь данное слово приобретает новое терминологическое значение на основе старого, которое за ним сохраняется.

Другим способом создания терминов является терминологизация архаизированных слов. В недавнем прошлом целые пласты тувинской лексики, связанные с социально-экономическими, религиозными, нравственными и другими историческими реалиями, считались архаизированными. Однако так называемые архаизированные слова не исчезают безвозвратно ещё и потому, что они потенциально существуют в тайниках материальной и духовной культуры народа и в благоприятных для них условиях могут в любое время возродиться вновь [10]. Таким образом, в компьютерном интерфейсе на тувинском языке эти слова также могут функционировать, например социально-экономический термин *шуудаң* [11] используется как обозначение почты в электронном виде: *Электроннуг шуудаңыңар адреси* [12] – '*Адрес Вашей электронной почты*'. В условиях широкого интереса к прошлому многие «устаревшие» термины могут «ожить» и приобрести специализированное употребление.

Общеизвестно, что парные слова, передаваемые графически через дефис, обозначают более общее отвлечённое понятие, чем значение их отдельных компонентов. В связи с этим потенциальные возможности терминологизации парных слов весьма широки. Однако, как показывает словарный материал, этот универсальный способ терминообразования практикуется недостаточно. Тем не менее таким способом разработаны следующие разряды терминообразующих парных слов: парные образования, компоненты которых имеют почти равнозначное употребление (типа *дериҕ-херексел* – 'устройство, оборудование' [13]), парные образования, где один из компонентов имеет более ограниченное употребление (типа *санал-онал* – 'комментарий, мнение' [14]), парные слова, состоящие из самостоятельно и несамостоятельного компонентов, т.е. один из элементов не имеет своего лексического значения или самостоятельно не употребляется (типа *амы-хуу* – 'личный' [15]).

Аффиксальный способ. Суффиксальное терминообразование имеет свои специфические особенности по сравнению с системой образования



общеобиходных слов. Терминология располагает более широким и разнообразным набором словообразовательных средств, чем общелитературное словообразование. Интересным примером служит слово *аайлалга*, образованное от глагола *аайлаар/аайла* 'приводить в порядок, наводить порядок' [16] присоединением аффикса *-лга*, образующее отвлечённые существительные от глагольных основ [17]. В интерфейсе на тувинском языке оно приобрело совершенно новое, ранее неизвестное терминологическое значение *настройка* – 'настройка личных данных в своей электронной почте, странице': *Ниипти аайлалга* [18] – 'Общие настройки'.

В тувинском языке продуктивные и непродуктивные аффиксальные средства всё ещё не получили чёткого разграничения. В терминообразовании особо выделяется ряд аффиксов, при помощи которых образуются слова-термины от именных и особенно от глагольных основ.

Синтагматический способ. Сложными терминами мы называем такие словосочетания, которые в результате слияния компонентов в единое целое превращаются в лексическую единицу назывного характера. Сложный термин представляет собой семантическое сращение, воплощающее в себе единство означаемого и означающего. Подобное словосочетание соотносится не со значением отдельных его компонентов, а с общим значением, которое чаще всего не выводится из значений компонентов. В тувинском компьютерном интерфейсе применены: сложный термин с аналитической глагольной конструкцией с 1-м компонентом деепричастия – *катап тургузар* ('восстановить'), сложный термин с отрицательной частицей «чок» – *айыыл чок чорук* ('безопасность') [19] и сложный термин, построенный как изафетная конструкция, – *четки богу* ('спам').

Особенностью тувинских сложных терминов является то, что они передаются на русском языке одним словом-термином. Русский однословный эквивалент сложного термина – наглядный показатель его лексикализации. Всё это говорит о семантическом слиянии компонентов в единое целое, в результате чего словосочетание превращается в лексическую единицу [20].

Создание терминов интерфейса – трудоёмкая и требующая постоянного анализа работа. На данный момент не существует единого стандарта для создания сайта на тувинском языке, и некоторые моменты вызывают спорные вопросы как у разработчиков, так и у обычных пользователей. При создании рассмотренного интерфейса применялись авторский подход, а также сравнительный анализ с другими тюркскими языками (татарский, якутский) и монгольским языком.

Создание национальных терминов – это творческий процесс, в ходе которого происходит трудоёмкая работа по подбору наиболее подходящих средств выражения того или иного научного или научно-технического по-



нения. Для так называемых младописьменных языков это, по существу, дело новаторское, требующее кроме тончайшего языкового чутья определённого мужества, энергии и любви к родному языку [21]. Сами слова-термины в большинстве своём – это не какие-то особенные слова. При ближайшем рассмотрении они оказываются обиходными словами, но употреблёнными в особой функции. Их особая функция заключается в одновременном обозначении наряду с обиходным значением какого-либо специально понятия. Иными словами, термин – это в основном общеупотребительное слово, применённое в какой-либо области знания для обозначения научного понятия.

* * *

1. Волченков Е.И. Стандартизация пользовательского интерфейса. – М., 2010. – С.41.
2. Оконешиников Е.И. Лингвистические аспекты терминологии языка саха. (На материале общей и отраслевой лексикографии). – Якутск, 2004. – С.76.
3. Щерба Л.В. Языковая система и речевая деятельность. – Л., 1974. – С.280.
4. Тенишев Э.Р. Тувинско-русский словарь. – Кызыл, 1968. – С.72.
5. Тенишев Э.Р. Указ. соч. – С.405.
6. Тенишев Э.Р. Указ. соч. – С.577.
7. URL: <https://www.vk.com> (дата обращения: 30.01.2017).
8. Тенишев Э.Р. Указ. соч. – С.88.
9. URL: <https://www.vk.com> (дата обращения: 30.01.2017).
10. Оконешиников Е.И. Указ. соч. – С.95.
11. Тенишев Э.Р. Указ. соч. – С.538.
12. URL: <https://www.vk.com> (дата обращения: 30.01.2017).
13. Тенишев Э.Р. Указ. соч. – С.159.
14. Тенишев Э.Р. Указ. соч. – С.367.
15. Тенишев Э.Р. Указ. соч. – С.159.
16. Тенишев Э.Р. Указ. соч. – С.27.
17. Исхаков Ф.Г., Пальмбах А.А. Грамматика тувинского языка: фонетика и морфология. – М., 1961. – С.168.
18. URL: <https://www.vk.com> (дата обращения: 30.01.2017).
19. Тенишев Э.Р. Указ. соч. – С.48.
20. Ойунский П.А. Вопросы терминообразования // Якутский язык: история и актуальные вопросы. – Якутск, 1986. – С.76.
21. Оконешиников Е.И. Указ. соч. – С.79.



Г.Л. Сабанцев,

*мл. научный сотрудник отдела
по инновационному развитию
марийского языка и культуры
ГБУК РМЭ «Республиканский
центр марийской культуры»
(Республика Марий Эл, г. Йошкар-Ола)*

А.В. Чемышев,

*зав. отделом по инновационному
развитию марийского языка и культуры
ГБУК РМЭ «Республиканский центр
марийской культуры»
(Республика Марий Эл, г. Йошкар-Ола)*

ЯНДЕКС.ПЕРЕВОДЧИК И ЯЗЫКИ РОССИИ

Перевод текста с одного языка на другой с помощью компьютерной программы, называемый машинным (автоматическим) переводом, в настоящее время получил широкое распространение, и в первую очередь, благодаря сервисам *Google Translate* и *Яндекс.Переводчик*.

Машинные переводчики могут быть реализованы на основе:

- правил;
- параллельных корпусов текстов;
- правил и параллельных корпусов текстов одновременно;
- морфологического, синтаксического и семантического разбора текстов.

Примером машинного переводчика, основанного на правилах, является система *Apertium*, разработанная при финансировании со стороны правительств Испании и Каталонии в Университете Аликанте (*Universitat d'Alacant*). Это свободное программное обеспечение, которое бесплатно издаётся разработчиками в соответствии с условиями *GNU GPL*. *Apertium* является системой машинного перевода, которая использует конечные преобразователи для всех своих лексических трансформаций, а также скрытые модели Маркова для выделения частей речи или устранения противоречий в категориях слов.

Google Translate (<https://translate.google.ru>) и *Яндекс.Переводчик* (<https://translate.yandex.ru>) – машинные переводчики, реализованные на основе параллельных корпусов текстов.



В последнее время «Яндекс» изменил алгоритм в своём машинном переводчике – дополнительно начал использовать правила.

Некоторыми компьютерными лингвистами ведутся работы по извлечению информации из массива неструктурированных текстов с использованием морфологического, синтаксического и семантического анализа. При этом машинные переводы будут являться «побочной работой» данных систем.

О Яндекс.Переводчике. *Яндекс.Переводчик* – это сервис автоматического перевода слов, фраз, целых текстов, а также веб-страниц. Сервис использует технологию машинного перевода, разработанную в «Яндексе».

На данный момент сервис предоставляет возможность перевода на 91 язык, в том числе на следующие языки России: башкирский, горномарийский, марийский, татарский, удмуртский.

Для перевода текстов достаточно выбрать язык исходного текста и язык, на который требуется осуществить перевод. Затем в поле слева следует ввести или вставить скопированный текст. Перевод выполнится автоматически и отобразится в поле справа.

Сервис автоматически определяет язык исходного текста. Достаточно начать набирать текст, и система сама определит и установит нужный язык в списке выбора языков.

Перевод производится автоматически по мере набора текста. Обновление перевода происходит, когда сервис определяет конец слова. Синхронный перевод обрабатывает текст длиной до 1 000 символов.

Опыт включения марийского языка в Яндекс.Переводчик. Тринадцатого июня 2016 г. при ГБУК РМЭ «Республиканский центр марийской культуры» был открыт Отдел инновационного развития марийского языка и культуры (далее – Отдел). Первой задачей нового отдела стало включение марийского языка в сервис Яндекс.Переводчик.

Организовано подписание договора между ООО «Яндекс» и Марийским научно-исследовательским институтом языка, литературы и истории им. В.М. Васильева на право использования базы 10-томного Марийско-русского словаря.

Осуществлён перевод так называемых «эталонных» текстов (510 строк) на марийский и горномарийский языки.

Описана грамматика марийского языка, в первую очередь морфология: http://wiki.fu-lab.ru/index.php/Формализованная_грамматика_марийского_языка

Началось создание корпуса параллельных текстов (на марийском и русском языках). Марийский национальный театр драмы им. Шкетана и Марийский республиканский театр-центр для детей и молодёжи предоста-



вили тексты около 50–60 пьес (в электронном виде) на марийском и русском языках, а газета «Марий Эл» – законы, переведённые на марийский язык. Также часть переведённой литературы, сказки, фольклор, загадки, пословицы и т.д. были отсканированы и распознаны при помощи программы *ABBYY FineReader*.

Для выравнивания параллельных текстов была использована программа *ABBYY Aligner*. Всем пользователям облачной среды для автоматизации перевода *SmartCAT* (<https://ru.smartcat.ai>) лицензия на *ABBYY Aligner 2.0* предоставляется бесплатно. Выровненные тексты экспортировались в файл *Translation Memory* (TMX-файлы) и передавались в отдел машинного перевода «Яндекс».

Шестого сентября 2016 г. «Яндекс» опубликовал, что они запустили машинный перевод для марийского и горномарийского языков. Таким образом, через 3 месяца интенсивной работы марийский язык появился в бета-версии сервиса *Яндекс.Переводчик*.

Как из бета-версии перейти в финальный релиз. При открытии *Яндекс.Переводчика* (<https://translate.yandex.ru>) можно заметить, что напротив башкирского, горномарийского, марийского и удмуртского языков стоит знак β – это означает, что машинные переводчики для данных языков находятся в бета-версии. При этом, например, напротив татарского языка такого знака нет: машинный переводчик для татарского языка за 1,5 года из бета-версии перешёл в финальный релиз.

Ведутся ли в «Яндекс» дальнейшие работы по удмуртскому языку, нам неизвестно. Скорее всего, нет, судя по интервью разработчика группы машинного перевода «Яндекс» А. Дворковича изданию «N+1» (<https://nplus1.ru/material/2016/12/28/small-languages-yandex>): «... при создании переводчика с удмуртского языка мы познакомились с группой энтузиастов, которые занимаются составлением его корпуса...».

Однако, наш Отдел с 6 сентября 2016 г. по настоящий день продолжает работу по улучшению марийского машинного переводчика: осуществляется сбор и выравнивание параллельных текстов, которые впоследствии передаются отделу машинного перевода «Яндекс».

Недавно в Отдел обратились башкирские активисты, недовольные качеством башкирского машинного переводчика в сервисе *Яндекс.Переводчик*, и попросили составить «инструкцию» с перечнем работ, направленных на улучшение работы башкирского машинного переводчика. Инструкция была составлена, волонтеры работают по сбору и выравниванию параллельных текстов (на башкирском и русском языках).

«Яндекс» ведёт работу по краудсорсу (<https://translate.yandex.ru/community>): скоро должны появиться «должности» модераторов, отвечающих за определённые языки.



В данный момент «Яндекс» продолжает работу по созданию машинных переводчиков для языков России. Полагаем, что в скором будущем мы увидим коми, коми-пермяцкий и другие языки в сервисе *Яндекс.Переводчик*. Но для этого одних усилий и желания разработчиков «Яндекс» недостаточно: в регионах работа должна опираться на центры компетенции (лаборатории, отделы и т.д.), работающие на профессиональной основе (3 лингвистов вполне достаточно). Всё время, как это зачастую происходит во многих регионах России, в вопросе сохранения и развития языков опираться на работу активистов, волонтеров, энтузиастов – некрасиво и неэтично!



В.И. Силин,

*д-р геогр. наук, гл. научный
сотрудник Института языка,
литературы и истории
Коми НЦ УрО РАН
(г. Сыктывкар)*

Л.И. Акулова,

*канд. биол. наук, доцент,
зав. кафедрой естественно-
научного образования
Сыктывкарского государственного
университета имени П.А. Сорокина*

**СОЗДАНИЕ ЭЛЕКТРОННЫХ СПРАВОЧНО-
БИБЛИОГРАФИЧЕСКИХ БАЗ ДАННЫХ
ПО ИСТОРИИ ГЕОГРАФИЧЕСКОГО ИЗУЧЕНИЯ
И КРАЕВЕДЕНИЮ РЕГИОНОВ
(на примере Республики Коми)**

Различные задачи историко-географических и краеведческих работ требуют от исследователя способности быстро ориентироваться в большом объеме опубликованной и фондовой литературы. Любое исследование должно начинаться со знакомства с библиографической литературой, а многие – и заканчиваться составлением разнообразных библиографических сводок.

В работе нами использовались различные виды библиографических справочников и указателей:

1. Дореволюционные универсальные и тематические.
2. Отраслевые: геологические, оленеводческие, ботанические и т.д.
3. Ведомственные: издания Коми НЦ УрО РАН, Сыктывкарского государственного университета имени П.А. Сорокина и др.
4. Персональные, посвященные деятельности отдельных учёных.
5. Посвященные памятным датам.

Исследователю необходимо как минимум знать основные справочники, в которых можно быстро и оперативно найти литературу по истории, естествознанию и современному состоянию природы и хозяйства родного края. При характеристике вопросов истории регионального естествознания (далее – ИРЕ) (экспедиции, жизнь исследователей, их вклад в естественные



науки) приходится пользоваться разнородными справочниками, хотя для Европейского Северо-Востока уже созданы первые специализированные справочно-библиографические пособия [1].

Краткий алгоритм работы историко-географа и краеведа с библиографическими указателями:

1. Постановка задачи (проблемы).
2. Знакомство с имеющимися в библиотеках указателями.
3. Выбор нужной литературы. Составление предварительного библиографического указателя.
4. Поиск литературы. Изучение литературы. Дополнение предварительного справочника и отсев ненужной информации.
5. Составление библиографического указателя по сформулированной проблеме.

Большое количество справочников, особенно дореволюционного издания, представляют библиографическую редкость и вместе с тем имеют научную ценность, поскольку иногда открывают неизвестные факты ИРЕ.

Первым своеобразным специализированным библиографическим справочником, в котором собрана различная литература по истории и экономике, в том числе Европейского Севера, являются многочисленные указатели в Географическо-статистическом словаре под редакцией известного русского путешественника и статистика П.П. Семёнова [2]. К сожалению, в этом словаре библиография, касающаяся определённого параграфа (названия губернии, города, села, реки и т.д.), дана в сокращённом виде, поэтому требуется значительная работа для составления полного списка использованной авторами словаря литературы. Положительной чертой этого издания является то, что в нём, помимо литературы, изданной в столице, приводится и региональная литература (издания губернских ведомостей и др.).

С 1838 г. в губерниях стали издаваться губернские ведомости (Архангельские, Вологодские, Пермские и др.), в «неофициальной части» которых печатались различные материалы, относящиеся к географии и истории Коми края. Ориентироваться в многочисленной литературе, печатающейся в Ведомостях, позволяют библиографические справочники [3]). Наиболее ёмкий и полный справочник, включающий литературу о Европейском Севере, в частности литературу о Коми крае, был предложен П.А. Дилакторским. Данный справочник был составлен в начале века, но опубликован лишь в 1921 г. [4]. Авторству П.А. Дилакторского принадлежит также интересный биографо-библиографический справочник «Вологжане-Писатели» [5], большая часть материалов которого впоследствии перекочевала в биографо-библиографический справочник «Вологжане-Краеведы» [6]. Из этих справочников можно получить интереснейшую информацию о краеведах дореволюционного времени и их работах.



Необходимо отметить ряд библиографических справочников дореволюционного времени, несущих информацию об отраслевых исследованиях, например справочник по ботанической литературе [7], справочник по Большеземельской тундре [8] и др. К тому же своеобразными сводками литературы являлись и являются многочисленные книги, посвящённые различным сторонам жизни и особенностям природы на Севере.

После революции и до Великой Отечественной войны справочных библиографических работ, посвящённых нашему краю, печаталось немного, они стали в основном отраслевыми, издавались для специалистов и были чаще всего разбросаны по разным изданиям (например, справочник Н. Терентьева по оленеводству [9] и др.). Необходимо отметить указатель литературы, посвящённый изученности геологии и полезных ископаемых Севера [10], в котором приводится характеристика литературы по геологии и минеральным ресурсам Севера европейской части России, изданной до 1940 года.

В послевоенный период количество указателей резко возросло, появились отраслевые (по ботанике [11], геологии [12] и др.), общие [13], библиотечные [14]. Наиболее известный и полный – справочник М.Г. Каракчиевой «Библиографический указатель литературы о Коми АССР». Этот справочник содержит ссылки на литературу, которые расположены по тематическим и временным разделам, поэтому в нём легко ориентироваться даже не очень подготовленному исследователю. К сожалению, справочник далеко не полный. Продолжением работы М. Каракчиевой являются библиографические ежегодные справочники «Литература о Республике Коми за... год». Эти справочники ценны тем, что помимо информации о книгах и статьях в журналах, они содержат также статьи из республиканских и районных газетах, а именно там содержится большой материал активно работающих краеведов. Целесообразно начинать исследования именно с просмотра этих справочников.

Краоведам и исследователям природы и хозяйства определённой территории необходимо иметь в виду, что многочисленные библиографические указатели издаются и в больших библиотеках. Например, библиотекой Коми научного центра издан ряд указателей литературы, опубликованной сотрудниками различных научных подразделений.

Начинающим исследователям полезно познакомиться с библиографическими указателями, посвящёнными определённым событиям или персоналиям. В таких указателях приводятся ссылки на статьи конкретного исследователя и литература о нём самом (например, в Коми НЦ УрО РАН выходит серия препринтов «Люди Науки»). С библиографией работ об отдельных исследователях, краоведах, исторических деятелях можно позна-



комиться в специализированных указателях [15]. Небольшими тиражами выпущены полезные и полные библиографические указатели в системе учебных пособий по спецкурсам. Так, значительная работа проведена в Сыктывкарском государственном университете имени П.А. Сорокина под руководством Л.П. Рощевской (например, указатели, посвящённые деятельности В.Ф. Попова [16] и др.).

Достаточно популярен, особенно среди начинающих свою работу историко-географов и краеведов, биографо-библиографический справочник «Исследователи Коми края» [17]. Большая работа проводится сотрудниками Национальной библиотеки Республики Коми по изданию учебно-справочных выпусков «Памятные даты на... год» и др. Эти справочники помогают учителям грамотно подготовиться к проведению различных мероприятий в школах. Подобные справочники издаются и в соседних регионах, например в Архангельске, Вологде, Перми, но из-за плохой связи между библиотеками они не всегда доступны.

Существует также большое количество видов библиографических указателей, имеющих самое непосредственное отношение к историко-географическим и краеведческим исследованиям. Несмотря на это, сегодняшнее состояние в этой области благоприятным назвать нельзя. Поэтому считаем целесообразным принятие следующих мер:

1. Необходимо подготовить и издать систематизированный справочник известных на сегодняшний день библиографических указателей, имеющих отношение к территории, хозяйству, культуре Коми края.

2. Желательно создать специализированные справочно-библиографические указатели на хронологической основе. Такие справочники, во-первых, внесут определённый вклад в изучение прошлого республики; во-вторых, они очень удобны для работы со школьниками.

3. Необходимо приступить к созданию электронных банков данных, относящихся к территории и истории края, в том числе к разработке алгоритма по составлению унифицированного справочно-библиографического указателя.

В данной статье нами представлена работа, над которой авторы-соавторы трудились много лет – это справочно-библиографический указатель по истории естественно-научного (географического) и краеведческого изучения территории Республики Коми, созданный на хронологической основе. Указатель выставлен на сайте Института языка, литературы и истории Коми НЦ УрО РАН по адресу: <http://illhkomisc.ru>

В 2006 г. в Коми государственном педагогическом институте (г. Сыктывкар) вышло подготовленное нами учебно-библиографическое пособие «История географических и краеведческих исследований Европейского



Северо-Востока России» (220 с.), в котором отражена библиография о событиях, происходивших на Европейском Севере с 870 года (путешествие норманнов) до 2004 г. (5 тыс. ссылок).

Работа в этом направлении была продолжена, и сейчас подготовлен электронный указатель литературы, опубликованной до 2016 года. В нём приводится более 8 тыс. ссылок, объём составляет около 800 страниц.

Структура электронного указателя включает в себя следующие блоки:

Блок 1. Свод материалов по географическому и краеведческому изучению территории Республики Коми. Информация приводится по хронологии событий. Например:

871 г.

Охтер – норвежец, дошёл морем до устья Северной Двины и нашёл там народ, говоривший на одном языке с финнами (страна Биармланд (Биармия). Вслед за Охтером в эту страну устремились и другие викинги. Главная цель их – пожить чужим добром. Эти походы отражены в исландских, датских и норвежских сагах.

Литература:

Андерсон Н.И. Об этимологии названий Биармия, Пермь, пермяки, зыряне // Труды десятого археологического съезда в Риге, 1896. – М., 1900. – Т.3. – С.125.

<...>

По многим событиям приводятся рисунки и фотографии.

Блок 2. Обобщающие работы по истории естественных региональных наук и истории освоения территории в послевоенный период.

Вавилов П.П. Коми учёные – республике // Красное знамя. – 1957. – 1 февр.

Вавилов П.П. Научная деятельность Коми филиала Академии наук СССР // Известия восточных филиалов Академии наук СССР. – 1957. – Вып.8. – С.117–125.

<...>

Блок 3. Рекомендуемая литература для изучения географического и краеведческого познания территории региона.

Авдеев А. Ижемские землепроходцы (XIX век: Заметки краеведа) // Красное знамя. – 1986. – 10 сент.

Авдеева Л.Р. Русские мыслители А.А. Григорьев, Н.Я. Данилевский, Н.Н. Страхов. – М., 1992.

<...>



Блок 4. Биографическая литература о современных исследователях.

Абрамова Н.А., Ильина Ф.Я. Выпускники Саратовского Университета // Библиографический словарь. Василий Иванович Чалышев, Лидия Михайловна Варюхина. – Саратов, 1989. – Вып. 1. – С.256–272.

Адуева В.А. Первооткрыватель «кладбища интинских крокодилов» (Г.А. Дмитриев) // Известия Общества изучения Коми края. – Сыктывкар, 2004. – № 3 (5). – С.38–42.

Блок 5. Библиографические указатели.

Алфавитный указатель литературного материала по Автономной области Коми. Составлен колонизационной экспедицией Коми н.-и. института. 1935. 66 л. // НА Коми НЦ УрО РАН. – Ф.1. Оп.11. Д.3.

Анатолий Александрович Куратов: библиогр. указ. лит. / Арханг. обл. науч. б-ка им. Н.А. Добролюбова, Помор. гос. ун-т им. М.В. Ломоносова; сост. Е.И. Тропичева. – Архангельск: ПГУ, 2001. – 47 с.

Проведя анализ электронного указателя, можно отметить следующее:

1. Наибольшее количество публикаций относятся к жизни и творчеству уважаемых в республике исследователей, выходцев из Коми края: К.Ф. Жаков, П.А. Сорокин. При этом в основном это публикации, созданные после 1990 года.

2. Немного публикаций, особенно в региональной печати, посвящено истории естественно-научного изучения периода XV–XVIII веков. Это объясняется малым количеством документов и отсутствием в те годы интереса к региону у центральных властей.

3. Жизнь и деятельность многих авторов, путешествующих по территории Европейского Севера и описывающих её в региональной прессе, практически не освещена. Достаточно большое количество литературы посвящено отдельным интересным исследователям, порой не много сделавшим для науки и республики.

4. Многие авторы повторяются в своих публикациях, но в этом есть и положительный момент – просвещение читателя и популяризация научных исследований.

Составитель считает свою работу нужной, надеется на то, что она поможет многочисленным исследователям-естественникам, историкам, краеведам, учителям, школьникам, и рассчитывает на помощь со стороны читателей. Большой массив приведённой литературы, возможность поиска по годам, именам, географическим объектам делает такой указатель необходимым в учебной, музейной, библиотечной работе. В перспективе составление подобных библиографий по другим регионам могло бы способствовать созданию единой электронной базы для изучения всей страны.



* * *

1. *Силин В.И.* История региональных географических исследований: Республика Коми: справ.-библ. учеб. пособие. – Сыктывкар, 1996. – Ч.1; *Силин В.И.* История региональных географических исследований: Республика Коми (1920–1944 гг.): справ.-библ. учеб. пособие. – Сыктывкар, 1998. – Ч.2; *Силин В.И.* История региональных географических исследований: Республика Коми (1944–1995 гг.): справ.-библ. учеб. пособие. – Сыктывкар, 1998. – Ч.3.
2. *Силин В.И.* Географический словарь П.П. Семёнова – краеведческий материал для учителя средней школы (на примере территории Республики Коми) // Очерки по истории географических исследований на территории Коми края. – Сыктывкар, 1998. – Ч.3. – С.103–137.
3. *Чубинский П.* Список статей, помещённых в Арх. губ. вед. по части истории, географии, этнографии и статистики губернии с начала издания их в 1838 г. по 1 янв. 1865 г. – Архангельск, 1865; Арх. губ. вед. – 1865. – Вып. 9–10; *Степановский И.К.* Вологодские губернские ведомости в период 50-летнего их существования с 1838 по 1888 г. – Вологда, 1888.
4. *Дилакторский П.А.* Опыт указателя литературы по Северному Краю с 1766 по 1904 г. – Вологда, 1921.
5. *Дилакторский П.А.* Вологжане-писатели. (Материалы для словаря писателей-уроженцев Вологодской губернии). – Вологда, 1900.
6. *Веселовские А. и А.* Вологжане-краеведы. – Вологда, 1923.
7. *Поле Р.Р.* Указатель лесоводственной и ботанико-географической литературы Северной России и Финляндии. 268 с. Петроград, 1915. Отт. из «Материалов по изучению русского леса». – 1915. – Вып. 2.
8. *Шидловский А.Ф.* Библиографический указатель литературы, касающейся Большеземельской тундры и оленеводства по материалам собранным АОИРС. – Архангельск, 1911.
9. *Терентьев Н.* Указатель литературы по оленеводству // Коми му. – 1926. – Вып. 1–2. С.97–98; Вып. 3. – С.52–54; Вып. 4. – С.42–47; Вып. 5. – С.51.
10. *Смирнов В.И.* Материалы о библиографии геологии и полезных ископаемых Северного края (кроме полярных островов). – Архангельск, 1940.
11. *Болотова В.М.* Флора и растительность Коми АССР (Библиографический указатель. Ч.1: Книги и статьи за 1941–1950 гг.). – Сыктывкар; М., 1956.



12. Геологическая изученность Коми АССР: в 5 т. Т.5. – М.; Сыктывкар, 1962–1967. – Вып. 1–5.

13. *Каракчиева М.Г.* Библиографический указатель литературы о Коми АССР. – Сыктывкар, 1963.

14. Каталог редких изданий краеведческой и национальной литературы (Национальная библиотека Республики Коми. Отдел краеведческой и национальной литературы). – Сыктывкар, 1994.

15. Святой Стефан Епископ Пермский (около 1330–1396 гг.): библиогр. указатель / сост. И.И. Табаленкова. – Сыктывкар, 1996.

16. Культурные связи русских и коми на Европейском Севере в XVII – начале XX вв.: метод. указания к спецкурсу и библиографический список трудов В.Ф. Попова для студентов IV курса исторического факультета ОЗО / сост. Р.Л. Попова, Л.П. Рощевская. – Сыктывкар, 1991.

17. Исследователи Коми края (Рекомендательный указатель литературы). – Сыктывкар, 1984.



З.А. Сиразитдинов,

*канд. филол. наук, ст. научный сотрудник,
зав. лабораторией лингвистики
и информационных технологий
Института истории, языка и литературы
Уфимского научного центра РАН
(Республика Башкортостан, г. Уфа)*

Л.А. Бускунбаева,

*канд. филол. наук, научный сотрудник
лаборатории лингвистики
и информационных технологий
Института истории, языка и литературы
Уфимского научного центра РАН
(Республика Башкортостан, г. Уфа)*

Г.Г. Шамсутдинова,

*лаборант-исследователь лаборатории
лингвистики и информационных технологий
Института истории, языка и литературы
Уфимского научного центра РАН
(Республика Башкортостан, г. Уфа)*

СТРУКТУРА СЛОВАРЕЙ В ЛЕКСИКОГРАФИЧЕСКОМ ПОДФОНДЕ МАШИННОГО ФОНДА БАШКИРСКОГО ЯЗЫКА

Создание электронных вариантов лингвистических ресурсов, таких как компьютерные словари, фонетические, грамматические и текстологические базы данных, размещённых в Интернете, не только расширяет сферу функционирования национального языка, но и открывает широкие перспективы для специалистов-филологов в плане сравнительных и сопоставительных исследований лингвистических реалий языков.

В последнем случае есть и сдерживающие факторы в использовании информационных ресурсов. Таковыми являются:

- 1) объём филологической информации;
- 2) достоверность лингвистических данных.

Эти факторы касаются в основном информационных данных Википедии, коммерческих организаций и частных лиц. Но такие проблемы не возникают для лингвистических ресурсов, созданных по печатным



первоисточникам в академических и научно-педагогических учреждениях, в структурах министерств образования и культуры и других органов власти, имеющих на начальной странице наименование создателя ресурса. Среди них особо выделяются комплексные лингвистические ресурсы, включающие в себя множественные базы данных [1]. Такие лингвистические ресурсы имеют большой и разнообразный филологический объём, их лексикографические базы дают «возможность поиска в десятках и сотнях словарей одновременно» [2]. Таким серьёзным ресурсом по башкирскому языку является лексикографический подфонд Машинного фонда башкирского языка (далее – МФБЯ), разработанный в лаборатории лингвистики и информационных технологий Института истории, языка и литературы Уфимского научного центра РАН [3].

Лексикографические базы данных МФБЯ построены на основе изданной печатной продукции с указанием авторов и года издания, за достоверность представленных данных лаборатория несёт ответственность. Базы данных этого подфонда сегодня состоят из 62 структурированных словарей общим объёмом словарных статей порядка 500 000 единиц.

В лексикографическом подфонде представлены академические и учебные словари: одноязычные и двуязычные, частотные, терминологические, фразеологические, синонимические, словари-справочники (названий населённых пунктов, улиц городов, водных и горных объектов Республики Башкортостан и др.). Все словари сгруппированы в 6 разделов:

- общефилологические двуязычные с башкирского языка;
- общефилологические двуязычные на башкирский язык;
- двуязычные терминологические словари;
- частотные словари;
- словари-справочники;
- одноязычные словари.

Отметим, что структуры имеющихся печатных словарей башкирского языка не являются одинаковыми, а стремление создать лексикографические базы, представляющие максимальную филологическую информацию и отражающие авторское представление (печатных вариантов словарей), не позволило создателям фонда унифицировать и представить единую структуру словарных данных.

Анализ структуры существующих словарей башкирского языка показывает следующее:

1. Двуязычные общие словари (башкирско-русский, русско-башкирский) имеют следующую структуру статей:

- 1) словарное слово;
- 2) показатель омонимии;



- 3) показатель многозначности;
- 4) дополнительная информация 1;
- 5) дополнительная информация 2;
- 6) дополнительная информация 3;
- 7) дополнительная информация 4;
- 8) перевод;
- 9) дополнительная информация на русском языке;
- 10) примеры употребления;
- 11) фразеологизмы.

II. Двухязычные терминологические словари.

1. *Двухязычные обычные терминологические словари* имеют следующую структуру словарной статьи:

- 1) словарное слово;
- 2) показатель омонимии;
- 3) показатель многозначности;
- 4) специальная помета;
- 5) перевод;
- 6) иллюстративный материал.

В словарях данного типа специальная помета и иллюстративный материал могут отсутствовать.

2. *Двухязычные толковые терминологические словари* имеют следующую структуру словарной статьи:

- 1) словарное слово;
- 2) показатель омонимии;
- 3) показатель многозначности;
- 4) специальная помета;
- 5) перевод;
- 6) толкование;
- 7) иллюстративный материал.

Специальная помета и иллюстративный материал также могут отсутствовать.

III. Двухязычные словари-справочники (словари названий населённых пунктов, водных и горных объектов Республики Башкортостан) имеют единую структуру:

- 1) словарное слово;
- 2) перевод;
- 3) вид объекта;
- 4) территория расположения (протекания);
- 5) этимология.



Пример представления словаря водных объектов:

Абалай	Абалай	рч	Бурз.	От антр. Абалай.
Абалак	Абалак	рч	Байм.	От диал. эбэлэк/абалак.
Абасай	Әбәсәй	оз	Чишм.	Возможно, от диал. эбәсәй

IV. **Частотные словари** имеют единую структуру данных:

- 1) словарное слово;
- 2) указатель части речи;
- 3) указатель омонимии;
- 4) частота употребления.

V. **Одноязычные словари.**

1. *Орфографический словарь башкирского языка* имеет следующую структуру словарной статьи:

- 1) словарное слово;
- 2) показатель омонимии;
- 3) показатель части речи;
- 4) дополнительная информация;
- 5) аффиксы словоизменения.

Например:

ағасак (ис.) (балык) -тар; -ғы(на).

2. *Толковый словарь башкирского литературного языка* имеет следующую структуру словарной статьи:

- 1) словарное слово;
- 2) показатель омонимии;
- 3) показатель части речи;
- 4) показатель многозначности;
- 5) толкование;
- 6) иллюстративный материал;
- 7) иллюстрация в виде фразеологизмов, пословиц и поговорок.

Приведённый анализ структур башкирских словарей показывает, что несмотря на различия все имеющиеся словари объединяет одно свойство, которое было подмечено ещё казахскими исследователями при создании информационной системы «Til – казына»: наличие опорного (реестрового) слова и наличие пояснительной (словарной) статьи (толкование значения слова), являющихся основным свойством любых словарей [4]. Это позволяет выделить обобщённую словарную структуру, которая в каждом конкретном случае будет представлена наличием или отсутствием каких-либо частей (информационных полей). Такая словарная структура, состоящая из 11 информационных полей, выглядит следующим образом:



- 1) заглавное слово;
- 2) омонимия;
- 3) многозначность;
- 4) дополнительная информация 1;
- 5) дополнительная информация 2;
- 6) перевод;
- 7) дополнительная информация 3;
- 8) толкование;
- 9) дополнительная информация 4;
- 10) дополнительная информация 5;
- 11) дополнительная информация 6.

Словари лексикографического подфонда открывают большие возможности для научных исследований. Мы надеемся, что они будут активно использоваться языковедами в сравнительных, сопоставительных и сравнительно-исторических исследованиях.

* * *

1. Словари русского языка. – URL: <http://gramota.ru> (дата обращения: 17.01.2017); Словари ABBYY LINGVO. – URL: <https://www.lingvolive.com/ru-ru> (дата обращения: 17.01.2017); Онлайн-словари FU-Lab (финно-угорских языков). – URL: <http://dict.fu-lab.ru> (дата обращения: 17.01.2017).

2. Федина М.С., Чемышев А.В., Степанов В.А. Финно-угорские языки в электронном информационном пространстве. – Сыктывкар: ГОУ ВО КРАГСиУ, 2015.

3. Машинный фонд башкирского языка. – URL: <http://mfbl2.ru> (дата обращения: 17.01.2017); Сиразитдинов З.А., Бускунбаева Л.А., Ишмухаметова А.Ш., Ибрагимова А.Д. Информационные системы и базы данных башкирского языка. – Уфа: Книжная палата РБ, 2013.

4. Жұбанов А.К. База данных «Тіл – қазына» казахского слова и её теоретические основы. – Алматы: Арыс, 2009. – С.213.



В.Н. Соловар,

д-р филол. наук, доцент,

зав. отделом хантыйской филологии

и фольклористики Обско-угорского

института прикладных исследований

и разработок (г. Ханты-Мансийск)

АКТУАЛЬНЫЕ ВОПРОСЫ СОВЕРШЕНСТВОВАНИЯ ПИСЬМЕННОСТИ ХАНТЫЙСКОГО ЯЗЫКА

Дискуссии по поводу проблем хантыйского алфавита велись ещё на финно-угорском конгрессе в 1995 г. в Дебрецене. Уже тогда исследователям было ясно, что ситуация с алфавитом, орфографией, используемых в учебниках хантыйского языка, газете и издаваемых книгах, требует анализа и изменения. В 1997 г. научные сотрудники Института филологии Сибирского отделения РАН (г. Новосибирск) предложили новый алфавит из 28 букв для основного диалекта, казымского, который распространён на большей части территории Ханты-Мансийского автономного округа – Югра (далее – ХМАО – Югра). Этот алфавит отражает фонемный состав казымского диалекта и не имеет лишних знаков (ранее использовался алфавит, в котором было более 40 букв). С этого времени в округе периодически собиралась национальная интеллигенция для обсуждения сложившейся ситуации, проводились семинары, и члены инициативной группы по разработке графики и орфографии хантыйского языка проводили рабочие совещания для обсуждения путей совершенствования алфавита. Был проведён анализ учебников хантыйского языка, анализ алфавита, используемого в единственной в округе газете, издаваемой на хантыйском языке. Рабочая группа столкнулась с сильным сопротивлением журналистов этой газеты, которые не хотели признать тот очевидный факт, что у них свой алфавит, свои неписанные правила. Национальная интеллигенция неоднократно обращала внимание на необходимость соблюдения принципов единообразного написания хантыйских слов в СМИ (открытое письмо канд. пед. наук Е.А. Немисовой опубликовано на сайте Обско-угорского института прикладных исследований и разработок <http://www.xn--hl1aaavgig.xn--p1ai> в разделе «Обсуждаем проблемы хантыйской письменности»). Фактически одна и та же фонема в газете передаётся 6 разными буквами. Изменять свою позицию газета не намерена и объясняет это тем, что при изменении алфа-



вита и разработке правил орфографии она потеряет своих читателей (хотя некоторые читатели выражали неудовольствие из-за отсутствия специальных знаков для хантыйского языка уже в начале 70 гг. XX в.). Этот единственный аргумент против нового алфавита активно поддерживается представителями власти – теми, кто считает себя хантами, но уже не пользуется языком этноса, не говорит и не пишет на этом языке. Не существует единых норм и правил, которыми могли бы руководствоваться хантыйские учителя и журналисты. Поэтому в каждом учебнике хантыйского языка разные авторы одни и те же слова писали по-разному, что создаёт трудности в усвоении учащимися хантыйского языка в школе.

На протяжении ряда лет ведётся планомерная работа по нормированию хантыйской письменности. В 2001 г. был проведён международный научно-практический семинар «Графика и орфография хантыйского языка» (г. Белоярский); в 2005 г. – окружное совещание «Актуальные проблемы графики и орфографии хантыйского языка» (г. Ханты-Мансийск). В период с 2005 по 2008 г. свою деятельность осуществляла комиссия по совершенствованию письменности хантыйского языка при Югорском государственном университете (председатель – д-р филол. наук Ю.В. Андуганов, сопредседатели – канд. филол. наук В.Н. Соловар и канд. филол. наук З.С. Рябчикова).

На существующие проблемы письменности хантыйского языка было обращено внимание Научно-координационного Совета (25 января 2012 г., 25 декабря 2013 г.). С 2012 г. при Департаменте образования и молодёжной политики ХМАО – Югры (Приказ Департамента от 05.09.2012) образована Рабочая группа по вопросам совершенствования графики и орфографии хантыйского языка (далее – Рабочая группа), которая провела 12 заседаний.

Было организовано три обучающих семинара: «Пути совершенствования графики и орфографии хантыйского языка» (2–7 сентября 2013 г., Институт развития образования), «Новая орфография хантыйского языка» (16–22 марта 2014 г., Институт развития образования), «Развитие письменности хантыйского языка» (27–29 мая 2014 г., Обско-угорский институт прикладных исследований и разработок), проведены курсы повышения квалификации учителей хантыйского языка (сентябрь 2016 г.), издан экспериментальный «Орфографический словарь хантыйского языка (казымский диалект)» В.Н. Соловар, подготовлен и апробирован на семинарах справочник «Правила хантыйской орфографии». В журнале «Вестник угроведения» опубликована статья Н.Б. Кошкаревой «Актуальные вопросы совершенствования хантыйской графики и орфографии» [1], в которой излагается история хантыйской письменности и формулируются принципы идеальных графики и орфографии. На сайте Обско-угорского института



прикладных исследований и разработок открыт раздел «Обсуждаем проблемы хантыйской письменности», в котором обсуждаются пути и перспективы развития хантыйской графики и орфографии. Получено 12 отзывов от ведущих специалистов в области изучения хантыйского языка, которые поддерживают деятельность Рабочей группы, подготовлено 10 книг на хантыйском языке, ожидающих издания.

В результате деятельности Рабочей группы с привлечением самой широкой общественности было принято решение о необходимости унифицировать письменность на хантыйском языке в соответствии с фонематическим принципом, признанным в теории языкознания оптимальным, т.к. он основывается на объективных законах языка, а не на субъективных факторах («мне это не нравится», «я так не хочу и делать этого не буду») и не на внешних обстоятельствах («это не своевременно»; в эпоху языкового строительства в СССР первоначальный выбор латиницы был обоснован ожиданием мировой революции; с переходом на кириллицу в 50-е гг. отсутствие дополнительных знаков, необходимых для отражения специфических звуков хантыйского языка, мотивировалось отсутствием в типографии хантыйской газеты соответствующих литер).

На протяжении 80 лет развития письменности на хантыйском языке было стихийно проведено 5 реформ. Последний вариант алфавита от 25 апреля 1990 г. ни в одном из существующих изданий не используется в полном объёме и последовательно. На его базе создано столько индивидуальных вариантов, сколько выходило публикаций на хантыйском языке. Создатель этого алфавита канд. пед. наук Е.А. Немысова, проанализировав его функционирование и причины широкого варьирования, предлагает использовать фонематический принцип как наиболее адекватно отражающий объективные законы хантыйского языка [2].

Графика и орфография на основе фонематического принципа будут долговечными, т.к. отражают языковую систему, которая меняется медленнее, чем модные политические веяния.

Таким образом, проведена большая теоретико-методологическая работа по совершенствованию письменности хантыйского языка с учётом новейших научных данных, полученных в том числе и на основе инструментальных методов (исследование звучащей речи на специальных приборах). На основе казымского диалекта хантыйского языка возможность использования усовершенствованной письменности поддерживается подготовленными орфографическими словарями и справочниками, а также её апробацией в уже изданных собраниях фольклорных текстов, учебной и научной литературе. Разработанные принципы применимы для всех диалектов хантыйского языка, т.к. отражают единые для языка нормы, и могут



быть рекомендованы для всех хантыйских диалектов (они успешно апробированы для шурышкарского, сургутского и ваховского диалектов во время семинаров).

Эмоциональное противодействие газеты и части общественности предсказуемо и естественно, т.к. любые нововведения принимаются с опаской, не все готовы затрачивать усилия на элементарное ознакомление с предложениями, отторгают их только в силу новизны. Проведённые семинары показывают, что всего за 2–3 дня желающие в полной мере овладевают грамотным письмом на хантыйском языке, т.к. правила просты и отражают законы хантыйского языка, которые интуитивно понятны каждому.

Продолжает издаваться газета, на страницах которой только изредка находится место для пропаганды грамотности.

Грамотное и единообразное написание на текущий момент представлено только в изданиях Обско-угорского института прикладных исследований и разработок, неграмотное в разных его формах продолжает тиражироваться за пределами этого института. Альтернативы фонематическому принципу нет, т.к. ему не противопоставлен никакой другой последовательный принцип.

Отсутствие нормированной письменности тормозит развитие хантыйского языка и культуры: невозможно подготовить и издать столь необходимые новые учебники, словари и грамматики академического типа, научные собрания фольклора. Известно, что стабильная письменная традиция консолидирует народ и способствует сохранению его культурных традиций.

Какой орган необходимо создать в Российской Федерации, чтобы утвердить усовершенствованный алфавит хантыйского языка? Почему до настоящего времени нет такого органа, а власти ХМАО – Югра не видят проблему в том, что целый народ стоит у порога безграмотности. После выступления заместителя губернатора на заседании научно-координационного совета от 25 декабря 2013 г. публикация книг на хантыйском языке на новом алфавите была приостановлена на целый год.

В школах округа в 2014 г. проведён опрос, в результате которого выявлено, что каждый учитель, преподающий родной хантыйский язык, имеет свой алфавит. В 2015 г. Департамент образования ХМАО – Югра наконец принял аргументы в пользу нового алфавита и издал приказ о проведении апробации учебников для 1–4 классов, изданных на новом алфавите ещё в 2008 г., однако успевших за этот период устареть. Работа учителей с 2015 г. по новому алфавиту в трёх школах округа (Казымской СОШ, Ванзеватской СОШ, Полноватской СОШ) уже даёт основания утверждать, что для учащихся 1–4 классов не представляется сложным научиться читать и писать по новому, ведь по сравнению со старым алфавитом количество



букв уменьшилось с 45 до 29, а новых букв появилось всего лишь 4, обозначающих 3 среднеязычных звука и 1 редуцированный. Но именно эти отсутствовавшие ранее буквы отражают специфические особенности хантыйского языка, и они касаются всех четырёх его диалектов. Конечно, учебно-методический комплекс и учебники в будущем будут переработаны в соответствии с новыми стандартами и программами по родному языку. Требуется большого внимания и литература, создаваемая на хантыйском языке, она отсутствует, нет и новой переводной литературы.

Сложившаяся ситуация с хантыйской письменностью в ХМАО – Югре требует неотложных мер. Подобные проблемы наблюдаются ещё в двух диалектах хантыйского языка, которые преподаются в Сургутском районе и Нижневартовском районе (сургутский и ваховский диалект); шурышкарский диалект функционирует в Ямало-Ненецком автономном округе, однако и здесь имеются такие же проблемы с письменностью.

Конечно, после принятия алфавита требуется работа по созданию системы проверки орфографии, электронных словарей, лексических баз данных, национального корпуса языка и многого другого, что способствует сохранению и развитию языка и культуры.

* * *

1. Кошкарева Н.Б. Актуальные вопросы совершенствования хантыйской графики и орфографии // Вестник угроведения. – 2013. – № 3. – С.47–78.

2. Немысова Е.А. Хантыйская орфография в газете «Ханты ясанг». – URL: https://ouipiiir.ru/sites/default/files/docs/e.a._nemysova_hantyyskaya_orfografiya_v_gazete_hanty_yasang.doc



А. Сорокина,

*студентка школы лингвистики
Национального исследовательского
университета «Высшая школа
экономики» (г. Нижний Новгород)*

СИСТЕМА ДЛЯ СРАВНИТЕЛЬНО-ИСТОРИЧЕСКОГО КОРПУСНОГО АНАЛИЗА ПЕРМСКИХ ЯЗЫКОВ

Электронные корпуса – незаменимый инструмент для лингвистических исследований. Как правило, они удобны для определения сочетаемости лексем и уточнения их значения, проверки определённых грамматических явлений. Подробные и хронологически сбалансированные корпуса дают возможность исследовать историю развития значений и конструкций [1].

В то же время для многих языков народов России корпуса, созданные на базе одного языка, почти не предоставляют инструментов для проведения диахронических или сравнительных исследований. Одним из возможных решений этой проблемы могло бы стать создание корпуса, объединяющего несколько близкородственных языков. Такой ресурс позволил бы лингвистам исследовать развитие значения слов и отдельных морфологических показателей в близких языках, сравнивать механизмы заимствования русских слов и способы их освоения языком.

Цель нашей работы заключалась в создании на материале пермских языков (коми, удмуртского) системы, связывающей два корпуса и позволяющей искать в них: а) когнаты и этимологически родственные морфологические показатели; б) грамматические категории со сходными функциями и семантикой, но этимологически разными показателями.

Для осуществления одновременного поиска по текстам двух языков необходимо, чтобы в двух корпусах этимологически близкие лексемы и морфемы были размечены одинаково. В качестве образца была взята морфологическая разметка, используемая в корпусе литературного удмуртского языка, объём около 7,3 млн словоупотреблений, омонимия не снята (<http://web-corpora.net/UdmurtCorpus>).

Таким образом, первым этапом работы стало создание размеченного корпуса коми языка. Для этого требовалось решить следующие задачи:



1) создание формального описания морфологии коми языка на основе грамматик, присваивание грамматическим категориям меток, соответствующих меткам аналогичных категорий в корпусе литературного удмуртского языка;

2) создание грамматического словаря с указанием частей речи, словоизменительных типов;

3) разметка текстов с помощью системы автоматического морфологического анализа *UniParser* [2].

Второй этап работы был посвящён составлению словаря соответствий для когнатов коми и удмуртского языков. Для этого автоматически сравнивались два словаря (30 и 23 тыс. лексем соответственно), этимологическое родство определялось методами машинного обучения на основе нескольких параметров:

1) **расстояние Левенштейна**, определяющее степень орфографической схожести слов. Так, паре сравниваемых лексем присваивалось значение, равное сумме «весов» операций, необходимых для превращения одной лексемы в другую. При этом были описаны регулярные переходы одной фонемы в другую, и вес операции был тем меньше, чем регулярнее было соответствие перехода. Соответствующие правила перехода основывались на материале описания фонологических систем пермских языков [3]. Порог приемлемого для пары лексем расстояния Левенштейна определялся в зависимости от длины слова;

2) **значение семантической близости переводов**. С помощью сервиса дистрибутивно-семантических моделей *RusVectors* для русских переводов с коми и удмуртского языков вычислялось значение семантической близости (от -1 до 1). Использовалась модель, обученная на национальном корпусе русского языка;

3) **длины лексем**.

Модель, определяющая этимологическую близость пары слов с учётом указанных параметров, была создана с использованием метода логистической регрессии. Обучающая выборка составила 90 лексем, тестовая – 30, процент ошибок модели – 8%.

Лексемы, заимствованные из русского языка (значительная часть обоих словарей), анализировались отдельно.

Поиск в текстах коми и удмуртского языков может осуществляться по лемме, словоформе или переводу. При поиске по словоформе или переводу в обоих корпусах находятся все контексты, содержащие (в строке текста или перевода соответственно) слово, совпадающее со строкой запроса. При поиске по лемме в текстах выбранного языка находятся контексты, содержащие слово из запроса, в текстах другого языка – содержащие его когнат, совпадающий или не совпадающий по форме (если он есть).



В целях развития системы прежде всего планируется увеличение количества текстов и расширение грамматических словарей. Отдельным направлением работы является улучшение поиска. Так, в настоящее время поиск по словоформе работает без учёта морфологических соответствий, т.е. одинаковые слова разных языков могут быть не найдены не из-за несоответствия лексем, а по причине, например, несоответствия падежных показателей (возможно, также этимологически связанных). Для решения этой проблемы, планируется составить нечто вроде морфологического словаря соответствий и применять его способом, схожим с применением лексического словаря.

Важным дополнением была бы возможность размещения на странице корпуса статистической информации, например подсчёта количества употреблений лексики в определённом контексте для коми и удмуртского языков, списков частотных словоформ и др.

* * *

1. *Плунгян В.А.* Корпус как инструмент и как идеология: о некоторых уроках современной корпусной лингвистики (Русский язык в научном освещении). – М., 2008. – № 2 (16). – С.15.

2. *Arkhangelskiy T., Belyaev O., Vydrin A.* The creation of large-scaled annotated corpora of minority languages using UniParser and the EANC platform // *Proceedings of COLING 2012: Posters.* – Mumbai: The COLING 2012 Organizing Committee, 2012. – Ch.9. – P.83–91.

3. *Тепляшина Т.И., Лыткин В.И.* Пермские языки // *Основы финно-угорского языкознания: Марийский, пермские и угорские языки.* – М.: Наука, 1976. – С.97–228.



Н.Н. Тимерханова,

*канд. филол. наук, доцент,
зав. кафедрой лингвистической
типологии и лингводидактики
Удмуртского государственного
университета (г. Ижевск)*

«ПО РОДНЫМ ПРОСТОРАМ» С РОДНЫМ ЯЗЫКОМ

В течение 2012–2017 гг. для воспитанников детских садов и школ коллективом авторов кафедры общей и физической географии института естественных наук и кафедры лингвистической типологии и лингводидактики института удмуртской филологии, финно-угроведения и журналистики Удмуртского государственного университета был подготовлен ряд пособий на удмуртском языке. Все пособия являются электронными учебными изданиями.

По географии были изданы:

- «География: Пичи нылпиослы энциклопедия» («География: Энциклопедия для маленьких детей». Ижевск, 2013 [1] (см. *рис. 1*));

- «Удмуртилэн ландшафтъёсыз: Туспуктэмъёслэз комментариосын альбом» («Ландшафты Удмуртии: Фотоальбом с комментариями». Ижевск, 2013 [2] (см. *рис. 2*));

- «Географиен олимпиада юанъёс» («Вопросы олимпиады по географии». Хельсинки, 2014 [3] (см. *рис. 3*));

- «Удмурт Элькунлэн географияез: шоръёзо дышетскон юртъёслы электрон дышетскон пособие» («География Удмуртской Республики: электронное учебное пособие для общеобразовательных учебных заведений». Хельсинки, 2014 [4] (см. *рис. 4*));

- «Удмурт Элькунлэн географияезъя ужъёсын дышетскон тетрадь: дышетскон юрттэт» («Тренажёр-практикум по географии Удмуртской Республики: учеб. пособие». Хельсинки, 2014 [5] (см. *рис. 5*));

- «Удмуртиысь туризм: этнотуризм, лулчеберетэн тодматскон, эмъяськон-шутэтскон но гуртысь туризм» («Туризм в Удмуртии: Этнотуризм, сельский, лечебно-оздоровительный и культурно-познавательный туризм в Удмуртской Республике». Хельсинки, 2016. URL: <https://sites.google.com/site/turizmuvudmurjrespublike> [6] (см. *рис. 6*)).



На данный момент готовится электронное учебное пособие к учебнику «Окружающий мир» (начальные классы) «По родным просторам. Удмуртия – край мой родниковый» с региональным содержанием для школ Удмуртской Республики.

Также вышли в свет по удмуртскому языку следующие пособия:

- «Удмурт кыл (Удмуртский язык): учебник-тетрадь для начинающих № 1». Ижевск, 2012. – В электронном формате с аудиоматериалами, упражнениями в игровой форме издано в 2015 г. [7] (см. *рис. 7*);

- «Удмурт кылын шонергожъяськонъя ужъёс: 9 презентацилэсь сузьет» («Упражнения по орфографии удмуртского языка: сб. из 9 презентаций». Хельсинки, 2016 [8] (см. *рис. 8*)). Эти материалы выложены на сайте, и наряду с ними предложены другие упражнения по орфографии и пунктуации удмуртского языка, в том числе созданные в Интернете (<https://sites.google.com/site/udmurtkylsykys/sutetskon/uzees>).

Кроме этого, для изучения удмуртского языка были подготовлены и размещены следующие материалы:

- дистанционный курс удмуртского языка с комментариями на русском языке (<http://e-learning.udsu.ru/course/view.php?id=32>);

- дистанционный курс удмуртского языка с комментариями на финском языке (<http://e-learning.udsu.ru/course/view.php?id=44>);

- дистанционный курс удмуртского языка с комментариями на венгерском языке (<https://sites.google.com/site/udmurtkylmadareesly>).

Для работы над развитием логического мышления детей 6–7 лет были подготовлены электронные учебные пособия:

- «Йырвизез сынан: Анимациен шудонъёс» («Разминка для ума: Анимированные игры». Хельсинки, 2014 [9] (см. *рис. 9*));

- «Визьбугор: Анимациен шудонъёс» («Умница: Анимированные игры». Хельсинки, 2014 [10] (см. *рис. 10*));

- «Визь куды: Анимациен шудонъёс» («Золотая голова: Анимированные игры», Будапешт–Бадачоньтёрдэмиц, 2014 [11] (см. *рис. 11*)). Данное электронное пособие было переведено на коми и марийский языки – «Вежӧр куд» (Сыктывкар–Будапешт–Бадачоньтёрдэмиц, 2014) [12] и «Шӧртньӧ вуй» (Йошкар-Ола–Будапешт–Бадачоньтёрдэмиц, 2014) [13].

Многие из названных электронных пособий созданы в программе *Microsoft Power Point*, т.к. работа с использованием данной программы интересна не только авторам, но и читателям, а применение её на занятиях приводит к ряду положительных эффектов:

- происходит визуализация материала, т.е. электронное пособие обогащает занятие наглядностью;

- электронный формат психологически облегчает процесс усвоения;

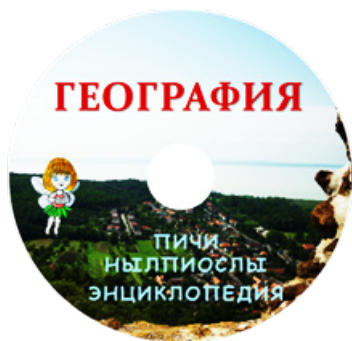


Рис. 1. «География: Пичи нылпиослы энциклопедия» («География: Энциклопедия для маленьких детей»)



Рис. 2. «Удмуртилэн ландшафтъясыз: Туспуктэмъяслэс комментарий» («Ландшафты Удмуртии: Фотоальбом с комментариями»)

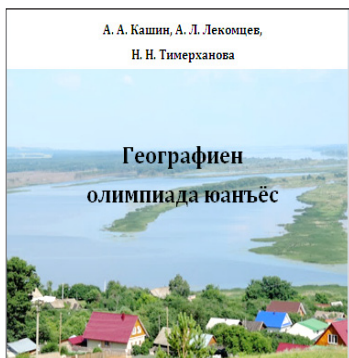


Рис. 3. «Географиен олимпиада юанъёс» («Вопросы олимпиады по географии»)



Рис. 4. «Удмурт Элькунлэн географияез: шоръёзо дышетскон юртъёслы электрон дышетскон пособие» («География Удмуртской Республики: электронное учеб. пособие для общеобразовательных учебных заведений»)



Рис. 5. «Удмурт Элькунлэн географияезъя ужъясын дышетскон юрттэт» («Тренажёр-практикум по географии Удмуртской Республики: учеб. пособие»)

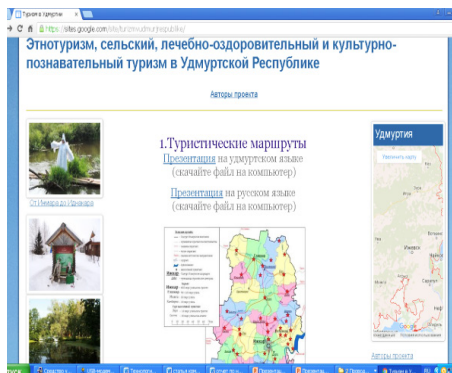


Рис. 6. «Удмуртиысь туризм: этнотуризм, лулчеберетэн тодматскон, эмъяськон-шугэтскон но гуртысь туризм» («Туризм в Удмуртии: Этнотуризм, сельский, лечебно-оздоровительный и культурно-познавательный туризм в Удмуртской Республике»)



Рис. 7. «Удмурт кыл (Удмуртский язык): учебник-тетрадь для начинающих № 1»



Рис. 8. «Удмурт кылын шонер-гожьяськонъя ужъёс: 9 презентацилэсь сузет» («Упражнения по орфографии удмуртского языка: сб. из 9 презентаций»)



Рис. 9. «Йырвизез сынан:
Анимациен шудонъёс»
(«Разминка для ума:
Анимированные игры»)



Рис. 10. «Визьбугор:
Анимациен шудонъёс»
(«Умница:
Анимированные игры»)



Рис. 11. «Визь куды: Анимациен
шудонъёс» («Золотая голова:
Анимированные игры»)

- у обучающихся пробуждается живой интерес к предмету познания;
- расширяется общий кругозор обучающихся;
- повышается производительность труда учителя и учащихся на занятии.

Кроме того, электронное пособие обладает дополнительными свойствами и возможностями:

1) имеет удобную систему навигации на основе гиперссылок, которая предоставляет возможность быстрого перехода от одного объекта изучения к другому;

2) содержит аудио- и видеофрагменты, включение которых в бумажный аналог пособия невозможно;



3) материалы пособия можно использовать в индивидуальном режиме или демонстрировать на занятиях и работать с ними в группе.

Применение на занятиях вышеназванных электронных пособий и материалов целесообразно для решения специальных практических задач:

- предъявления новой информации, изучения нового материала;
- закрепления пройденного, отработки универсальных учебных действий;
- повторения и практического применения универсальных учебных действий;
- обобщения, систематизации знаний;
- организации самостоятельной работы учащихся по формированию основополагающих знаний школьного курса, коррекции и учёта знаний учащихся.

Мы считаем, что учащимся будут интересны не только знакомство с теоретическими материалами вышеназванных электронных изданий, но и работа с тренажёром-практикумом, который позволяет отрабатывать темы, изученные на уроках, а также работа с контролирующими тестами и заданиями, которые представлены в разных красочных и интересных формах. Ко всему прочему каждый ученик с данными учебными материалами может работать в индивидуальном темпе и по индивидуальной программе.

В данных электронных пособиях используются различные приёмы медиадидактики и технологические приёмы, что делает подачу материала более разнообразной и интересной, например: анимированные кроссворды, сканворды, филворды, приёмы сорбонка, экран, листание, паззлы, всплывающие окна, лупа, соты, мультимедиапроектор, волшебный конверт, шаблон *DragAndDrop*, красочные тесты с триггерами, различные анимационные приёмы и др. Некоторые из пособий созданы в жанре игры-путешествия, например «География: Энциклопедия для маленьких детей», «По родным просторам. Удмуртия – край мой родниковый».

Говоря о вышеназванных электронных учебных пособиях в целом, следует отметить, что их использование на занятиях в учебном учреждении или в домашних условиях может дать хорошие метапредметные результаты освоения учебного материала: овладение начальными сведениями о сущности и особенностях объектов, процессов и явлений действительности (природных и др.) в соответствии с содержанием учебного предмета; овладение базовыми предметными и межпредметными понятиями, отражающими существенные связи и отношения между объектами и процессами. Предметными результатами освоения могут стать: сформированность уважительного отношения к России, родному краю, природе нашей страны; осознание целостности окружающего мира, освоение основ экологической



грамотности, элементарных правил нравственного поведения в мире природы и людей, норм здоровьесберегающего поведения в природной среде; освоение доступных способов изучения природы; развитие навыков установления и выявления причинно-следственных связей в окружающем мире; овладение основами логического и алгоритмического мышления, пространственного воображения (пособия по логике).

Таким образом, можно сказать, что созданные за последние годы электронные образовательные ресурсы на удмуртском языке (а также и на русском) помогут обучающимся познавать различные аспекты окружающего мира, при этом изучая, усваивая язык.

* * *

1. *Кудрявцев А.Ф., Лекомцев А.Л., Тимерханова Е.Р., Тимерханова Н.Н.* География: Пичи нылпиослы энциклопедия (География: Энциклопедия для маленьких детей) / отв. ред. Н.Н. Тимерханова, научный ред. А.Ф. Кудрявцев. – Ижевск, 2013. – [CD-ROM. 570 слайдов].

2. *Кашин А.А., Кудрявцев А.Ф.* Удмуртилэн ландшафтёсыз: Туспук-тэмъёслэсь комментариосын альбом (Ландшафты Удмуртии: Фотоальбом с комментариями) / отв. ред. Н.Н. Тимерханова. – Ижевск, 2013. – [CD-ROM. – 150 слайдов. – На удмуртском и русском языке].

3. *Кашин А.А., Лекомцев А.Л., Тимерханова Н.Н.* Географиен олимпиада юанъёс. – Хельсинки: М.А. Кастренлэн Мерлыкес, 2014 (рус. назв.: Кашин А.А., Лекомцев А.Л., Тимерханова Н.Н. Вопросы олимпиады по географии). – Хельсинки: Общество М.А. Кастрена, 2014; фин. назв.: Kašin A., Lekomtsev A., Timerhanova N. Maantiedon kysymyksiä koululais- ja opiskelijakilpailuja varten. – Helsinki: M.A. Castrénin seura, 2014).

4. Удмурт Элькунлэн географияез: шоръёзо дышетскон юртъёслы электрон дышетскон пособие / А.Ф. Кудрявцев, А.Л. Лекомцев, Н.Н. Тимерханова но мукет; кылкутись ред. Н.Н. Тимерханова. – Хельсинки: М.А. Кастренлэн Мерлыкес, 2014. – [DVD. – 427 слайд, 3 видеофильм] (рус. назв.: Кудрявцев А.Ф., Лекомцев А.Л., Тимерханова Н.Н. География Удмуртской Республики: электр. учеб. пособие для общеобразовательных учебных заведений. – Хельсинки: Общество М.А. Кастрена, 2014; фин. назв.: Udmurtian maantiedon tehtäväkirja: Sähköinen oppimateriaali yläkouluja varten / A. Kudrjavtsev, A. Lekomtsev, N. Timerhanova. – Helsinki: M.A. Castrénin seura, 2014).

5. *Лобыгин А.Н., Симонова Н.А., Тимерханова Н.Н.* Удмурт Элькунлэн географияезъя ужъёсын дышетскон тетрадь: дышетскон юрттэт. –



Хельсинки: М.А. Кастренлэн Мерлыкес, 2014 (рус. назв.: Лобыгин А.Н., Симонова Н.А., Тимерханова Н.Н. Тренажёр-практикум по географии Удмуртской Республики: учеб. пособие. – Хельсинки: Общество М.А. Кастрена, 2014; фин. назв.: Lobygin A., Simonova N., Timerhanova N. Udmurtian maantiedon tehtäväkirja: oppimateriaali. – Helsinki: M.A. Castrénin seura, 2014).

6. *Тимерханова Н.Н., Кудрявцев А.Ф., Лекомцев А.Л.* Удмуртиысь туризм: этнотуризм, лулчеберетэн тодматскон, эмъяськон-шутэтскон но гуртысь туризм = Туризм в Удмуртии: Этнотуризм, сельский, лечебно-оздоровительный и культурно-познавательный туризм в Удмуртской Республике. – Хельсинки: Общество М.А. Кастрена, 2016. – [CD]. – URL: <https://sites.google.com/site/turizmvudmurjrespublike>

7. *Тимерханова Н.Н.* Удмурт кыл (Удмуртский язык): учеб.-тетрадь для начинающих № 1. – Ижкар: Удмурт. ун-т, 2012.

8. *Тимерханова Н.Н.* Удмурт кылын шонергожъяськонъя ужъёс: 9 презентацилэсь сузъет (Упражнения по орфографии удмуртского языка: сб. из 9 презентаций). – Хельсинки: Общество М.А. Кастрена, 2016. – [CD]. – URL: <https://sites.google.com/site/udmurtkylsykys/sutetskon/uzees>

9. *Тимерханова Н.Н., Тимерханов В.Р.* Ёырвизез сынан: Анимациен шудонъёс: 6–7 аресъем пинальёслэсь логика малпаськонзэс азинтйсь ужъёс: дышетскон юрттэт. – Хельсинки: М.А. Кастренлэн Мерлыкес, 2014 (рус. назв.: Тимерханова Н.Н., Тимерханов В.Р. Разминка для ума: Анимированные игры: Задания для развития логического мышления детей 6–7 лет: учеб. пособие. – Хельсинки: Общество М.А. Кастрена, 2014; фин. назв.: Timerhanova N., Timerhanov V. Aivojumppaa: Harjoituksia logiikan ja päättelykyvyn ke-hittämiseksi 6–7-vuotiaille lapsille: oppimateriaali. – Helsinki: M.A. Castrénin seura, 2014).

10. *Тимерханова Н.Н., Бородин А.Е., Широбокова С.Н.* Визьбугор: Анимациен шудонъёс: 6–7 аресъем пинальёслэсь логика малпаськонзэс азинтйсь ужъёс: дышетскон юрттэт. – Хельсинки: М.А. Кастренлэн Мерлыкес. – 2014 (рус. назв.: Тимерханова Н.Н., Бородин А.Е., Широбокова С.Н. Умница: Анимированные игры: Задания для развития логического мышления детей 6–7 лет: учеб. пособие. – Хельсинки: Общество М.А. Кастрена, 2014; фин. назв.: Timerhanova N., Borodina A., Širobokova S. Älypää: Harjoituksia logiikan ja päättelykyvyn kehittämiseksi 6–7-vuotiaille lapsille: oppimateriaali. – Helsinki: M.A. Castrénin seura, 2014).

11. *Тимерханова Н.Н., Лекомцев А.Л.* Визь куды: Анимациен шудонъёс: 6–7 аресъем пинальёслэсь логика малпаськонзэс азинтйсь ужъёс: дышетскон юрттэт. – Будапешт–Бадачоньтёрдэмиц: NH-Collegium Fennougricum, 2014 (рус. назв.: Тимерханова Н.Н., Лекомцев А.Л. Золотая голова: Анимированные игры: Задания для развития логического мышления



детей 6–7 лет: учеб. пособие. – Будапешт–Бадачоньтёрдемиц: NH-Collegium Fenno-Ugricum, 2014; венг. назв.: Timerhanova N., Lekomcev A. Tudorka: A logikus gondolkodás fejlesztése 6–7 éves gyermekek számára: tananyag. – Budapest–Badacsonytördemic: NH-Collegium Fenno-Ugricum, 2014).

12. *Тимерханова Н.Н., Лекомцев А.Л., Федина М.С.* Вежөр куд: 6–7 арöса челядьлы логика сөвмöдан уджъяс: велöдчан пособие. – Сыктывкар–Будапешт–Бадачоньтёрдемиц: NH-Collegium Fenno-Ugricum, 2014.

13. *Тимерханова Н.Н., Лекомцев А.Л., Гусева Э.В.* Шöртнö вуй: Задания для развития логического мышления детей 6–7 лет: учеб. пособие. – Йошкар-Ола–Будапешт–Бадачоньтёрдемиц: NH-Collegium Fenno-Ugricum, 2014.



УДК 372.8:91(045)

Н.Н. Тимерханова,

*канд. филол. наук, доцент,
зав. кафедрой лингвистической
типологии и лингводидактики
Удмуртского государственного
университета (г. Ижевск)*

А.Ф. Кудрявцев,

*канд. геогр. наук, директор
института естественных наук
Удмуртского государственного
университета (г. Ижевск)*

ЭНЦИКЛОПЕДИЯ ПО ГЕОГРАФИИ НА УДМУРТСКОМ ЯЗЫКЕ

В 2013 г. вышло в свет электронное издание «География: Пичи нылпиослы энциклопедия» («География: Энциклопедия для маленьких детей» [1] (см. *рис. 1*)), которое является результатом совместной научно-методической работы кафедры экономической географии географического факультета и кафедры удмуртского языка и методики его преподавания факультета удмуртской филологии Удмуртского госуниверситета (ныне кафедра общей и физической географии института естественных наук и кафедра лингвистической типологии и лингводидактики института удмуртской филологии, финно-угроведения и журналистики). Это учебно-научное издание, подготовленное в рамках проекта «“География” на удмуртском языке», предназначено для детей раннего возраста, владеющих удмуртским языком.

В данной электронной энциклопедии авторский коллектив (А.Ф. Кудрявцев, А.Л. Лекомцев, Е.Р. Тимерханова, Н.Н. Тимерханова) стремился отразить самые важные и интересные географические объекты и некоторые факты, связанные с ними, в доступной для детей форме, но в то же время авторы преследовали цель дать ребёнку основополагающие энциклопедические знания по географии. Поэтому, кроме теоретического материала, в пособие включены эксперимент, моделирование (см. *рис. 2*), простые анимационные игры (см. *рис. 3*), викторина из разнотипных заданий (см. *рис. 4*) и физкультминутки (см. *рис. 5*).

Для того чтобы работать с энциклопедией было удобно, некоторые разделы находятся внутри более общего раздела. Так, биосфера рассматривается в связи с конкретным материком и определённой территорией, а землетрясения и вулканы входят в раздел «Рельеф» и др. Такое деление



поможет ребёнку выявить взаимосвязь явлений и фактов. Приведём в качестве примера общее содержание энциклопедии (на русском языке) и содержание второго уровня, которые, на наш взгляд, должны быть удобной, красочной и интересной навигацией (см. рис. 6 и 7).

Необходимо отметить, что при создании энциклопедии была проведена большая терминологическая работа, поскольку многие термины-неологизмы в удмуртском языке ещё не устоялись. Очень аккуратно подбирались названия географических объектов и явлений; при этом выбирались наименования, известные широкому кругу читателей (и взрослым, и детям). Для читателей, слабо владеющих удмуртским языком, к мало знакомым либо знакомым только старшему поколению терминам созданы гипертекстовые «страницы» словаря с переводом некоторых слов с удмуртского языка на русский (см. рис. 8).

Электронная энциклопедия создана в программе *Microsoft Power Point*, т.к. её возможности весьма обширны, и в то же время с данным продуктом сможет познакомиться практически любой пользователь персонального компьютера, на котором установлен набор программного обеспечения *Microsoft Office*, а также его сможет использовать на занятиях любой воспитатель в дошкольном учебном учреждении или учитель в школе. Необходимо также отметить, что тиражирование учебного издания на компакт-дисках позволяет использовать и соответственно видеть преимущества электронной книги и является более доступным по цене для большей аудитории читателей. Таким образом, обеспечивается широкий доступ к данному учебному ресурсу.

Рассматриваемая электронная энциклопедия в отличие от своих бумажных аналогов обладает дополнительными свойствами и возможностями: 1) имеет удобную систему навигации на основе гиперссылок; 2) включает в себя аудио- и видеофрагменты. Материалы пособия можно использовать в индивидуальном режиме или демонстрировать на занятиях и работать с ними в группе.

Как уже было отмечено, в рассматриваемой электронной энциклопедии широко используется гипертекстовая система – разнообразная информация, располагающаяся не только на разных слайдах, но и в разных файлах, в Интернете. Преимуществом гипертекста является возможность быстрого перехода по гиперссылкам, которые в данном случае оформлены или в виде текста на том же слайде, но скрытом при помощи анимации, или на отдельном слайде, т.е. связаны с местом в документе (особенно такие ссылки были необходимы при оформлении в другом дизайне), или же в другом файле при необходимости создания отдельных частей презентации для прямого доступа к отдельно взятой части энциклопедии, а также для того, чтобы файлы открывались быстро и не был замечен переход (т.к. в энци-

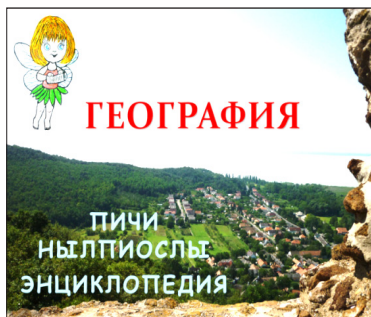


Рис. 1. Электронное издание «География: Пичи нылпиослы энциклопедия» («География: Энциклопедия для маленьких детей»)

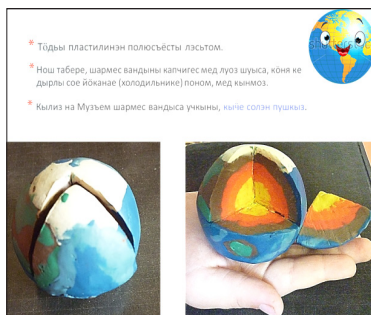
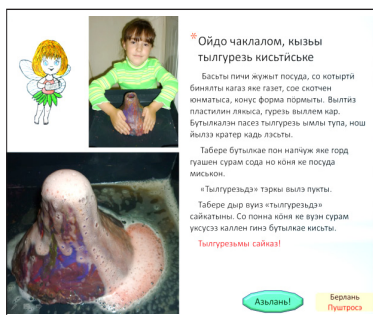


Рис. 2. Эксперимент

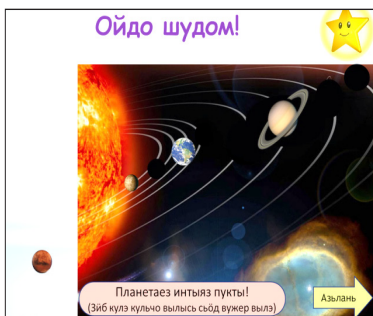


Рис. 3. Анимационные игры



Рис. 4. Викторина

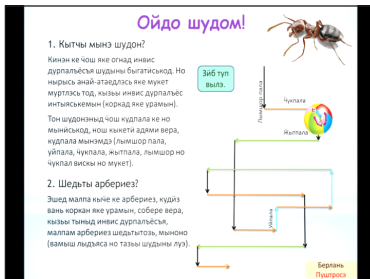
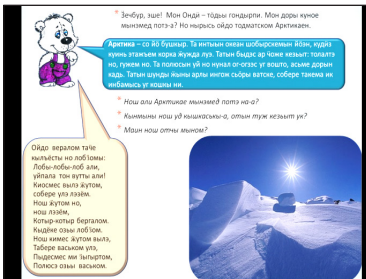


Рис. 5. Физкультминутки



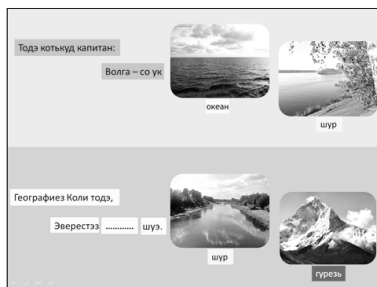
Рис. 6. Общее содержание энциклопедии



Рис. 7. Содержание второго уровня



а)



б)



в)

Рис. 12. Задания для самоконтроля и закрепления материала (а-в)

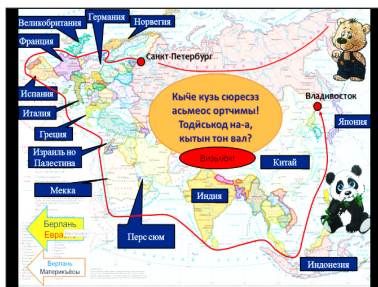


Рис. 13. Виртуальное путешествие в рассмотренном электронном издании



Рис. 14. Информация об Удмуртии



клопедии около 570 слайдов, то могли возникнуть проблемы с открытием файла). Кроме этого, имеются гиперссылки на веб-страницы.

Гиперссылки в пособии представлены в виде отдельно оформленного текста, отдельных слов (рис. 8), фигурных стрелок, управляющих кнопок или определённых графических изображений. На некоторых страницах энциклопедии (экране компьютера) имеется одна гиперссылка либо одновременно располагается несколько одинаковых или разных гиперссылок (см. рис. 9), каждая из них определяет свой маршрут «путешествия», направляя в нужную часть энциклопедии (см. рис. 10). Наряду с графическими и текстовыми гиперссылками в электронном пособии имеются гиперссылки на мультимедиаресурсы, например на звук (звуки, издаваемые животными и птицами: крик дельфинов, голоса китов, пыхтение ежа и др., а также шум водопада, вьюги, раскаты грома, бурление лавы в вулкане и т.п. (см. рис. 11)), видео (шум орбиты Земли), анимацию (простую и созданную через триггеры). То есть в электронной энциклопедии используется гипермедиа.

Как отмечает И.Г. Захарова в учебном пособии «Информационные технологии в образовании» (и с чем нельзя не согласиться), «использование в электронных изданиях различных информационных технологий (АОС, мультимедиа, гипертекст) даёт весомые **дидактические преимущества** электронной “книге” по сравнению с традиционной:

- в технологии мультимедиа создаётся обучающая среда с ярким и наглядным представлением информации, что особенно привлекательно для школьников;
- осуществляется интеграция значительных объёмов информации (до 700 Мб) на едином носителе;
- гипертекстовая технология благодаря применению гиперссылок упрощает навигацию и предоставляет возможность выбора индивидуальной схемы изучения материала;
- на основе моделирования процесса обучения становится возможным дополнить учебник тестами, отслеживать и направлять траекторию изучения материала, осуществляя таким образом обратную связь» [2].

Для закрепления изученного материала и самоконтроля в энциклопедии созданы слайды с вопросами по ранее изученным материалам, а также красочные тренировочные задания, которые гармонично входят в электронную энциклопедию в качестве подсистемы. Их основой являются слайды с гиперссылками и анимацией, имеющими форму тестов, кроссвордов, зоологической карты и др. (см. рис. 12).

В основном читателю предлагается традиционное изложение текста справочной статьи со статичным рисунком или фотографией, что особенно хорошо подходит для читателей, имеющих вербальный тип мышления.



Но наряду с таким изложением имеются и анимированные иллюстрации (анимация *Power Point*), которые более привлекательны для детей с образным типом мышления. Благодаря анимированному ряду им легче ориентироваться в связях между явлениями, видеть динамику действий, процессов и явлений, а также сравнивать их.

В понимании географии как обучающего предмета о разнообразии Земли важнейшую роль играет сравнительный метод познания. На наш взгляд, наиболее удачной формой реализации этого метода является виртуальное путешествие. В энциклопедии представлен вариант такого путешествия вокруг Евразии – крупнейшего материка Земли – из Петербурга до Владивостока. Главная героиня фея Майя и её друзья, привязанные к определённым географическим объектам, например Опоссум Опи из Северной Америки, броненосец Броня из Южной Америки, львёнок Лёви из Африки, коала Коли из Австралии и др., знакомят детей с соответствующими географическими объектами, а обучающихся вместе с Майей пересекают три океана и несколько природных зон, посещают около десятка стран (см. *рис. 13*) и попутно знакомятся с компонентами окружающего земного мира, сравнивая их.

Большее внимание уделено в энциклопедии России и родному краю – Удмуртии (см. *рис. 14*). Обучающиеся знакомятся с географическим расположением, рельефом, климатом, природой региона, особенностями флоры и фауны, водоёмами. Не остаётся без внимания информация о населении, административном делении республики и населённых пунктов, республиканской символике, об особенностях сельскохозяйственного и промышленного производства.

В заключение хотелось бы сказать, что в рассмотренном электронном издании авторы попытались совместить современные носители информации и способы её представления, востребованные среди подрастающего поколения, с традиционным, проверенным веками содержанием географического знания. Подобный синтез позволит, на наш взгляд, заложить первоначальный интерес у детей младшего и дошкольного возраста к увлекательной науке географии, являющейся одним из краеугольных камней общего развития и кругозора любого человека.

* * *

1. Кудрявцев А.Ф., Лекомцев А.Л., Тимерханова Е.Р., Тимерханова Н.Н. География: Пичи нылпиослы энциклопедия (География: Энциклопедия для маленьких детей) / отв. ред. Н.Н. Тимерханова, науч. ред. А.Ф. Кудрявцев. – Ижевск, 2013. – [CD-ROM. – 570 слайдов].

2. Захарова И.Г. Информационные технологии в образовании: учеб. пособие для студ. высш. учеб. заведений. – 4-е изд., стер. – М., 2008.



М.С. Федина,

канд. филол. наук, доцент, начальник

*Центра инновационных языковых
технологий ГОУ ВО КРАГСиУ*

Д.А. Левченко,

инженер-программист

*Центра инновационных языковых
технологий ГОУ ВО КРАГСиУ*

ИЗ ОПЫТА СОЗДАНИЯ КОМИ МЕДИАТЕКИ

Документирование языка – достаточно новая область лингвистики, которая понимается как «долговременная многоцелевая фиксация языковых данных» [1], а основной её целью является сбор и фиксация первичных данных по малым или неизученным языкам для их использования в дальнейших научных исследованиях и учебно-методических целях.

Изначально данная область лингвистики разрабатывалась теми исследователями, которые занимались языками, находящимися на грани исчезновения и чаще всего не имевшими ни электронной письменности, ни каких-либо других электронных ресурсов. В настоящее время многие диалектологи используют современные методы электронного документирования языка в связи с тем, что для диалектов, во-первых, также свойственно отсутствие документирования, во-вторых, они также находятся под серьёзной угрозой исчезновения, а в-третьих, даже диалекты тех языков, которым ничего (или в менее значительной степени) не угрожает, могут предоставить очень интересную информацию для теоретических лингвистических и иных исследований.

При документировании в основном записываются тексты разных жанров: различные легенды, рассказы, сказки, байки, анекдоты, стихотворения, фольклорные праздники и постановки, диалоги на повседневные бытовые темы, песни и др. Но наибольшую ценность имеют записи устной спонтанной речи, собранные в полевых условиях при интервьюировании на различные темы. Устная спонтанная речь является основной, базовой для человека во всех отношениях: она исторически исходна, естественна и функционально многоаспектна. Устная речь используется человеком на протяжении всей жизни, в то время как письменную форму могут использовать не все (или используют намного реже). Устная речь является зеркалом современ-



ной жизнедеятельности человека, в ней отражаются и индивидуальные, и коллективные особенности говорящих на данном языке в данное время и в данном месте. Но исследовать устную речь без фиксации крайне сложно, поэтому документирование языка ставит своей целью предоставление информации для будущих междисциплинарных исследований языков в частности и лингвистических практик в целом.

При этом важнейшими задачами документирования являются не только сбор и качественная фиксация языкового материала, но и предоставление широкой аудитории доступа к нему в электронной форме и надёжное долговременное хранение собранных данных и их периодическая обработка по требованиям современных форматов. Значение языкового документирования огромно. «Для местных жителей записанные тексты интересны тем, что привязаны к их сегодняшней жизни, быту, знакомым людям и судьбам. Кроме того, сам факт, что их языком интересуются учёные, что эти материалы будут доступны во всём мире, несомненно, поднимает престиж языка, может способствовать пробуждению интереса к собственному языку у молодых. Для будущих поколений – это сохранение культурного наследия. Конечно же, собранные и обработанные материалы имеют огромную ценность для всех специалистов, изучающих язык в самых разных аспектах» [2].

Необходимость современного документирования диалектов коми языка и сохранения данных в виде электронного корпуса устной речи признана всеми учёными-филологами и неоднократно обсуждалась, в том числе и в Коми научном центре УрО РАН [3].

Центром инновационных языковых технологий была предпринята попытка создания ресурса коми устной диалектной речи – коми медиатеки (<http://videocorpora.ru>). Она представляет собой электронную справочно-информационную систему на основе электронного архива аннотированных записей (видео, аудио, расшифрованные тексты на коми языке, в переводе на русский и английский языки) диалектных текстов коми языка с возможностью поиска языковых единиц по различным параметрам. Коми медиатека – это ещё не полноценный корпус устной речи, разные виды которого уже представлены в Интернете, например: проект «Рассказы о сновидениях и другие корпуса звучащей речи» (<http://www.spokencorpora.ru>), устный подкорпус национального корпуса русского языка (<http://www.ruscorpora.ru/search-spoken.html>), устный корпус осетинского языка (<http://www.ossetic-studies.org/ru/texts>). На данный момент коми медиатека – это обработанная коллекция текстов на ижемском диалекте коми языка, взятых из различных архивов и собранных в ходе полевых экспедиций в рамках международного научно-исследовательского проекта «Коми-ижемцы: Создание аннотированного электронного корпуса для дальнейших исследований по коми



языковым сообществам российского Крайнего Севера» в Ижемском, Со-сногорском районах Республики Коми, Ненецком автономном округе, Мурманской области. Основной целью этого проекта было социолингвистическое исследование территорий проживания ижемцев, сравнительное исследование локальных говоров ижемского диалекта и создание электронного архива аннотированных записей [4]. Участниками проекта были д-р филол. наук, профессор финно-угорского языкознания Института современных языков университета г. Уппсала Р. Блокланд (Швеция), д-р филол. наук, руководитель лаборатории саамских языков Фрайбургского университета им. Альберта Людвиг М. Рисслер (Германия), канд. филол. наук, доцент, начальник Центра инновационных языковых технологий Коми республиканской академии государственной службы и управления М. Федина, аспирант Н. Партанен (Университет г. Хельсинки, Финляндия – Университет г. Гамбург, Германия), магистрант, переводчик Бюро официального перевода Центра инновационных языковых технологий Коми республиканской академии государственной службы и управления В. Чупров, инженер-программист Центра инновационных языковых технологий Коми республиканской академии государственной службы и управления Д. Левченко (Сыктывкар, Республика Коми, Россия), магистрант М. Федина (Университет г. Йохенсуу, Финляндия).

Перед нами стояло несколько задач:

- 1) собрать в естественных условиях и качественно зафиксировать на современных носителях языковой материал;
- 2) предоставить к собранному и обработанному материалу доступ в электронной форме широкой аудитории, для чего, во-первых, перевести аудио- и видеоматериалы в письменную форму (транскрибировать); во-вторых, сделать перевод на русский и английский языки для понимания содержания текстов носителями других языков; в-третьих, создать удобный для пользователей электронный ресурс;
- 3) обеспечить надёжное долговременное хранение собранных данных, возможность их периодической обработки по требованиям современных форматов и пополнение ресурса другими материалами.

За период с 2014 по 2016 г. в ходе 4 экспедиций было опрошено 196 ижемцев из разных регионов, сделано 157 интервью продолжительностью от 20 мин. до 1 часа каждое. Кроме того, восстановлено 16 записей из старых фондов, недоступных широкому кругу пользователей. Возраст информантов был разным – от 5 до 80 лет. Вся информация об информантах представлена с их согласия следующим образом:

1. Карта с указанием места жительства информанта (см. *рис. 1*).

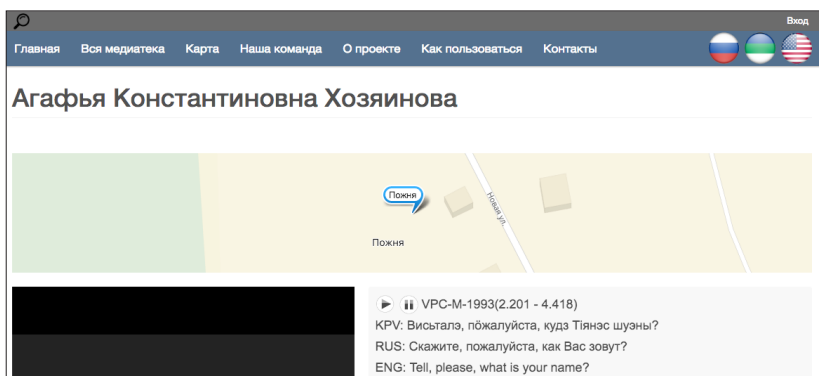


Рис. 1. Карта с указанием места жительства информанта

В данной области отображается подробная информация о каждом участнике интервью: фотография, фамилия, имя, отчество, дата рождения, место рождения, место жительства (на момент интервью) (см. рис. 2).

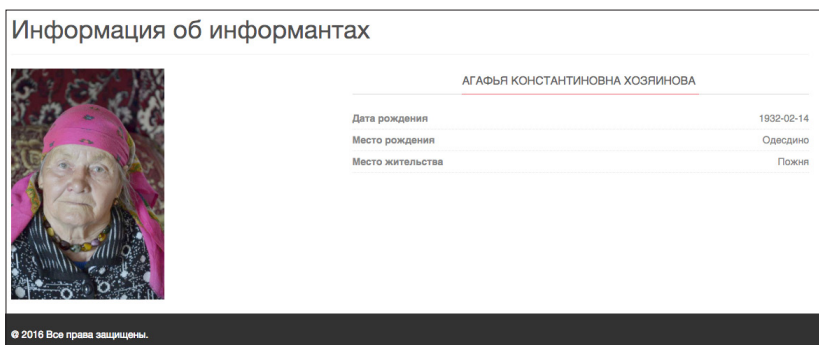


Рис. 2. Информация об информанте

В медиатеке представлены различные тексты: песни, частушки, театральные и детские постановки, архивные записи, интервью, анекдоты, рассказы и т.д. Основной объём материала составляют интервью с информантами на различные темы: о семье, о себе, об истории рода, истории переселения, работе, увлечениях, учёбе, творчестве, народной медицине, культуре, оленеводстве и т.д. Интервью брались в естественных для информантов условиях: в детском саду, школе, на улице, дома, в мастерской, на работе, на отдыхе, в лесу. Причём по форме они представляют собой интервью как с одним информантом, так и с двумя (супружеская пара, родственники,



хорошие соседи) и более информантами. На данный момент на <http://videocorpg.ru> загружены не все тексты, работа по пополнению этого ресурса продолжается.

Одной из основных сложностей в создании баз или корпусов устной речи является транскрипция. Имеются разные решения этой проблемы, которые можно наблюдать в разных корпусах. Мы исходили в первую очередь из того, что «транскрибируя устный дискурс, следует основываться на верности реальной аудиозаписи... Необходимо избегать искушения подгонки под привычные шаблоны письменного языка. То, чего не может быть в письменном языке (или мы думаем, что его там не может быть), вполне может встречаться в устном. В особенности это касается речи субъектов, относящихся к не совсем стандартным с точки зрения исследователя социальным группам – детей, людей с нейropsychическими расстройствами, людей “не своего” социального слоя, и т.п.» [5]. С другой стороны, основная цель – сделать этот материал доступным для восприятия пользователями – продиктовала необходимость не использовать научную фонетическую транскрипцию (финно-угорскую (ФУТ) или международную (МФА)), которая понятна только специалистам, а передать аудиоматериалы приближенно к коми графике. Разумеется, это будет неточная передача качества отдельных звуков, например качество [Эö/Ӗэ], [Ч/Чть/ТЬ] и т.д. Но и в ФУТе и МФА возможно отобразить не все переходные варианты звуков, аллофоны.

В коми медиатеке мультимедийный файл и текстовая расшифровка взаимосвязаны. Область взаимодействия с мультимедийным файлом и область текстовой расшифровки с элементами управления выглядит следующим образом (см. *рис. 3*).

1. Область взаимодействия с мультимедиафайлом. Для запуска видео или аудио необходимо нажать левой кнопкой мыши по центру области, для остановки – повторно нажать левой кнопкой мыши по центру области.

В нижней части области находятся элементы управления (см. *рис. 4*).

Слева направо: текущее время, продолжительность видео, кнопка выбора языка субтитров, регулировка уровня звука, кнопка развёртывания видео на весь экран.

2. Область текстовой расшифровки с элементами управления. Здесь представлена текстовая расшифровка мультимедийных файлов, которая разбита на предложения или фразы. В каждом сегменте присутствуют элементы управления (старт, пауза), название слоя, время и перевод на трёх языках.

Было очень сложно транскрибировать диалоги и полилоги, особенно когда происходило наложение реплик при одновременном разговоре нескольких людей. Так как в текстовом формате наложение реплик отобразить не-

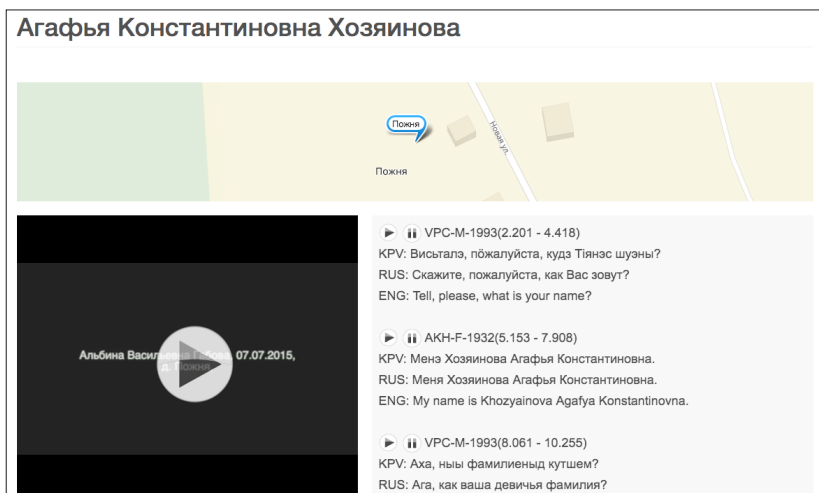


Рис. 3. Область взаимодействия с мультимедийным файлом и область текстовой расшифровки с элементами управления



Рис. 4. Элементы управления

возможно, мы попытались показать весь речевой поток по временной шкале – «*Timeline*», что достаточно удобно и для работы с диалогами (см. рис. 5).

Для отображения данной возможности необходимо нажать «Показать *Timeline*».

При создании коми медиатеки были использованы современные языковые и программные средства, максимально отвечающие поставленным задачам, в том числе обеспечивающие скорость обработки запросов, надёжное долговременное хранение и создание резервных копий собранных данных, возможность их периодической обработки по требованиям современных форматов и пополнение ресурса другими материалами.

Перечень использованного программного обеспечения:

- операционная система: *Ubuntu 16.04 server*;
- база данных: *MySQL 5.7.16*;
- веб-сервер: *Apache/2.4.18*;
- платформа сайта: *CMS Drupal 7.53*;
- языки программирования: *PHP, JavaScript*.



Рис. 5. Шкала «Timeline»

Одно из основных программных средств метаразметки аудио- и видеофайлов в проекте «Коми медиатека» – *ELAN*. «Программа *ELAN* является на сегодняшний день одной из лучших программ многоуровневого аннотирования мультимедийных данных. *ELAN* поддерживает визуализацию аудио- и/или видеосигналов одновременно с полученными аннотациями; временную привязку аннотаций к медийному потоку; сложные связи аннотаций друг с другом; неограниченное количество задаваемых пользователем уровней аннотации (*Tiers*); различные шрифты и кодировки; экспорт данных в виде текстовых файлов табличного вида (*tab-delimited text*); импорт и экспорт между *ELAN*, *Praat*, *ToolBox*, *Shoebox* и другими популярными лингвистическими программами; развитые поисковые опции» [6]. Результатом обработки исходных мультимедиафайлов в данном проекте являются файлы с расширением *.eaf* (по своей структуре полный аналог *.xml*). Полученные данные импортируются в базу данных проекта для дальнейшей обработки и хранения.

Однако данный ресурс предназначен для хранения и обработки информации не только по одному диалекту коми языка – ижемскому, но также и по другим диалектам. В будущем коми медиатека должна стать ресурсом, где представлен весь спектр коми диалектов с добавлением слоёв научной транскрипции и коми орфографии с привязкой диалектного словаря.



* * *

1. *Архипов А.В.* Документирование малых языков: научные и технические аспекты // Языковое разнообразие в киберпространстве: российский и зарубежный опыт. – М., 2008. – С.76–83.
2. *Архипов А.В.* Указ. соч.
3. *Некрасова Г.А.* Электронный диалектный корпус как ресурс сохранения и изучения коми диалектов // Финно-угорский мир. – 2010. – № 1. – С.13–16.
4. *Федина М.С.* Документирование ижемского диалекта коми языка // BiZ-Vote. – 2016. – № 3. – С.18–21.
5. *Кибрик А.А., Подлесская В.И.* К созданию корпусов устной русской речи: принципы транскрибирования. – URL: http://www.philol.msu.ru/~otipl/new/main/people/kibrik-aa/files/Corpora_speech_transcription@S&I_2003.pdf
6. *Блокланд Р., Рисслер М., Партанен Н., Федина М., Чемышев А.* Использование цифровых корпусов и компьютерных программ в диалектологических исследованиях: теория и практика // Актуальные проблемы диалектологии языков народов России: материалы XIV Всероссийской науч. конф., посвящённой 90-летию со дня рождения известного башкирского языковеда, диалектолога, тюрколога, д-ра филол. наук С.Ф. Миржановой (20–22 октября 2014 г.). – Уфа, 2014. – С.252–255.



С. Ферхеес,

*аспирант школы лингвистики
Национального исследовательского
университета «Высшая школа
экономики» (г. Москва)*

ЭВИДЕНЦИАЛЬНОСТЬ И ПЕРФЕКТ В РУТУЛЬСКОМ ЯЗЫКЕ (на материале говора с. Кина)

В нахско-дагестанских языках широко представлена категория эвиденциальности, или грамматическое маркирование источника сведений говорящего о сообщаемой им ситуации. Представление этого признака на географической карте отображает определённое ареально-генетическое распределение внутри языковой семьи, которое требует дальнейшего исследования. В то время как в большинстве нахско-дагестанских языков эвиденциальность выражается (в том числе и) перфектной формой глагола, считается, что в рутульском языке (и в некоторых других) в глагольной парадигме этот признак отсутствует. Статья посвящена вопросу наличия данного признака во временных формах, как перфект в рутульском языке, а также вопросу о том, что выражает рутульский перфект, если не эвиденциальность. В первом разделе обсуждается существующая литература по этой теме, второй раздел даёт характеристику системы прошедшего времени в рутульском языке (конкретно в ныне неопisanном диалекте с. Кина), в третьем разделе описаны результаты полевого исследования, проведённого автором данной статьи в с. Кина летом 2016 года. Представленные данные исследования и их значение для типологии нахско-дагестанских перфектов обобщены в заключении.

Эвиденциальность в нахско-дагестанских языках.

В нахско-дагестанских языках есть два основных способа выражения категории эвиденциальности*: 1) перфектная форма глагола; 2) частицы косвенной речи или инференции. Они представляют собой маленькие с типологической точки зрения системы, в которых различается в первую очередь косвенная засвидетельствованность. В некоторых языках также маркируется прямая засвидетельствованность (обзор об эвиденциально-

© Ферхеес С., 2017

* Бывают ещё аналитические конструкции с глаголами («найти» или «останавливаться» в качестве вспомогательного глагола), но про них пока мало известно.



сти в нахско-дагестанских языках см. в работе [1]). Косвенная засвидетельствованность включает в себя информацию из чужих слов, инференцию из очевидных последствий, а также инференцию на основе какого-то личного знания говорящего [2]. Как описано в том числе в источнике [3], выражение эвиденциальности перфектной формы глагола на самом деле является ареальным признаком большой территории, окружающей носителей нахско-дагестанских языков. При этом эвиденциальность как грамматическая категория считается «чувствительной» к заимствованию в ситуациях языкового контакта.

Типология перфекта. Под термином «перфект» мы понимаем форму глагола типично аналитического строя со вспомогательным глаголом, которая противопоставлена менее маркированному перфективному прошедшему и выражает одно из значений, которое ассоциируются с перфектом в типологической литературе [4].

Перфект известен тем, что он семантически нестабилен. У этой формы, как правило, не одно, а несколько значений, и, как отмечено В.А. Плунгяном [5], нестабильность, наверное, его самый универсальный признак. При этом, как правило, набор его возможных значений весьма разнообразен. Из этого следует, что для описания языка ярлык «перфект» не достаточен, т.к. наименование само по себе мало говорит о конкретной семантике некоторой формы. В европейских языках, например, более распространены значения «текущей релевантности» [6], а в нахско-дагестанских языках преобладает эвиденциальность. Как ранее было отмечено, например, С.Г. Татевосовым [7], значения текущей релевантности в нахско-дагестанских языках присутствуют, но они имеют весьма маргинальный характер. Рутульский язык в этом плане является особенным на уровне языковой семьи, что и будет доказано ниже.

Ареально-генетическое распределение. На *рисунке* показано распределение разных способов выражения эвиденциальности в нахско-дагестанских языках*. На карте представлены данные, которые описаны в научной литературе, созданной на современном этапе изучения нахско-дагестанских языков. На данный момент нельзя окончательно утверждать о наличии или отсутствии признака, основываясь на материалах изученных работ. Существует вероятность того, что для некоторых языков признак эвиденциальности ещё не описан или же, наоборот, представлен, но не имеет доказательной основы о содержании грамматикализованного характера.

* Карта была создана автором данной статьи с помощью пакета *lingtypology* для R, проработанного Г.А. Мороз, преподавателем Национального исследовательского университета «Высшая школа экономики» (см. подробности о пакете (на английском языке): <https://agricolamz.github.io/lingtypology>).

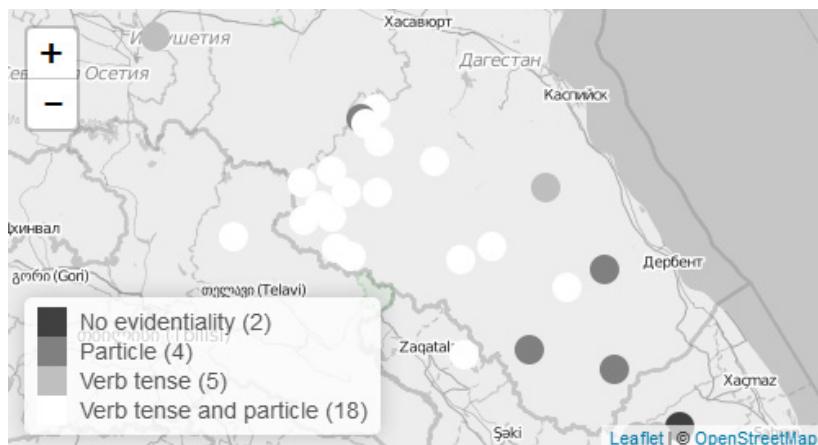


Рис. Эвиденциальность в нахско-дагестанских языках

Подробно рассмотрев данные рисунка, нетрудно заметить, что в большинстве языков эвиденциальность выражается временными формами глагола, так же как и частицами. К тому же стоит отметить, что лезгинские языки, используемые на юге Дагестана и на севере Азербайджана, отличаются от общего дагестанского типа*. Вероятно, можно говорить об отсутствии эвиденциальности в изолированных будухском и удинском языках, основываясь на данных научных источников. В рутульском, лезгинском и табасаранском языках употребляются только частицы, а в крызском, наоборот, наблюдается использование только временных форм глагола.

Исключениями являются агульский и арчинский языки. Последний находится в окружении и контакте с другими дагестанскими языками, что могло бы объяснить присутствие в нём грамматикализованного маркера эвиденциальности. С другой стороны, в агульском языке категория эвиденциальности в глаголе была открыта относительно недавно [8]**. Именно поэтому можно говорить о возможности неописанных ранее фактов нали-

* Отмечается также, что ингушский и чеченский языки отличаются от дагестанских отсутствием эвиденциальных частиц.

** В цахурском языке эвиденциальная семантика сложным образом перемешана с другими категориями (см.: Майсак Т.А., Татевозов С.Г. Кодирование эпистемического статуса средствами морфосинтаксиса (на материале цахурского языка) // Вопросы языкознания. – 1998. – № 1. – С.60–88; Maisak T.A., Tatevosov S.G. Beyond evidentiality and mirativity: Evidence from Tsakhur // L'énonciation médiatisée II. / eds. by Z. Guentchéva, J. Landabaru. – Louvain/Paris: Peeters, 2007. – С.377–406).



чия данной категории в остальных лезгинских языках. В отношении рутульского языка это невозможно, т.к. есть несколько источников, довольно подробно описывающих существующие формы глагольной парадигмы.

Система прошедшего времени в рутульском языке. В говоре с. Кина корневая система прошедших времён глагола выглядит следующим образом (см. табл.).

Т а б л и ц а

Прошедшее время в рутульском языке (диалект с. Кина)

Форма	w-ezas ('дойть')
Аорист	w-ezɪr
Перфект	w-ezɪr a
Плюсквамперфект	w-ezɪr aj
Имперфект	w-ezar aj

Примечание: Таблица основана на обзоре глагольных форм говора с. Кина, подготовленном Н.Р. Добрушиной во время экспедиции в 2016 году.

В действительности рутульский язык имеет гораздо большее количество форм прошедшего времени, но в таблице представлены самые частотные. Особый интерес вызывает разница между аористом (который равен перфективному деепричастию) и перфектом (который состоит из перфективного деепричастия с прибавлением связки *a* – 'есть, имеется, находиться'), потому что аналогичные формы в других дагестанских языках выражают оппозицию по эвиденциальности.

В работах С.М. Махмудовой [9] и М.Е. Алексеева [10] о рутульском и лучекском говорах финитная форма без связки не упоминается. По М.Е. Алексееву [11], в лучекском говоре 'аорист' состоит из перфективного деепричастия (*'terminative gerund'*) и связки *i*, но в говоре с. Кина формы с этой связкой оказываются малоупотребительными (см. ниже). У С.М. Махмудовой перфективное деепричастие со связкой *i* описано как «простое прошедшее результирующее», а «простое прошедшее» является перфективным деепричастием с добавлением *ij* [12]. Та же форма описана М.Е. Алексеевым как прошедшее общее, которое выражает «действие, *хабитуально* происходящее в прошлом» [13]. Поэтому кининский говор в этом плане оказывается ближе к мухадскому, описанному Г.Х. Ибрагимовым [14], для которого отмечено «прошедшее общее», формально совпадающее с деепричастием, на *-ir*. Форма со связкой описана как «недавно прошедшее недлительное». Та же форма описывается С.М. Махмудовой как «преждепрошедшее неопределённое результирующее» [15], а М.Е. Алексеевым как «перфект» [16]. Как уже упоминалось выше, помимо связки *a*



в рутульском имеется связка *i* со значением ‘есть, является’. Она также выступает в глагольных конструкциях, но в данных кининского говора употребление данной связки встречалось недостаточное количество раз. Исходя из этого в данной статье рассматриваются только формы со связкой *a**.

При элицитировании простых изолированных предложений о действиях, завершённых в прошлом, носители могут использовать перфект или аорист. Например:

- (1) *za-d xila-bir jidvir* (a)
я-ERG рука-PL помыть.CVB (AUX.PRS)
‘Я помыл руки.’

Такие случаи носители языка разных возрастов комментировали, что оба варианта допустимы и различие сложно объяснить, однако варианты со связкой часто переводятся со словом «уже».

Эвиденциальность. По результатам проведённых исследований более сложные примеры с разными эвиденциальными контекстами подтвердили мнение, что засвидетельствованность говорящего не играет роли при выборе формы прошедшего времени. Использовались 2 анкеты: для элицитирования эвиденциальности и для разных значений перфекта. Ни один пример с явным эвиденциальным контекстом не элицитировал определённую форму, которую нельзя было заменить на другую из-за параметра засвидетельствованности.

Не менее существен и следующий факт: носители языка при формулировании отмечали, что при использовании аориста осуществляется только передача факта, а употребление перфекта подчёркивает, что, событие произошло только что, или делает акцент на достижении результата действия. Соответственно, известный «эффект первого лица» не возникает, например, в предложениях типа (1) с перфектом**.

В то же время в эвиденциальных контекстах употреблялись имеющиеся в рутульском эвиденциальные частицы. Они являются непарадигматическими, т.е. факультативно добавляются к спрягаемой форме глагола.

* Следует отметить, что в данном исследовании связка пишется отдельно, но у носителей языка она произносится с варьирующей степенью слитности. Упомянутое расхождение отражается также в дескриптивной литературе. В одних источниках авторы указывают связку энклитически, а другие отмечают как отдельную словоформу.

** Эффект первого лица – ограничение на употребление перфекта с субъектом первого лица, которое вытекает из эвиденциальной семантики. Сочетание косвенной засвидетельствованности с действующим первым лицом даёт значение, что говорящий делал что-то неосознанно (см., напр.: Aikhenvald A.Y. Evidentiality. – New York: Oxford University Press, 2004. – С.219).



- (2) *šeher-d-e q'ixti-bir jidq'ir a-xwa*
 город-AD начальство-PL приехать.CVB AUX.PRS-REP
 'В город приехало начальство, говорят.'

Несомненно, факт использования носителями языка необязательных выражений эвиденциальности в ответах на представленную анкету подтверждает, что если бы в глаголе была грамматикализованная эвиденциальность, результаты отразили бы это явление.

Перфект и нарративы. Также стоит отметить, что в повествовании используется аорист, являющийся нехарактерным явлением для нахско-дагестанских языков. Перфект часто является предпочтительной формой в нарративах о событиях, свидетелем которых говорящий не был. При элицировании нарративов использовался только аорист, в отличие от изолированных предложений. В примере (3) представлена часть такого нарратива.

- (3) *hanu-wa le<w>šur dahar, s'iwX'ir yara-a.*
 DEM.OBL-ERG <M>взять.AOR камень бросать.AOR змея-IN
 'Он взял камень и бросил [его] на змею.'
had w-iq'ir
 DEM M-умереть.AOR
 'Та умерла.'

В данном примере при замене аористной формы на перфект, по мнению говорящих, подчёркивается, что это событие произошло только что. В общей типологии перфекта это называется «*hot news perfect*», или перфект недавнего прошедшего [17]. Спонтанные нарративы также преимущественно построены на аористе.

Результатив. В кишинском говоре перфект также имеет значение результатива – состояние, которое является результатом действия в прошедшем. Из-за этого некоторые глаголы используют перфект для обозначения состояния в настоящем:

- (4) *key-mav had saxir a*
 будить-PROH он спать.CVB AUX.PRS
 'Не буди его, он спит (=заснувший есть).'

Всё вышесказанное позволяет сделать вывод о том, что у рутульского перфекта действительно нет эвиденциального значения, что для нахско-дагестанского языка довольно необычно. Данная форма главным образом выражает только что завершённое действие (или «*hot news perfect*», который относится к семантической зоне текущей релевантности), а с определёнными глаголами указывает на результирующее состояние завершённого действия. Следовательно, перфект не выступает как основная по-



вестовательная форма, т.к. его главная функция – подчёркивать отдельность определённого действия как свершившегося только что. В этом отношении рутульский язык отличается от большинства нахско-дагестанских языков, в которых главное значение перфекта – косвенная засвидетельствованность. Разумеется, это обстоятельство подчёркивает релевантность более подробных описаний употребления глагольных форм.

Семантика перфекта в нахско-дагестанских языках представляет собой многогранную и малоизученную тему. Рисунок показывает некоторый ареально-генетический кластеринг внутри семьи языков, однако для составления более точной ареальной картины необходимы более подробные сведения и сравнительные исследования употребления данных форм в разных языках, что в итоге позволяет установить происхождение таких грамматических категорий, как эвиденциальность, в данной языковой семье.

* * *

1. *Forker D.* Evidentiality in Nakh-Daghestanian languages // *The Oxford Handbook of Evidentiality*. – 2016. – P.52–68.
2. *Plungian V.A.* The place of evidentiality within the universal grammatical space // *Journal of Pragmatics* – 2001 – № 33. – P.351–352.
3. *Aikhenvald A.Y.* Evidentiality. – New York: Oxford University Press, 2004.
4. *Dahl Ö.* Tense and aspect systems. – Oxford: Basil Blackwell ltd, 1985; *Dahl Ö.* Tense and aspect in the languages of Europe. – Berlin; New York: Mouton de Gruyter, 2000.
5. *Плунгян В.А.* К типологии перфекта в языках мира: предисловие // *Казанский Н.Н.* Типология перфекта. – СПб.: Наука, 2016. – С.7–38.
6. Подробное описание семантического домена «текущей релевантности», или current relevance см.: *Ritz M.E.* Perfect tense and aspect // *The Oxford Handbook of Tense and Aspect*. – 2012. – P.881–907.
7. *Tatevosov S.G.* From resultatives to evidentials: Multiple uses of the Perfect in Nakh-Daghestanian languages // *Journal of Pragmatics*. – 2001. – № 33. – P.451.
8. *Майсак Т.А., Мерданова С.Р.* Категория эвиденциальности в агульском языке // *Кавказоведение*. – 2002. – № 1. – С.102–112.
9. *Махмудова С.М.* Грамматические категории и грамматические классы в рутульском языке: автореф. дис. ... д-ра филол. наук. – Махачкала, 2001.
10. *Alekseev M.E.* Rutul // *The indigenous languages of the Caucasus*. – 1994. – № 4. – P.213–258.



11. *Alekseev M.E.* Op. cit.
12. *Махмудова С.М.* Указ. соч. – С.146–147.
13. *Alekseev M.E.* Op. cit. – P.231.
14. *Ибрагимов Г.Х.* Рутульский язык. Синхрония и диахрония. – Махачкала: Народы Дагестана, 2004.
15. *Махмудова С.М.* Указ. соч. – С.146.
16. *Alekseev M.E.* Op. cit. – P.231.
17. *Ritz M.E.* Op. cit. – P.883.



Г.И. Фридман,

начальник отдела технологии

шрифта НПП «ПараТайп»

(г. Москва)

А.В. Королькова,

арт-директор

НПП «ПараТайп» (г. Москва)

ШРИФТ КАК ОСНОВА ЭЛЕКТРОННОЙ ПИСЬМЕННОСТИ

Современное общество невозможно представить без цифрового шрифта – основного инструмента коммуникации. В последние годы к классическим сферам использования шрифтов – книгам, газетам, навигационным указателям, плакатам, рекламным буклетам, упаковке и другим печатным материалам – добавились экраны разнообразных электронных устройств и Интернет. Производство цифровых шрифтов из ремесла превратилось в самостоятельную индустрию, во многом влияющую на нашу жизнь. Современные шрифты не только позволяют передавать содержание текста, но и способны наполнять его дополнительным смыслом и активно привлекать внимание.

Сегодня используется уже более 40 тыс. шрифтовых гарнитур (более 200 тыс. начертаний), и число их постоянно растёт. Новые шрифтовые форматы, их поддержка разработчиками операционных систем и браузеров существенно расширяют возможности шрифтов и, соответственно, сферу их применения. Однако, несмотря на постоянное развитие стандарта *Unicode*, языковая поддержка в подавляющем большинстве шрифтов оставляет желать лучшего. Чтобы оценить последний тезис, рассмотрим *доступные шрифтовые источники* и заодно уточним степень поддержки в них как минимум титульных языков Российской Федерации:

- шрифты, распространяемые под открытой лицензией;
- шрифты, поставляемые в рамках операционных систем;
- шрифты, предлагаемые коммерческими словолитнями.

Мы намеренно опускаем группу «бесплатных» шрифтов, распространяемых без лицензий. Как правило, они существенно ограничены по составу знаков и набору требуемых свойств, что значительно сужает возможности их использования.



Шрифты, распространяемые под открытой лицензией.

К первой группе, прежде всего, отнесём шрифты *SIL International* [1], распространяемые Международным некоммерческим институтом лингвистики под открытой (*OFL*) лицензией. Это небольшая группа шрифтов достаточно высокого качества, отличающихся большим составом знаков и дополнительной поддержкой языковых особенностей. Однако по составу они ограничены стандартом *Unicode*.

Другой источник – самая крупная на сегодняшний день библиотека шрифтов, распространяемых компанией *Google* под свободной лицензией. Это более 800 гарнитур, из которых 9,5% поддерживают стандартную кириллицу (в рамках кодовой страницы *MS 1251*), а 6,5% – расширенную кириллицу (в основном в рамках страницы *Cyrillic* стандарта *Unicode*). По оценкам экспертов, около 40% сайтов мировой сети используют этот шрифтовой сервис *Google* [2].

Третий, более близкий нам источник – супергарнитура *PT Sans*, *PT Serif*, *PT Mono*. Подробная информация о суперсемействе, а также актуальные версии шрифтовых файлов представлены на сайте Паратаип [3]: «Шрифты, помимо стандартного набора знаков для работы с языками Западной и Восточной Европы, а также комплекта стандартной кириллицы, включают знаки всех алфавитов титульных языков РФ, что делает их уникальным и чрезвычайно важным инструментом современных компьютерных коммуникативных сред».

Современный дизайн, наличие наряду с базовыми узких и мелкокегельных начертаний, высокое экранное качество, широкий знаковый состав и поддержка языковых особенностей – всё это определило активное распространение и использование *PT*-шрифтов. К примеру, только с помощью сервиса *Google* базовыми начертаниями *PT Sans* пользуются более 30 тыс. сайтов.

Однако только эти шрифты не могут обеспечить все актуальные потребности. Нерешённым остаётся вопрос создания специализированных начертаний для школьных учебников, навигации, интерфейсов и т.д.

Шрифты, поставляемые в рамках операционных систем.

Шрифты этой группы можно использовать «в личных целях», а при коммерческом использовании могут возникать некоторые ограничения. В их число входят знакомые многим *Arial*, *Helvetica*, *Times New Roman*, а также многие другие шрифты, подготовленные для разных операционных систем. Присутствуя на большинстве компьютеров, обладая представительным знаковым составом, эти шрифты широко востребованы. Как правило, их состав знаков соответствует стандарту *Unicode* и не поддерживает все титульные языки РФ. Однако в последнее время появляются и исключения.



Помимо набора гласных русского языка с ударениями, кириллических гласных с долготами и дополнительной диакритикой, башкирских, болгарских и сербских альтернативных форм, в новых операционных системах *Apple* постепенно проявляется поддержка целого ряда дополнительных знаков, не входящих в стандарт *Unicode*. При этом используются определённые в Паратаип коды из пользовательской зоны (*PUA Unicode*). Таким образом, *Mac OS*, начиная с версии *Sierra*, будет более полно поддерживать следующие языки: ительменский, кильдинский, нивхский, селькупский, чеченский, эвенкийский, эвенский, хантыйский.

В связи с этим стоит напомнить, что *PT Sans* и *PT Serif* входят в базовый комплект шрифтов, поставляемых с *Mac OS*, начиная с версии *X*, и рекомендованы для использования в *Windows* (начиная с версии *Vista*) компанией *Microsoft*. Учитывая подобный интерес, можно сделать вывод о том, что после включения недостающих знаков расширенной кириллицы в *Unicode* системные шрифты довольно быстро станут поддерживать титульные языки РФ. Это утверждение актуально и для ряда знаков акцентированной латиницы, необходимой для поддержки нескольких титульных языков РФ, использующих латинскую письменность [4].

Отдельного внимания заслуживает опыт формирования комплекта шрифтов создателями отечественных операционных систем.

Так, в 2016 г. совместно с разработчиком операционной системы *Astra Linux*, НПО РусБИТех, Паратаип представил новые общедоступные шрифты *PT Astra Sans* и *PT Astra Serif*, разработанные на основе *Pt Sans* и *PT Serif*, модифицированные таким образом, чтобы они подходили под стандарты оформления документов, эталоном ёмкости для которых является *Times New Roman* [5]. В рамках операционной системы шрифты уже используются в делопроизводстве ряда министерств и ведомств, а также в некоторых областях РФ.

Шрифты, предлагаемые коммерческими словолитнями.

Сразу отметим, что к этой группе относятся в основном коммерческие шрифты, цены на которые варьируются в достаточно широком диапазоне и зависят от лицензионной политики конкретной словолитни.

Учитывая тот факт, что около 5% населения нашей планеты используют кириллическую письменность, многие словолитни мира выпускают новые шрифты с поддержкой кириллицы. В большинстве случаев это стандартная кодовая страница *MS 1251*. В стандартах некоторых крупных словолитен расширение кириллицы касается и казахского языка. Некоторые небольшие студии предлагают свои стандарты, ограничивая поддержку языков по числу пользователей (самое распространённое ограничение – 1 млн человек), поскольку подготовка и поддержка дополнительных знаков



шрифта приводит к существенным затратам. Проанализировав возможности разрешения проблемы шрифтового многообразия для младописьменных языков [6], *мы констатируем необходимость участия в её решении государственных органов РФ.*

И тем не менее сколько шрифтов всё-таки нужно?

Здесь необходимо обратиться к современной типографике, которая даёт чёткое представление о назначении шрифта. Для различных носителей информации и разных сфер применения лучше всего подходят шрифты с определёнными параметрами, такими как пропорции знаков, апертюра, контраст, насыщенность, причём эти параметры могут быть выражены в конкретном цифровом виде. Есть многочисленные примеры шрифтов, разработанных специально для решения конкретной задачи и имеющих наилучшие для этого параметры. Однако в подавляющем большинстве случаев это примеры шрифтов латинской графической основы, наиболее развитой в этом отношении. Шрифтовая палитра, доступная отечественному пользователю, несмотря на её современное разнообразие, испытывает недостаток шрифтов для некоторых конкретных целей, например для навигации и указателей или для экранного интерфейса. Кроме того, есть области применения, для которых можно найти всего 1–2 кириллических шрифта, зачастую являющихся адаптацией существующих латинских. Права на такие шрифты, как правило, принадлежат западным компаниям. Ещё одной важной проблемой является неочевидность выбора шрифта для той или иной цели. Даже графические дизайнеры часто не знают, какой шрифт в большей степени подойдёт для решения стоящей перед ними задачи. Как следствие – важная информация воспринимается конечным пользователем хуже, чем могла бы.

Таким образом, при определении необходимого набора требуемых обществу шрифтов необходимо исходить именно из функциональных требований к ним, ясно обозначив сферу применения. Это позволит, во-первых, применять для конкретных целей шрифты с наиболее подходящими параметрами; во-вторых, снизит зависимость от зарубежных разработчиков и правообладателей; в-третьих, явное обозначение областей применения поможет пользователю точнее решать поставленные задачи.

На наш взгляд, с точки зрения функциональной востребованности шрифтов необходимо решить 5 важных для отечественного пользователя задач:

1. **Шрифт для книг.** Минимальное количество начертаний – 2 (нормальное, курсивное). Сфера применения – набор основного текста в художественной литературе, учебниках, детских книгах, электронных книгах (с высоким разрешением экрана). Основные параметры: шрифт с засечками (антиква между переходной и шотландской), достаточно широкие пропор-



ции, средние межбуквенные расстояния, средняя высота строчных знаков, средняя длина выносных элементов. Кроме поддержки государственного и титульных языков РФ (базовый состав знаков), необходимы следующие дополнения: минускульные цифры, значительный набор знаков препинания и пробелов. Ближайшие аналоги: гарнитура Школьная (*Schoolbook*), шрифты *Charter* и *Georgia*.

2. Шрифт для новостей. Минимальное количество начертаний – 2 (нормальное, полужирное). Сфера применения – набор основного текста в периодических изданиях, корпоративных стилях и т.п. Основные параметры: шрифт с засечками (малоконтрастная антиква), узкие пропорции, большие межбуквенные расстояния, большая высота строчных знаков, короткие выносные элементы. Знаковый состав – базовый. Ближайшие аналоги: *Swift*, *Arnhem*, *Breve*.

3. Навигационный шрифт. Минимальное количество начертаний – 2 (нормальное и инвертированное). Сфера применения – навигация, ориентирование, дорожные знаки. Основные параметры: шрифт без засечек (открытый нейтральный гротеск), широкие пропорции, большие межбуквенные расстояния, большая высота строчных знаков, средние выносные элементы. Знаковый состав – базовый. Ближайшие аналоги: *ALS Direct*, *PT Sans Caption*, *Frutiger*, *FreeSet*.

4. Шрифт для интерфейса. Минимальное количество начертаний – 2 (нормальное и полужирное). Сферы применения: интерфейс и текст на экране, операционные системы, сайты и т.п. Основные параметры: шрифт без засечек (нейтральный гротеск), узкие пропорции, большие межбуквенные расстояния, большая высота строчных знаков, короткие выносные элементы. Знаковый состав – базовый. Ближайшие аналоги: *Lucida*, *San Francisco*, *Roboto*.

5. Моноширинный шрифт. Минимальное количество начертаний – 2 (нормальное, полужирное). Сферы применения: набор кода при программировании, набор электронных таблиц. Основные параметры: моноширинный шрифт без засечек, большая высота строчных знаков, короткие выносные элементы. Знаковый состав: кроме базового знакового состава, необходим дополнительный набор спецсимволов. Ближайшие аналоги: *Input*, *Consolas*.

Все шрифты должны быть подготовлены как для печати на бумаге, так и для полноценной работы на экране, в том числе низкого разрешения.

При наличии такого комплекта шрифтов отечественный пользователь будет иметь инструмент для решения большинства функциональных задач. Однако его создание потребует значительных усилий и затрат. Поэтому для того, чтобы сделать функциональный комплект шрифтов не коммерческим, а общедоступным, необходима государственная поддержка.



* * *

1. URL: <http://scripts.sil.org/cms/scripts/page.php>
2. URL: <https://fonts.google.com>
3. URL: <http://paratype.ru/public>
4. Языки меньшинств в компьютерных технологиях: опыт, задачи и перспективы // Материалы международной конференции. – Йошкар-Ола, 2011. – С.96.
5. URL: <http://www.paratype.ru/cinfo/news.asp?NewsId=3469>
6. Якунов Э.Л. Алфавиты современной России // Три столетия русского гражданского шрифта (1708–2008): сборник. – М.: Пашков дом, 2008. – С.173.



Е.В. Хазова,

*переводчик Бюро официального
перевода Центра инновационных
языковых технологий
ГОУ ВО КРАГСиУ*

Л.Д. Матвеева,

*редактор Бюро официального
перевода Центра инновационных
языковых технологий
ГОУ ВО КРАГСиУ*

О.Л. Исакова,

*ведущий редактор Бюро официального
перевода Центра инновационных
языковых технологий
ГОУ ВО КРАГСиУ*

ФУНКЦИОНИРОВАНИЕ КОМИ ЯЗЫКА В ОФИЦИАЛЬНЫХ ИНТЕРНЕТ-ИСТОЧНИКАХ

Современный коми язык проявляет себя в разных областях жизни общества: это и повседневный язык общения, и литературные произведения, и официальные документы, и информация юридического и общественно-значимого характера на телевидении, радио и в Интернете. С 1992 г. коми язык является одним из государственных языков в Республике Коми. В связи с этим цель нашего исследования – проанализировать функционирование современного коми языка в официальных интернет-источниках. На сегодняшний день это важный фактор для анализа. Язык живёт и развивается, если общество активно использует его в повседневной жизни и профессиональной деятельности.

Коми язык как государственный присутствует в Интернете на официальных сайтах государственных органов власти, учреждений, общественных объединений. Это обусловлено положениями Федерального закона «Об информации, информационных технологиях и защите информации» от 27.07.2006 № 149-ФЗ [1], в соответствии с которым государственные органы и органы местного самоуправления представляют официальную информацию о своей деятельности на русском языке и на государственном языке соответствующего субъекта в составе Российской Федерации.



Основными республиканскими сайтами, на которых размещается общественно значимая информация на русском и коми языках, являются официальный интернет-портал Республики Коми (<http://rkomi.ru>), сайт Государственного Совета Республики Коми (<http://gsrk.ru>) и сайт Министерства национальной политики Республики Коми (<http://minnats.rkomi.ru>).

Официальный портал Республики Коми функционирует на двух языках, новостная лента обновляется ежедневно как на русском, так и на коми языке. Это даёт возможность читателям, жителям республики своевременно и в доступной форме получать необходимую информацию. Но следует отметить и тот факт, что версия сайта на коми языке существенно отличается от русской. Она представлена в более сокращённом варианте, например, в ней отсутствуют блоки о правительстве, судебной системе, приоритетных национальных проектах, деятельности Антитеррористической комиссии в Республике Коми, о том, какие вопросы чаще всего задают жители республики, о праздновании 95-летия Республики Коми, переселении граждан из аварийного жилья. Также отсутствуют блоки, в которых размещается информация о деятельности Правительства республики, стратегии развития региона до 2020 г., Финно-угорском этнокультурном парке, государственном и муниципальном заказе, а также ряд других блоков («Электронное правительство», «Перевод государственных и муниципальных услуг в электронный вид», «Сетевое издание “Перечень правовых актов...”», «Журнал “Бюллетень органов государственной власти Республики Коми”», «Мониторинг эффективности деятельности органов местного самоуправления», «Книга Почёта Республики Коми», «Доска Почёта Республики Коми «Лучший в профессии», «Книга памяти Республики Коми, «Объявления», «Типовые условия использования общедоступной информации...»).

Для социально значимых сайтов важным фактором является своевременное размещение и обновление информации. Подробный анализ текстов в имеющихся разделах на коми языке обнаруживает устаревшую информацию: название раздела «Администрация Главы Республики Коми» в старой редакции – «Администрация Главы и Правительства Республики Коми»; в подразделе «Депутаты Государственной Думы» нет информации об О.В. Савастьяновой; в разделе «Местное самоуправление» обнаруживается информация о том, что руководителем администрации городского округа «Сыктывкар» является Р.В. Зенищев (на коми языке текст звучит таким образом: «Сыктывкар» кар кытшлön администрацияса юралысь – Роман Валерьевич Зенищев).

Большое количество событий происходит в спортивной и культурной жизни республики, в финно-угорском мире и других сферах. Информирование читателей интернет-портала о них также важно, однако в под-



разделе «Съезд коми народа» отсутствует информация о X и XI съездах, в подразделе «Международные связи» раздела «Финно-угорский мир» отсутствует информация о VII Всемирном конгрессе финно-угорских народов 2016 года.

В современном мире для пользователей интернет-ресурсов наиболее важным фактором, кроме своевременного обновления и достоверности информации, является функциональность и доступность ресурса. Нужно отметить, что для пользователей, читателей интернет-портала на сайте сделана кнопка «Версия для печати», которая позволяет им сохранить нужную информацию. Но в то же время пользователь испытывает определённые трудности при поиске самой информации. Например, некорректно отображается раздел «Архив новостей» – коми или англоязычная версия сайта автоматически переходит на русскоязычную. При открытии новостей и просмотре новостной ленты в одном блоке с опубликованными новостями открываются и анонсы мероприятий. Конечно, поиск облегчает календарь мероприятий, размещённый в разделе «Информация о мероприятиях», но в то же время данный календарь находится внутри раздела, что приводит к большой трате времени на поиск указанного календаря, а впоследствии и на поиск интересующей информации. Возможно, размещение календаря мероприятий отдельным блоком сделало бы поиск более удобным. Необходимо отметить, что использование буквы Ё и I из латинского сегмента Юникода вместо кириллических приводит к некорректной передаче слов и невозможности поиска, например: КОМИ РЕСПУБЛИКАСА ВÖР-ВА ОЗЫРЛУН (http://rkomi.ru/km/left/info/prir_res).

Сайт **Государственного Совета Республики Коми** (<http://gsrk.ru>) также предоставляет пользователям общественно значимую информацию. С точки зрения поиска информации он доступен и функционален. При переходе с русского на коми язык сохраняется найденный раздел, выделен календарь мероприятий, что также упрощает поиск информации. По своей структуре версии сайта на русском и коми языках соответствуют друг другу, ежедневно обновляется новостная лента, но всё же некоторые блоки в комиязычной версии отсутствуют. Например, нет блоков, в которых речь идёт об изданиях Государственного Совета, противодействии коррупции, закупках товаров, работ, услуг, Контрольно-счётной палате Республики Коми, Общественной палате Республики Коми, Молодёжном парламенте. Рассмотрев разделы сайта, можно обнаружить и некоторое несоответствие версий на русском и коми языках. К примеру, в разделе «Законотворчество» («Оланпастэчас») в русском варианте 5 подразделов, в коми – всего лишь 3; в разделе «Деятельность» («Каналан Сöветлöн удж») в русской версии – 6 подразделов, в коми – 2. Также несоответствие наблюдается



в таких разделах, как «Правовые основы деятельности» («Уджлөн инёда подувьяс»), «Состав и структура» («Пырысьяс да тэчас»), «Пресс-центр» («Юбртан шöрин»). К сожалению, на сайте имеется и устаревшая информация, касающаяся раздела «Контакты» («Йитöдьяс»). В качестве примера можно привести небольшой фрагмент. В русской версии руководителем Секретариата Председателя Государственного Совета Республики Коми и его заместителей является Ю.В. Наговицына, а в коми версии – О.К. Сычугова. На коми языке данный текст звучит таким образом: «Каналан Сöветса Веськöдлысьлөн да сійöс вежысьяслөн секретариатса юрнуöдысь – Ольга Константиновна Сычугова» (<http://gsrk.ru/pages/communications.html>). Информация данного раздела важна для граждан, обращающихся в Государственный Совет. Кроме того, устаревшая информация размещена и в подразделе «Избирательные округа» («Бöрйысян кытшьяс») раздела «Общие сведения» («Подув»). В русской версии опубликовано Постановление Государственного Совета Республики Коми «О схеме одномандатных избирательных округов для проведения выборов депутатов Государственного Совета Республики Коми V созыва» от 19.02.2015 № V-9/16, а в коми версии – Постановление Государственного Совета от 11.11.2010 № IV-8/59. Текст звучит следующим образом: «Коми Республикаса Каналан Сöветлөн депутатъясöс бöрйысьöмьяс дырйи öти мандата бöрйысян кытшьяс. Коми Республикаса Каналан Сöветлөн 2010 во вöльгым тöлысь 11 лунся IV-8/59 №-а Шуöм дорö содтöд» (<http://gsrk.ru/pages/constituencies.html>).

Сайт **Министерства национальной политики Республики Коми** функционален с точки зрения использования, поиска информации, имеет версию на коми языке, правда, состоящую из одного раздела «О министерстве» (<http://minnats.rkomi.ru/km>). Новостная лента на коми языке обновляется не регулярно. На момент написания статьи последняя новость была опубликована 12 декабря 2016 г. под заголовком «В 2017 году в Коми пройдёт VI съезд финно-угорских народов России», на коми языке – «2017 воын Комиын лоас Россияса финн-йöгра войтырлөн чукöртчылöм» (<http://minnats.rkomi.ru/km>).

По нашим исследованиям, из всех **муниципальных районов и городских округов Республики Коми** комиязычную версию официального сайта имеют только 2 района: МР «Печора» (<http://www.pechoraonline.ru/kmi>) (переведено 50% информации) и МР «Сыктывдинский» (<http://syktyvdin.ru/km/static/page>) (переведено 40% информации).

Если говорить об **общественных объединениях**, нужно отметить официальный сайт межрегионального общественного движения «Коми войтыр» (<http://komivoityr.com>). Ранее вся общая информация и новостная



лента здесь была представлена на русском и коми языках соответственно. Однако на момент проведения мониторинга (январь 2017 г.) сайт был закрыт.

Свой комиязычный сайт создали педагоги **Государственной общеобразовательной школы-интерната «Гимназия искусств при Главе Республики Коми» имени Ю.А. Спиридонова** (<https://sites.google.com/site/moucitejkomiazykailiteratury/home>). Из представленной здесь информации на коми язык переведена одна четверть. У прочих общественных объединений версии сайтов на коми языке отсутствуют.

Для улучшения ситуации предлагаем муниципальным образованиям создать адекватные комиязычные версии официальных республиканских и районных сайтов, которые будут регулярно пополняться актуальной информацией о важнейших событиях в жизни района или городского округа. Государственному Совету Республики Коми, Администрации Главы Республики Коми, Министерству национальной политики, тем муниципальным районам, которые имеют коми версии официальных сайтов, желательно пересмотреть комиязычные разделы и привести коми версии в соответствие с русскими. Наиболее эффективным способом решения данной задачи является, по нашему мнению, привлечение специалистов – коми филологов.

Считаем, что живой коми язык должен присутствовать и активно функционировать во всех областях жизни общества, чего можно достичь благодаря электронному пространству, в частности официальным интернет-источникам, в которых публикуется самая актуальная общественно значимая информация.

* * *

1. Об информации, информационных технологиях и защите информации: федеральный закон от 27.07.2006 № 149-ФЗ // Собр. законодательства РФ. – 2006. – № 31 (ч.1). – Ст.3448.



УДК 81:004

А.А. Васильева,
*доцент кафедры общих
 проблем управления механико-
 математического факультета
 МГУ имени М.В. Ломоносова
 (г. Москва)*

КЫДЗИ ВЕЛӨДНЫ ТӨДТӨМ КЫВ ИНТЕРНЕТ ОТСӨГӨН (АС ОПЫТ СЕРТИ)*

Позьб-ө велөдны төдтөм кыв сөмын интернет отсөгөн? Тайö юалöмыс майшөдлө унабс. Ме видлі тайöс вöчны и велөдны коми кыв. Та йылысь и кöсья неуна висьталыштны – кызди ме велөді коми кыв. Ме дженъыда пасья велөдчан этапяс йылысь.

Велөдөм водзын төді сөмын некимын кыв, коми төдсаяс эз вöвны. Босьті öтувезйысь Е.А. Цыпановлысь «Коми кыв» небöгсö [1]: тайö вöлі медводдза гижөд, кодöс сетіс *Google*. Небöгас 20 урок.

Кöч тöлысь чöж лыдди медводдза 8 урок. Корси «Контактын» социальной везйысь коми кыв велөдан группас да аддзи «Ме велөда коми кыв – Я учу коми язык» группасö. Ыстöгъяс пыр аддзи «Сёрни-басни» группасö, пыри и сэтöн практикала.

Йирым тöлысь чöж лыдди нöшта 5 урок. Босьтчи уна практикуйтчыны öтувезйын: вуджөдны дженъыд текстяс, гижны комментарийс комиöн. Казялі, мый кужа нин гөгөрвоны юбртöмъяс, дженъыд биографияс.

Вöльгым – öшым тöлысьяс чöж ковмис уна уджавны да, лыдди сөмын кык урок. Вöльгым медводдза лунъяснас аддзи öтувезйысь Михаил Лебедевлысь «Яг Морт» поэмасö, лыдди сійöс электроннöй кывкуд отсөгөн. Бара уна практикуйтчи социальной вез пыр. Öшым тöлысь помсянь босьтчи отсавны школаса велөдчысьяслы «Ме велөда коми кыв» группаас.

Төвшөр тöлысь чöж лыдди медбөръя вит урок (медбөръя уджыясыс кындзи). Аддзи важ математика велөдчан небöгъяссö, велөді коми латинича да важ математика терминология, орччөді сійöс выль кывкудыясыс терминологиякөд. (Сэні вöліны кыв вожалöмъяс да ми Öнью Лавкөд да Сергей Елфимовкөд босьтчим тэчны ассыным терминология.) Лыдди коми кывбуряс да блогяс.

© Васильева А.А., 2017

* Печатается в авторской редакции.



Урасьом тӧлысьнас вӧчи медбӧръя уджъяссӧ. Та бӧрын водзӧ лыд-
дыси да лӧсьӧдӧ математика терминология.

Скайп менам абу да, комиӧн кужа сӧмын гижны-лыддысьны.
Позьӧ-ӧ велӧдны тадзи выль кыв? Ме вочавидза – да, позьӧ.

Пасйӧд. Менсьым текст редактируйтӧс Ӗньӧ Лав.

* * *

1. *Цыпанов Е.А.* Коми кыв. Самоучитель коми языка. – Сыктывкар:
Коми книжное изд-во, 1992.



Е.А. Макарова,
*переводчик Бюро официального
перевода Центра инновационных
языковых технологий
ГОУ ВО КРАГСиУ*

Е.В. Кожевина,
*переводчик Бюро официального
перевода Центра инновационных
языковых технологий
ГОУ ВО КРАГСиУ,
аспирант Институт языка,
литературы и истории
Коми НЦ УрО РАН*

БЛОГОСФЕРАЫН КОМИ КЫВЛӨН ПЕТКӨДЧӨМ*

Быд ловъя кыв сөвмө олöмыскөд тшöтш. Ф. Де Соссюр гижö: «Нельзя найти такого языка, который находился бы в состоянии покоя и неподвижности» [1]. Сидзкö, медым кывйыс олiс, сылы колö вежсьыны олöмыскöд тшöтш, а медым олöмын вежсьöмъяс петкöдчисны кывворын, сийöс колö пыртны олöмлөн став юкөнö. Öнiя олöмын зэв тöдчана юкөнөн лööтуввез. Коми кывйөн кутiсны гижны контактын, мукöд сайт вылын, википедияын статьяяс йöзöдöны комиён. Талун коми кыв вылын нуöдöны унакодö блог. Тайö статьяын ми видлалам коми блогосфера артмöдöм, блогосфераын коми кывйөн вöдитчöм да сетам вöзиймъяс коми блогъяс сöвмöдöм серти.

Медводз пасьям, мый этшöм блогыс, блогерыс да блогосфераыс. Öтуввез терминъяс кывкудйын шуöма, мый блог – тайö «сетевой дневник одного или нескольких авторов, состоящий из записей в обратном хронологическом порядке. Слово произошло от английского weblog – «веб-журнал». С помощью сервиса блогов можно создать свой онлайн-дневник, читать и комментировать дневники других пользователей, принимать участие в сообществах на интересные темы и создавать свои сообщества. Совокупность блогов называют блогосферой» [2]. Мöд кывкудйыс видзöдлам, кодöс шуöны блогйысьён. «Блоггер – это владелец личного блога... Он обладает всеми правами на то, чтобы удалить, изменить или

© Макарова Е.А., Кожевина Е.В., 2017

* Печатается в авторской редакции.



исправить содержимое страниц своего веб-ресурса. Только он знает все ключи и пароли для доступа к своему сайту...» [3].

Медводдза юалом: мыйсянь заводитчис коми блог нуёдмыс?

Дзик неважён на Россияса финн-угор отуввезйын чужан кывъя блогыс вöлі сöмын удмуртъяслөн. Удмуртияс том йöз заводитисны гижны отуввезйын да нуöдны сетевöй дневникъяс да журналъяс удмурт кыв вылын. Мариечьяс тайö уджсьыс оз жö кольччыны. Мари блогер Елена Кукина весиг нуöдiс мари блогерьяслыс школа, мый бöрын артмис куим мари блог. Коми блогерьясöс та вылö эз велöдны, но и найö кутисны нуöдны блогъяс чужан кыв вылын [4].

Медводдза коми блогьясыс вöліны коми фольклор туялыс Алексей Рассыхаевлөн, кодi заводитис нуöдны блогсö 2010 воын LiveJournal (Живой журнал) платформа подув вылын, да Коми Республикаын Россияса МЧС-лөн медшöр веськöдланiнын пресс-службаса уджалыс Елена Лодыгиналөн («Зырянка»), кодi аслас блогын юöртiс коми кывйöн сы йылыс, мый колö вöчны виччысьтöм лоöмторьяс дырйи (ытва, пöжарьяс, терактыяс дырйи...). Е. Лодыгина заводитис гижны аслас журналын 2011 вося урасьöм тöлысьын, но талун кежлö дугдiс гижны аслас блогын, а А. Рассыхаев ассьыс блогсö нуöдö водзö [5].

Чужан кывйöн сiдзжö заводитис гижны коми лингвист Öньö Лав да ышöдiс та вылö ассьыс ёртьясö да тöдсаясö. Сылөн кывьяс серти, удмуртъяс вылö видзöдöмөн сiйö мöвпыштiс, кызди позьö ышöдны комиясöс лöсьöдны ассьыныс блогьяс, да думыштiс, мый медводз лöсьöдас ассьыс блог, а сэсся нин öддзöдас ёртьясö. Сылөн кывьяс серти, медводз сiйö вöзйöма Алёна Нестеровалы («Тувсовья») нуöдны блогсö комиöн. Та бöрын нин чужöма уна коми блог [6].

Мöд юалом: кутшöм технической позянлуньяс эмöсь сы вылö, медым нуöдны татшöмсяма блогьясö?

Öнi эм уна технической позянлун, медым лöсьöдны ассьыд блог. Оз нин ков аслыд вöчны кутшöмкö программа. Эмöсь сiдз шусяна платформаяс, мыйяс отсöгöн блогйысьяс лöсьöдöны ассьыныс блогьяснысö. Платформа отсöгöн позьö бöрйыны дизайн да тема. Со сöмын некымын платформа: *Blog.com*, *WordPress.com*, *LiveJournal.com*, *Blogspot.com*, *Weebly.com*.

Отуввезйын бура гөгöрвöдöны, кызди лöсьöдны, нуöдны блогьяс да кызди оформитны найöс: вöзйöны выль шаблоньяс, юöрсö йöзöдан ногьяс, вежлавны дизайн да эффектъяс. Тайö абу сьöкыд технической удж, кодöс позьö вöчны конструктор отсöгöн [7].

Коми блогйысьяс тшöкыджыка вöдитчöны *Blogspot.com* платформаöн, но эмöсь и найö, кодьяс лöсьöдöны блогьяс *LiveJournal.com*



платформа отсөгөн. Тайос позьо аддзыны коми блогйысьяслөн отувья списокусь, кодос йозьодма «ВКонтакте» социальной везлөн «Коми блогйысь» котырын [8].

Но медся тодчана, мый öнi эм позянлун гижны комиöн. Та выло лосьодма коми раскладка. Компьютерö раскладка пуктанногö да сийöн вöдитчанногö позьо аддзыны <http://wiki.fu-lab.ru/index.php> ыстог серти. Öнi «ö» да «i» шыпасьяслы позьо бöрийны торья клавишася да коми кывья юöрсö лоас кокныдджык аддзыны сы понда, мый тайö шыпасьяс-сö пыртöма юникодö.

Коймöд юалöм: кодьяс нуöдöны коми кывья блогьяс, мый йылысь гижöны?

Ёнджыкасö коми кыв вылын блогйысьöны Сыктывкарса университет помалысьяс, кодьяс босьтисны коми филологиясь тöдöмлуньяс, кызвыннас тайö гижысь войтыр, эм некымын велöдысь. «Коми блогйысь» котырын пасйöма коми кывья блогьяс выло ыстöдыас [9].

Блогьяс серти тыдалö, мый блогйысьяс – восьса, нинöмön абу дзескөдöм йöз. Найö гижöны сы йылысь, мый налы окота: олöм, асланысьöлöмкылöмьяс йылысь. Найö весиг вермöны ворсыштны кыв нормаясöн, пыртны выль кывьяс.

Сиджö ми аддзам, мый блогьясö кызвыннас нуöдöны аняяс. Медьёна зильöны Евгения Лобанова, Елена Рассыхаева, Ксения Мартюшова, Наталья Дёмина, Алена Нестерова.

Мукöд блогйысьяс сетöны ним буретш блогыслы, шуам, Елена Афанасьевалөн «Куим ю» блог, Öнб Лавлөн «Арся» блог. А мукöдыс нимтöны блогьясö асланыс псевдонимьяс серти: Елена Хозяиновалөн «Кебраса» блог, Елена Лодыгиналөн «Зырянка» блог.

Блогйысьяс гижöны асланыс удж йылысь («Ме – Танюк», «Кола» блогьяс), коми культура да фольклор сöвмöм йылысь (Алексей Рассыхаев), олöмысь казтылöмьяс («Арся», «Тусь», «Кебраса», «Асывыв Европатi шöйтöм», «Алёна Тувсовья»), вөр-ва да сийöс видзöм йылысь («Ставыс лоас бур», «Висервожса ныв»). Унаöн аслас блогьясын йöзöдöны кывбурьяс («Алёна Тувсовья», «Кывбурьяс», «Куим ю», «Изьваса ныв», «Менам видзöдлас: кывбурьяс», «Эжва йывса» блогьяс), вичко да эскöм йылысь («Миян вичко» блог). Поводдя да виччысьтöм лöмторьяс йылысь юбр позьо аддзыны Елена Лодыгиналөн блогысь. Блогьясын сиджö петкöдчö общественно-политическöй олöм (Вячеслав Попов, «Ставыс лоас бур»). Быд блогын сийö либö мöд ногön тыдовтчö коми кыв видзöм-сöвмöдöм понда майшасьöм.

Блогьясын петкöдчö, кызди сöвмö ловья коми кывйыс. Öнiя олöмысь тыдалö, мый кывйыс кутас овны, кутам кö ми сйөн вöдитчыны отуввезйын, лосьодны телепередачаяс, лэдзны дискьяс, сёрнитны быд лун и быдлаын.



Бөрья юалом: мый позьб вочны, медым видзны да сөвмөдны коми блогосферасө, мый вочома нин?

2016 вося вольгым тöлысын нуодісны республикаса «Коми блогйысь» конкурс. Конкурслөн шөр могыс – велөдны коми йöзöс комион гижны öтуввезйын, корсьны выль позянлунъяс коми кывйöн вöдитчöм выльö. Тайö отсалас сөвмөдны коми кыв да культура, а йöзыс асланыс блогъясын вермасны петкөдлыны ассьыныс енбинысö, уджнысö, видзöдласьяснысö, сёлөмкылöмьяснысö. Тайö конкурс йылысь «Юрган» телеканал вылын сёрнитіс А. Рассыхаев [10]. Татшөм юөртöмыс вермас ышöдны мукöд коми йöзöс лöсьöдны ассьыныс блогъяс.

Сіджö А. Рассыхаев аслас сёрниын тöдчöдіс, мый «Коми блогйысь» конкурсö котыртöма татшөм жö сям мари конкурс серти [11]. Мари кывъя блогъяс серти тайö конкурсыс эз нин öтчыдысь вöвлы. Налөн конкурсыс шусьö «Звёзды Марнета» [12].

«Коми блогйысь» конкурс серти жюри видлаліс, мый йылысь да кутшөм кывйöн гижöны статьяс, да видзöдісны, мед эськö эз вöвны орфография, грамматика, пунктуация боксянь тырмытöмторъяс. А сіджö, кызди юөрсö оформитöма лист бок вылын: кутшөм фото да видеоматериал мыччöма [13]. И со Коми культура шöринын шуисны медбур блогйысьяслысь нимъяссö. Кольöм во вöлі бөрийöма 5 номинация:

1. Донъялана (аналитикаа) медбур да колана гижöд.

2. Торъя лөөмтор йылысь медбур гижöд.

3. Коми му артмöмсянь 95 воы сиöм чужанін радейтöм йылысь медбур гижöд.

4. Медзіль коми блогйысь.

5. Медбура мичмөдöм блог.

«Донъялана (аналитикаа) медбур да колана гижöд» номинацияын «Муслун» методика» гижöдысь вермысьнас шуöмаöс Евгения Лобановаöс, «Торъя лөөмтор йылысь медбур гижöд» юкöнын – «Тöдысьяс да ми» пасйöдысь Наталья Дёминаöс, «Коми му артмöмсянь 95 воы сиöм чужанін радейтöм йылысь медбур гижöд» номинацияын – бара жö Евгения Лобановаöс «Бабö, благсöвит менö!» гижöдысь. «Медзіль коми блогйысьöн» жюри шуöма Алёна Нестероваöс, коді нуөдö «Алёна Тувсовья» блог. «Медбура мичмөдöм блог» номинацияын вермöма Елена Рассыхаева [14].

2017 воын тайö конкурсö көсйöны нуöдны водзö. А мед унджыкөн лыддясны коми блогысь гижöдъяссö, быд тöлысь медбур пасйöдъяссö ку-тасны йöзöдны тшöтш «Коми му» да «Йöлöга» газетьясын [15].

Коми кывъя блогйысьяслы позьб сетны вöзйöмьяс. Öти-кö, блог нуöдысьяслы колö ышöдны том йöзöс лöсьöдны öтуввезйын ассьыныс



коми кывъя блогъяс, медым тадзи видзны да сѳвмѳдны коми кывсѳ. Мѳд-
кѳ, вѳзйыны блог нуѳдысьяслы видзѳдны да комментируйтны мукѳ коми
блог, а сѳдзжѳ аслыныс зѳльны быдсяма конкурсьяс котыртѳмын, нуѳдѳмын
да сѳні участвуйтѳмын.

Кывкѳртѳд пыдди позьѳ шуны, мый коми кывъя блогъяссѳ дась-
тѳмлы да нуѳдѳмлы подулыс пуктѳма. Гижѳны танѳ том и верстьѳ йѳз
олѳмлѳн уна лоѳмтор, асланыс сьѳлѳмкылѳмъяс, мукѳд йѳзлы вѳзйѳмъяс
йылысь. Уна сикас тема вылѳ гижѳм сетѳ позянлун пыртны ѳтуввезйѳ да
паськыда вѳдитчыны олѳмлѳн разнѳй юкѳнын вѳдитчан кывворѳн.

* * *

1. *Фердинанд де Соссюр*. Заметки по общей лингвистике / пер. с фр.;
общ. ред., вступ. ст. и коммент. Н.А. Слюсаревой. – М.: Прогресс,
2000. – С.47.

2. Словарь терминов. Блог. – URL: <http://your-hosting.ru/terms/b/blog/>
(дата обращения: 20.01.2017).

3. Кто такой блоггер – отрывок из энциклопедии. – URL: <http://www.workformation.ru/kto-takoj-blogger-listaem-enciklopediyu.html> (дата
обращения: 20.01.2017).

4. *Сергеева Л.* Анбур – блогосфера. – URL: <http://uralistica.com/profiles/blogs/anbur-blogosfera> (дата обращения: 20.01.2017).

5. *Сергеева Л.* Указ. соч.

6. *Сергеева Л.* Указ. соч.

7. Как создать свой блог в Интернете. – URL: <https://uguide.ru/kak-sozdat-svoj-blog> (дата обращения: 20.01.2017).

8. Коми блогйысь. – URL: <https://vk.com/club78645075> (дата обра-
щения: 20.01.2017).

9. Коми блогйысь...

10. URL: [https://vk.com/impeers=362663497&sel=140138665&z=](https://vk.com/impeers=362663497&sel=140138665&z=video11948594_456241760%2F283b520181bb3a105e)
[video11948594_456241760%2F283b520181bb3a105e](https://vk.com/impeers=362663497&sel=140138665&z=video11948594_456241760%2F283b520181bb3a105e) (дата обращения:
20.01.2017).

11. URL: [https://vk.com/impeers=362663497&sel=140138665&z=](https://vk.com/impeers=362663497&sel=140138665&z=video11948594_456241760%2F283b520181bb3a105e)
[video11948594_456241760%2F283b520181bb3a105e](https://vk.com/impeers=362663497&sel=140138665&z=video11948594_456241760%2F283b520181bb3a105e) (дата обращения:
20.01.2017).

12. Конкурсе звѳзды Марнета-2016. – URL: [http://yoshkarola.bezfor-
mata.ru/listnews/konkurs-zvezdi-marneta-2016/45292938](http://yoshkarola.bezformata.ru/listnews/konkurs-zvezdi-marneta-2016/45292938)

13. Артѳев А. В Сыктывкаре определяют лучших коми блогеров // <http://finugor.ru/news/v-seti-opredelyat-luchshih-komi-bloggerov> (дата обра-
щения: 20.01.2017).



14. Коми блогйысь. Юргö коми кыв. – URL: <http://komimu.com/komi-mu/komi-mu/архив/2016/№-48-18249/1492-коми-блогйысь> (дата обращения: 16.01.2017).

Блогйяс лыддьöг:

- 1) «Алёна Тувсовья». – URL: <https://tuvsovja.blogspot.ru>
 - 2) «Арся». – URL: <https://svaik.blogspot.ru>
 - 3) «Асыывыв Европаті шöйтöм». – URL: <https://asyveuro.blogspot.ru>
 - 4) «Берба». – URL: <https://verba168052.blogspot.ru>
 - 5) «Висер вожса ныв». – URL: <https://erassyhaeva.blogspot.ru>
 - 6) Вячеслав Попов. – URL: <http://blog.bnkomi.ru/blog/343890>
 - 7) «Зырянка». – URL: <http://zyrjanka.livejournal.com>
 - 8) «Изъваса ныв». – URL: <https://tatianakirpichenko.blogspot.ru>
 - 9) «Кебраса». – URL: <https://kebrasa.blogspot.ru>
 - 10) «Кола». – URL: <https://kolananebog.blogspot.ru>
 - 11) «Коми кывбурьяс». – URL: <http://komikyvbur.blogspot.ru>
 - 12) «Куим ю». – URL: <http://kuimju.blogspot.ru>
 - 13) «Ме – Танюк». – URL: <https://tanyuk.blogspot.ru>
 - 14) «Менам видзöдлас: кывбурьяс». – URL: <https://chushka-k.blogspot.ru>
 - 15) «Ната». – URL: <https://skorodumsanata.blogspot.ru>
 - 16) «Тусь». – URL: <http://tusjuk.blogspot.ru>
 - 17) «Шонді». – URL: <https://menamshondi.blogspot.ru>
 - 18) «Эжва йывса». – URL: <https://ezhvajyvs.blogspot.ru>
 - 19) «Юргö коми кыв». – URL: <https://bystrovaele.blogspot.ru>
15. *Сергеева Л.* Указ. соч.



Ю.С. Сычева,

*магистрант Сыктывкарского
государственного университета
имени Питирима Сорокина*

**«АРТ» ЖУРНАЛЛОН
«КОМИ ГИЖЫСЬЯС ЛЫДДЬОНЫ АССЫНЫС КЫВБУРЬЯС»
МЕДВОДДЗА МУЛЬТИМЕДИА ПРИЛОЖЕНИЕ ЙЫЛЫСЬ***

2008-ӧд воӧ «Арт» журнал вӧчис лыддысьысьяслы козин – петіс «Коми гижысьяс лыддӧны ассыныс кывбурьяс» нима медводдза аудио приложение, кӧні юргӧны Владимир Тиминлӧн, Галина Бутыревалӧн, Михаил Елькинлӧн, Евгений Козловлӧн, Нина Обрезковалӧн, Анжелика Елфимовалӧн, Алена Ельцовалӧн кывбурьяс.

Миян мог – тӧдмавны коми студентъяслыс «Арт» журналлӧн мультимедиа приложение вылӧ видзӧдлассӧ.

2009-ӧд воӧ Сыктывкарса канму университетса студентьяс Екатерина Карич, Татьяна Кушманова, Вера Цывунина, Алена Шомысова «Коми литература велӧдӧмын компьютернӧй технологияс» лабораторияын босьтчисны вӧчны «Ӧнія коми поэтьяс» нима мультимедиа проект. Подувнас лоис «Арт» журналлӧн «Коми гижысьяс лыддӧны ассыныс кывбурьяс» CD-приложение. Windows Movie Maker программаын найӧ вӧчисны фильм, кӧні быд поэты сӧма торъя ролик. Микросюжетьяс артмӧны поэтлӧн дженьыд биографиясы, кывбур чукӧрьяслӧн кышӧд серпасьясыс, кывбур юргӧмысь. Рефренӧн лоӧ «Гажа ворсӧмьяс» шылад (30-ӧд трек). Кывбурьяс лыддӧмсӧ колльӧдӧны студентьясӧн чукӧртӧм фотографияяс, кодьяс, студентьяслӧн чайтӧм серти, петкӧдлӧны гижӧдлыс образностьсӧ, лирика геройлыс сӧлӧмкылӧмьяссӧ. Фильм помын вайӧдӧма вӧдитчан литература да проект дасьтысьысьяслӧн фотография. Фильм мунӧ 7 минут да 38 секунда.

Ми корим Екатерина Хазоваӧс (Карич) да Алена Старцеваӧс (Шомысова) казтывны проект дасьтӧм йылыс. Со мый найӧ гижисны:

«Диск серти уджыс миян муніс кыпыда да лӧсьыда, ӧд ми кывзім ӧнія поэтьяслыс медбур кывбурьяссӧ да вермим петкӧдлыны, кутшӧма найӧ воӧны миянлы сӧлӧм вылӧ да мый ми аддзам, найӧс кывзігӧн.



Ми бөрйим сэтшөм вөр-ва серпасьяс, медым стөча петкөдчис кывбурлөн оланруыс, коми мулөн мичлуныс, мортлөн сьөлөмкылөмьясыс. Татчө лөсялісны и Өтувезйыс серпасьяс, и асланым чужан сиктъяслөн фотояс. Найө төдчөдісны юргыс кывбурьяслыс мичсө, сьөлөмкылөмьяслыс джудждасө. Чайтам, мый татшөм удж пыр «Арт» журнал проектлөн коланлуныс да мичыс воссис нөшта на ёнджыка да паськыдджыка».

Екатерина Хазова (Карич)

«Коми поэтыс йылыс презентация вөчөм муніс зэв гажаа, сы вөсна мый котырын вөліны зэв зіль да кыпид руа нывьяс: Катя Карич, Вера Цывунина да Таня Кушманова. Медводз ми корсим кывкудъясыс бөрйөм гижысыс йылыс юөр, сэсся кывзім «Арт» журнал бердын лэдзөм аудиодискъясыс поэтыслыс кывбурьяссө да на дорө нин корсим серпасьяс, снимокъяс. Сэсся Катя да Таня шуисны, презентация пө важтор нин, колө вөчны видеоролик. И асьныс нин ситөртісны уджнымөс *Windows* киностудия отсөгөн. Чайта, видеороликным артмис зэв лөсьыд. Сэні и поэтыслөн портретъяс, и кывбурьясныс юргөны авторъяс гөлөсөн. Видеороликсө видзөдігөн чайтсьө, быттё поэтыс висьтасьө тэныд, и кылан сылыс ловшысө».

Алена Старцева (Шомысова)

Төдчө, мый студентъяслы ёна кажитчөма татшөм удж. Найө төд-масисны коми кывбурьясөн, сьөлөм пыр нуисны найөс да козьналісны миянлы неыджыд, но тырбура дасьтөм да колана коми литература йылыс фильм.

Тайө вөлі 2009 воын, и миянлы лоис интереснө, мый өнія студентъяс мөвпалөны «Арт» журналлөн CD-приложение йылыс. Та могыс Пити-рим Сорокин нима Сыктывкарса канму университетса коми студентъяс пөвстын вөлі нуөдөма анкетируйтөм да гижөм анкетаяссө видлалөма. Анкетаын вөлісны татшөм юалөмьяс:

1. Кывлінныд-ө ті «Арт» журнал йылыс?
2. Лыддянныд-ө ті «Арт» журнал?
3. Мый тіянлы медся ёна кажитчө «Арт» журналын»?
4. Энө кө лыддьөй, көсьянныд -ө төдмасьны журналөн?
5. Төданныд-ө ті, мый «Арт» журналлөн эм мультимедиа приложение (дискъясөн)?
6. Кодсянь та йылыс төдмалінныд?
7. Кывзылінныд-ө дискъясыс, кызди коми гижысыс лыддьөны ас-сьыныс гижөдьяссө?
8. Кывзінныд кө, мый воис сьөлөм выланыд?



9. Кыдзи ті чайтанныд, колёны-ё татшом петасьясыс?

10. Вёзйой, кыдзи позё вёдитчыны татшом озырлунён.

Анкетируйтём вёлі нуёдом 39 студент пёвстын. Вочавидзёмьяс петкёдлісны, мый 3 (7,7%) на пиысь эз кывлыны «Арт» журнал йылысь. 15 (38,5%) студент оз лыддьы «Арт» журнал, 19 (48,7%) лыддьёны шоча, ёти студент (2,6%) лыддис ётпырысь, 4 (10,3%) студент гижисны, мый лыддьёны «Арт» журнал. Коймёд юалём вылё вочавидзис 17 студент. Журналын студентьяслы медся ёна кажитчёны:

- 1) серпасьяс – 10 студент (58,8%);
- 2) художествоа гижёдьяс – 7 студент (41,2%);
- 3) том гижысьяслён кывбурьяс – 3 студент (17,6%);
- 4) сійё, мый гижёма кык кывйён – 3 студент (17,6%);
- 5) оформление – 3 студент (17,6%);
- 6) научнйой статьяс – 3 студент (17,6%);
- 7) гижысьяслён биография – 2 студент (11,8%);
- 8) кышод – 2 студент (11,8%);
- 9) дискьяс – 1 студент (5,9%);
- 10) художествоа гижёдьяс видлалём – 1 студент (5,9%).

(% вайёдом 17 студентьясысь, кодьяс вочавидзисны юалём вылё)

Студентьяс, кодьяс эз лыддьыны «Арт» журнал, кёсйёны сійён тёдмасыны. Сёмын ёти студент гижис: «Торья кёсйём абу».

Мультимедиа приложение йылысь тёдёны 34 студент (87,2%). Дискьяс йылысь найё тёдмалісны:

- 1) школаын – 18 студент (52,9%);
- 2) университетын – 16 студент (47%);
- 3) «Арт» журнал редакцияын – 1 студент (2,9%);
- 4) тайё анкетасысь – 1 студент (2,9%).

30 (88,2%) студент кывзыліс дискьяс, кёні гижысьяс лыддьёны ассыныс гижёдьясё.

8-ёд юалём вылё (Кывзінныд кё, мый воис сёлём выланыд?) студентьяс вочавидзисны:

- «Очень теплые и душевные стихотворения, которые хочется перечитывать постоянно»;
- «Кутшом кужомён, мичаа, интонацияён лыддьё авторыс»;
- «Алена Ельцовалён кывбурьяс; сылён гёлёсыс зэв лёсыд, кывбурсё лыддьё гажаа, кыпыда, колана интонацияён»;
- «Мёным воис сёлём вылё, кыдзи ачыс гижысь лыддьё ассыс кывбурсё, кутшом сёлёмкылёмьяс мёнам вёрзёны»;
- «Зэв лёсыда йитчёма музыкаыс висьталыськёд»;
- «Кыдзи авторьяс лыддьёны ассыныс кывбурьясё»;



- «Нина Обрезковалон кывбурьяс»;
- «Сьолон вылам воисны Владимир Васильевич Тиминлон кывбурьясыс».

100% студент тодчод, мый татшом мультимедиа приложениеас колоны. Студентьяс вэйоны найон вэдитчыны коми литература урокьяс вылын, институтын, конференцияс дырйи, коми праздникяс лосьодигон. Найос позь кывзыны гортын да литературной рытьяс вылын. Сидзжэ студентьяс вэйоны пыртны мультимедиа приложениеяссэ отувезйэ. Некемын студент гижисны, мый кутасны наон вэдитчыны школаын уджалогон.

Анкетируйтом серти позь вочны татшом кывкортдыяс. Унджык респондентяс тодны «Арт» журналлон приложение йылысь «Ония коми поэтьяс» видеоролик серти. Оз став студентяс тодны «Арт» журнал йылысь да сылон CD-приложениеас йылысь, коть и тодмасьом борын шуоны, мый найэ лосьыдысь, коланадысь. Мыйла таdzi артмэ, од и школаын урокьяс вылын, и университетын занятиес дырйи зиль велодысьяс тодмодны велодчысьяс выль петасьясон, ловзэодны велоданторсэ поэтьяслон юргысь голбсьясон. Но, анкета серти, джын юасьом студентяс оз тодны на татшом приложение йывсысь, унджыкыс тодмалом борас оз вэдитчыны наон. Студентьяс вэйоны пыртавны мультимедиа приложениеяссэ отувезйэ. Сэк жэ пасьям, мый отувезйын унатор нин эм, а «Ловья кыв» сайтэ пыртэма «Арт» журналлысь приложениеяссэ, но студентьяс та йылысь эз тодны. Тыдовтчэ, мый велодчысьяс пырджык вичысьоны, мый налы индасны, петкодласны кытысь мый корсьны, ышодасны найос мыйкэ вочны.

Дерт, они дискьясон оз нин ставон вэдитчыны, од борья гаджетьясын дисководьяс абуос нин. Велодчигон «Ловья кыв» сайтэ пырны позянлуныс абу жэ пыр, од уналаын Интернет оз ёна бура уджав, весиг университын. Сайт программаас абу скачивайтан функциясы, и тайэ ёна векнёдэ сийон вэдитчанногсэ. «Ония коми поэтьяс» мультимедиа фильм уна во чэж разодисны школаясэ практика вылэ мунысь студентьяс оти папкаын коми литературысь мукэд уджыаскэд – кафедралон воли традиция сетны школаяслы электронной козин. Коми литератураысь электронной образовательной ресурсьяс дасэтом йылысь петавлис дасысь унджык статья тайэ уджон веськодлысьлон Е.В. Остаповалон [1].

Коми литература паськодэмогысь 2016-од вося ода кора толысь 9-од лунэ «ВКонтакте» социальной везйын лосьодэма «Коми литература радейтысьяс» котыр. Котырын эм «Ловья кыв» сайтлон торья гижэдыас вылэ ыстэ. Талун кезлэ тайэ котырэ пырис 128 морт. Медся видлалана материал: И.А. Куратовлон «Коми кыв» кывбур, А.Е. Ванеевлон «Ловья ва» кывбурьыс юкон, А.В. Ельцовалон «Бара муна, чужан сиктой, муна...»



кывбур, видео, кӧні француз группа «Chet Nuneta» сылӧ «Варыш поз» сыланкыв, Е.В. Козлова йылысь юӧр, студентъясӧн дасътӧм В.В. Пахорукова йылысь «Кыпид сьӧлӧм» фильм, «Арт» журналлӧн медводдза мультимедиа приложение сери лӧсьӧдӧм «Ӧнія коми поэтыяс» фильм.

* * *

1. *Остапова Е.В.* Мультимедийные литературно-образовательные ресурсы на коми языке // Родной язык в современных условиях двуязычия: сб. статей / под общ. ред. Е.В. Остаповой, С.Н. Терентьевой. – Сыктывкар: Кола, 2014. – С.267–274.



J.-P. Hakkarainen,
Project Manager
National Library of Finland

YOU CAN'T HAVE ONE WITHOUT THE OTHER: THE ROLE OF (ELECTRONIC) LIBRARIES IN THE LINGUISTIC ECO-SYSTEM*

The National Library of Finland has executed the Uralic Languages Project in 2012–2016. During the project, more than 30 000 documents in 18 different Uralic languages were made available for both, the research communities and the great public. The end-users can access the digitized content openly at the Fenno-Ugrica repository without any restrictions.

The project was a success and the both user-groups have widely benefited from our work. In my paper, I will 1) discuss the emerging role of libraries in the new era of research partnership, 2) clarify what practical procedures the libraries should take in order to support the linguistic research and 3) launch a step-by-step action plan for libraries for achieving the targets within their new role.

Good practices are often blocked by the library strategies

There is quite a high number of good practices, which can and should be followed by the libraries in order to achieve the abovementioned goals. The major problem seems still to be that the libraries do have their strategies and digitization policies how to spend their many to make content available for the public. However, in many cases these strategies don't necessarily meet the requirements of the research communities, especially if the data of digitized content hasn't been shared openly to the researchers. Libraries should revise their strategies in order to make more data available, because it'll bring wider opportunities to greater audiences.

Librarians should change their mindsets

The first step towards the better practices is the change in mindset. The new skills in librarian's profession are needed in order to reduce the gap between the researchers and information specialists. Quite often, the digitized content is made available in a format that won't help the further use in research. The researchers don't need pictures, but they need data out of pictures.



In my opinion, the librarians should understand, at least to a certain stage, how the data is going to be used in research. The key factor to success is the capability to produce data that is re-usable in research. In order to make the data more available for the scholars' benefit, the librarians should operate as data curators or data stewards, as a data expert who would have enough skills to produce the data out of digitized content. This doesn't mean only that the librarians would produce more data out of the digitized content, but the principles of language documentation and open research data must be understood properly.

Engage the academia to your project

Stop working in your silo. Don't work only for the sake of your organization. The libraries must serve various masters. The services what the libraries offer for the readers are no longer enough for the linguistics research, for instance. The primary key factor for success is the engaging the researchers to your work flow. For instance, the researchers were an essential part of our team, when we were planning the Uralic Languages Project. They took part even before the money was raised. The researchers did participated in planning (choosing the digitized content), decision-making (were part of the steering group), communications (researcher did communicated on our behalf), networking (participation in workshops, seminars and conferences) and they also distributed the data.

By engaging the academia into library's project, we benefitted a lot. We knew exactly what sort of needs they had related to the digitized content. The project members also participated regularly to the scholarly events by arranging and taking part in workshops, seminars and conferences. Through this channel, we managed to spread the audience and more research started to get interested in our work and our work started to get more aligned with the work of the linguistics.

Learn the language documentation and data principles

Libraries are the part of scientific world, but aren't the science itself. The libraries and modern-day librarians are the intermediators between the collections and the researchers. To be fully capable to co-operate with the academia, we needed to learn some basic principles of their work and in language documentation. As it is indicated often, the language documentation is no longer understood as an approach to make descriptive grammars etc. In these days, it does have the aspect of supporting the language communities to improve the mastering of their languages. This aspect must clearly be understood in the libraries, which digitized and published the language material in their repositories.

Besides, the language documentation can easily be adopted into the work flow of libraries as they already possess capabilities to do so. Within the Uralic Languages Project, we understood the language documentation through five components:



- releasing the material openly in repositories, including the relevant metadata;
- mobility of language content;
- producing the added value, for instance by annotation and transcribing the material;
- providing an open access to the archived material;
- sharing the data for third-part systems.

Some of these targets are easily achieved in libraries, but for some of them, the libraries must take over some new procedures.

The released data must be good enough to be used in research. The data of Uralic Language Project wasn't always able to distribute the golden standard data for research communities as such, but we considered our data as a raw data that must be cross-checked by the researcher before using it. However, the libraries can secure that the data does meet some standards when it is released. Good guidelines can be found, for instance in FAIR principles of OpenAire initiative, which get as close as possible to giving quality criteria for research data. The acronym FAIR stands for Findable, Accessible, Interoperable and Reusable (data) and all the libraries must think over again how the FAIR acronym could be met within their organizations.

Action Plan for libraries

The proposed Action Plan must be understood as a loose guideline for libraries, which like to co-operate with the research groups in linguistics:

1. Engage the scholars into your project ask what they want and how they want the material to be accessed invite them to the decision-making make the co-operation as visible by engaging the scholars in communication.
2. Revise the procedures at your home institution improve the work flow in order to meet the needs of the researchers learn new skills and become a data steward spread your new knowledge.
3. Make the data available make your data available for the public. Consider the FAIR principles, when producing the data. Follow the language documentation principles upon your work



N. Partanen,
MA, Aspirant

CHALLENGES IN OCR TODAY: REPORT ON EXPERIENCES FROM INEL*

Introduction

This paper discusses some experiences and workflows from the *INEL project* (<https://inel.corpora.uni-hamburg.de>) (Grammatical Descriptions, Corpora, and Language Technology for Indigenous Northern Eurasian Languages). The long-term project is funded within the framework of the Academies' Programme, which is coordinated by the Union of German Academies of Sciences and Humanities. INEL focuses on languages from two language families: Uralic (Selkup, Kamas, Nenets and Komi) and Altaic (Dolgan, Evenki and Siberian Tatar). It also focuses on Ket, an isolate language**.

Digital preservation has many advantages for preserving written cultural heritage, although it is not a necessity in itself, as paper has also proven to be a long-enduring medium. However, digitalization increases the discoverability and the possibilities for reuse of these materials. The next step is to bring these materials into common and scientific use. Opening these datasets is still in its early phases, and many related practices are underdeveloped. In Scandinavia, public and national libraries have recently begun to offer digitalized materials for public use, although the focus has been mainly on publications before 1910 [1] (<http://www.dlib.org/dlib/july16/paakkonen/07paakkonen.html>). Within language documentation projects, the concept of open data is barely discussed, and the most of the data is behind variously restricted licenses or with limitations to only academic use. This is unfortunate, since incorrectly structured and licensed materials will require considerable future effort and negotiation before they can be taken into active research use.

In order to do enable better usability, content must be accessible. In the case of written documents, this refers mainly to original text data. Optical Character Recognition (henceforth OCR) tools are irreplaceable in efforts to digitalize these resources. There are already broader and more systematic comparisons of different OCR tools available; see, for example, Tafti et. al. [2],

© Partanen N., 2017

* Печатается в авторской редакции.

** Thanks to Rogier Blokland, Jussi-Pekka Hakkarainen and Alexandra Kellner for much appreciated feedback when writing this article.



but I aim to contribute to the discussion by pointing out certain particular needs of language documentation, which may be relevant for the wider linguistic community as well. Within the INEL project [3] we have done an evaluation of currently available tools, and despite our satisfactory results, one must also acknowledge the existence of several alarming bottlenecks, unsuitable technical combinations and dead ends within the existing workflows.

There have been multiple large digitalization projects run by both academic institutions and libraries. Some of these, such as *Endangered Archives* (<http://eap.bl.uk>) run by the British Library, have focused mainly on audio and photographic media. Others, namely *Fenno-Ugrica* (<https://fennougrica.kansalliskirjasto.fi/>) and the *Komi Kyv* (<http://komikyv.ru>) maintained in Syktyvkar, are good examples of the systematic digitalization of written resources. Since many projects have produced vast quantity of scanned pages, the principal question is how we can use this data.

In INEL, we have been working with the digitalization of Selkup, Dolgan and Kamas materials. Our focus has been primarily on published fieldwork materials and other spoken primary data. With smaller languages, it is likely that the recorded and transcribed materials exceed the printed ones in size; generally speaking, the ratio of these two can be used as one way to estimate the usage situation or vitality of the language. The texts we work with in INEL tend to fall into two distinct categories: handwritten manuscripts and published printed transcriptions. This data is also unusually complex because the transcriptions are usually related to various archival audio recordings, but the relation may not be entirely transparent. For example, in the case of some Selkup transcriptions, the recordings available are different versions from those on which the existing transcriptions are based. The same transcriptions may have been published in several publications in somewhat different versions and using various transcription conventions. One recording may be stored in different copies in various archives.

This is also a very particular case for OCR, since the scripts used are often not included in available models, the number of diacritics (possible added by hand) can make the scripts extremely complex, and the fonts used may be highly customized. The OCR challenges could be very comparable to the OCR of printed early modern manuscripts, although there are also clear differences, especially concerning the volume of the texts. The copyright questions also get very complex when it comes to legacy materials, since the rights of the speakers, original researchers, archives and publishers may all interact in many ways, and may be impossible to reconstruct. That said, it seems we are still far away even from solving the copyright issues involved with a very simple set of data, so the question should not be made overly complicated at this point. It is obvious that as free license as possible, preferably Public Domain, would allow the widest reuse.



OCR workflow

Most of the time, the first step after digitalization is to find some solutions for performing OCR on the text. There are specific tasks for which a graphical user interface is particularly suitable. One of these is the post-processing of OCR results. Another is manual correction of the detected text regions. Both of these tasks require a visual comparison between the recognized text and the original image. There are plenty of different tools that offer the possibility of manually correcting OCR'd text, and virtually all of these have very similar user interfaces. The page image is on the left, the text field on the right, and moving the cursor around the text field has some kind of correspondence on top of the image. This raises the question of why so many projects have ended up deciding to develop new tools for post-correction. *Fenno-Ugrica* (<https://github.com/NatLibFi/ocrui-frontend>) has their own tool, as does *Deutsches Textarchiv* (<http://www.deutsches-textarchiv.de/doku/software>). *PoCoTo* (<http://thorstenv.github.io/PoCoTo>) has been developed in CIS Munich within a larger EU project, and, in contrast to the others, it seems to have features related to batch correction of systematic mistakes.

Compared to this, the initial training of an OCR model is very well suited to a command line environment, especially with respect to the transparency and reproducibility of the training process.

Current solutions

The most critical commonly lacking features are:

- Lack of find and replace with regular expressions [4]
- Lack of good XML export

The first issue is connected to the automatization of the proofreading process, since some mistakes are always systematic. The export format is more serious issue. The following attributes would ideally available alongside the output format:

- Recognized word (or character)
- Coordinates on page
- Recognition confidence
- Some style information
- Information about the OCR system used
 - Version
 - System settings
 - OCR model used (with version)
 - Information about page processing, if that was performed

As a format, some kind of XML seems to be the current standard. However, certain formats often employed, such as plain text files, Word documents or PDFs are clearly not satisfactory. The National Library of Finland has also developed their own export format which contains ALTO XML, metadata and the text in



plain text, and which was created to suit the needs of users who benefit from access to local copies [1] (<http://www.dlib.org/dlib/july16/paakkonen/07paakkonen.html>). It can be expected that new solutions for distributing this kind of data will emerge, since users will certainly have the need to download the data locally and to update it with the newest resources once they are added.

At the moment, the OCR market is dominated by *ABBYY FineReader* (<https://www.abbyy.com/finereader>) and *Tesseract* (<https://github.com/tesseract-ocr/tesseract>) [5]. The former is entirely commercial software, which makes it very problematic from the perspectives of open science and software. Tesseract, and its derivations such as *OCRicola*, are open source, but they offer only the core functionality of OCR recognition and thereby must be used in conjunction with other tools, and, at least in INEL, we have not yet found a satisfactory combination.

The main problem with ABBYY-based solutions is that the two versions of this program, Desktop and Engine, differ very greatly from one another and it is not possible to combine their functionalities. One is used through a graphical user interface, and the Engine version is meant to be used in more automatized manner from the command line. The main functionality that Abbyy Desktop lacks is XML export. This means there is no export format that would directly store the page position information. However, in the Desktop version, it is possible to train new language models. The problem is that the models trained in Abbyy Desktop cannot be used in Abbyy Engine. With rare scripts and endangered languages, the current language models are insufficient and some additional training is necessary in order to achieve a good output. It would already be a half-way solution to be able to use the models trained in Abbyy Desktop also in Abbyy Engine.

Why Abbyy Desktop does not have an XML export function is certainly a valid question. For a long time, I suspected this was due to corporate business interests: there are some reasons it is more lucrative not to offer XML export for a casual user. Perhaps it would allow them to do too much, instead of relying on the company's more expensive options. However, after using Abbyy for a longer while, I have begun to suspect that there is also a softer reason. Could it be that even Abbyy does not know how to store user-edited position information? When a user edits a text in Abbyy, there are moments where a word entirely loses its highlighting, as if the software would not have information about its position any longer. There are similar problems with the Revizor editor developed by National Library of Finland in the Fenno-Ugrica project. Especially at the line boundaries the program seems to lose its information about the correct line. This all signals that it must be extremely difficult to store word position information in user-edited text. The assumption is very logical in the sense that the user can



easily copy and paste text around, delete some words and retype them, and somehow the software would still need to keep track of positions. If the task is too difficult for even the industry leader, we may in deeper trouble than we thought, as it is unlikely that smaller open source projects could reinvent this function. I remain optimistic that in the longer run, the open source community will emerge as the winner. One demand for this is certainly to stop building a new software for every project and cooperate more between different involved parties.

In INEL, we are currently working through Abbyy Desktop plain text file and HTML exports, which both keep the paragraph structure and page numbering. It has been possible to distinguish page numbers as sequentially growing numbers that are the only items on the line, and chapter numbers as those that are on the same line as the title or the first line of the text; in any case, not alone. However, these definitions are not very generic and also require manual adjustment. The data retrieved this way is aligned with Russian sentences using hunalign (<https://github.com/danielvarga/hunalign>), again with some manual supervision, and converted into a Toolbox file with a simple script that transforms the file structure to fit Toolbox demands. This is eventually imported into FLE_x, with line, chapter and page information being stored on the note tier. The solution selected in INEL works for the careful digitalization of individual texts one by one, but would not be satisfactory for use on large scale digitalization projects.

OCR model training

OCR model training can be carried out at different levels. Abbyy has been used by us on the character and character combination level. Since the wordform information is not used at all, there is no interference from the Russian word frequency and character combination information that the software could use with Russian text. New combinations can also be trained, which can be especially helpful when teaching Abbyy to recognize rare and new characters that are not used in Russian orthography. The next step would be to feed the manually corrected word forms from the Dolgan corpus to the Dolgan Abbyy model. One benefit of this relatively shallow model is that it does not get mixed up by dialectal variation in the word forms. If the model has exact information about all possible word forms, the recognition quality can possibly be improved even further. However, according to Silfverberg and Rueter, word lists may still be superior to morphological analyzers [6]. It must also be taken into account that a more sophisticated model may also struggle more when the text it encounters is less standard.

OCRicola is a software built on the basis of Tesseract and has most of the same functionalities. The main difference between the tools is that OCRicola is able to use a finite state transducer (HFST) to generate the word forms used in



recognition. In theory the morphological analyzer is also able to generate forms that are never present in the corpus but may occur in a new text.

Challenges with the Tesseract/OCRicola model are related especially to fonts and availability of a large enough training corpus. With OCRicola the question arises also about the existence of an available Finite State Transducer, but at least within the *Giellatekno* (<http://giellatekno.uit.no>) infrastructure a large number of Uralic languages are already covered.

Contrary to the Tesseract/OCRicola model, Abbyy takes care of most of the above-mentioned tasks by itself without user interference. This also makes the model development non-transparent. It is possible to use another Cyrillic model as a backbone and build upon that by adding the special characters needed for other languages. This makes Abbyy model training very broad and general, and allows even arbitrary matching between characters and recognized forms, which is a great benefit when compared to Tesseract/OCRicola, which demands that the trained character also be present in the font.

Indeed, recognition of languages such as English, Russian and German is almost flawless. In the same vein, it must be stated that even with new languages Abbyy often reaches very good results with a relatively small investment in the training. The only concrete problems with Abbyy are the lack of transparency behind the model training, lack of interoperability between different versions and some issues with missing export formats. As already mentioned, the lack of XML export format results in loss of information with the word and line positions on the page.

I have often encountered the idea that page position information is not necessary. There are several situations where this information is very critical. For example, different page elements such as page numbers are very easy to recognize from the layout when their position is known. The same goes for paragraph information, since line indent coordinates, when known, make paragraph detection rather reliable. Detecting it from unstructured lines is, on the other hand, rather hopeless.

It can be argued that there are cases where this information can be disregarded. When working with texts like the Bible or Koran, the original verse structure is enough to distinguish different elements. Some texts follow essentially the same model where each utterance can be identified as a part of a closed set. Some data in INEL is also like this, such as the translated sentence lists collected from many different languages in Siberia. The rarer and more susceptible to subjective reading and interpretation the text is, more valuable it is to be able to reproduce the original text or to easily match the recognized words to positions on the page. Especially with old typefaces or handwritten manuscripts, the matter of interpretation tends to become more significant.



There are other good discussions about the steps of the OCR process; for example, Holley [7] goes through it step by step. Although that description is some years old, the principle has not changed, and table 1 in her paper provides a good overview of all necessary phases (<http://www.dlib.org/dlib/march09/holley/03holley.html>). I personally still find it questionable how well the current tools actually store information about the entire workflow. The measure of this would be whether we are able to associate an arbitrary token in the final recognized text with the matching coordinates in the original manuscript, microfilm or whatever the source has been. At least in regard to INEL workflows, the answer is unfortunately no.

There are experiences that a text resulting from OCR can already be corrected to some degree using simple ngram-based post-processing [8]. This is plausible, especially with data that is relatively well recognized, although the worst-recognized parts should probably be approached by redoing the OCR [9]. I would still be cautious about eliminating the need for manual post-correction, which, although impractical for millions of titles, can still be relatively cost-effective for hundreds or thousands of proofread books in a year.

Within INEL, the OCR is not a direct goal in itself, but the text extracted from the documents will be associated with the original recordings. There are also cases where the original recording has been lost or was never made. In the latter case, the document on which the OCR has been carried out is the closest primary data source which we can have. Thereby there must be a close connection between the original text document and the later XML file. The XML format currently used in INEL is the one produced by the *EXMARaLDA* (*Extensible Markup Language for Discourse Annotation*) System. The EXMARaLDA tools have a high level of interoperability with existing standards used in language documentation. At the *Hamburger Zentrum für Sprachkorpora* (<https://corpora.uni-hamburg.de>) we already have a long practice of publishing online different visualizations for spoken data (see *Demo corpus* on the HZSK website (<https://corpora.uni-hamburg.de/drupal/de/islandora/object/spoken-corpus:demo>)).

Publishing the texts is naturally much simpler when there are no annotations. One method of publishing is simply rendering the text as plain HTML. For a normally structured book, appropriate dose of CSS already makes the reading experience in this format very enjoyable. On the other hand, if it is possible to render the text into HTML, it is also possible to produce other formats, such as EPUB files, which are very good for reading on mobile devices.

There are also some types of documents that would benefit from a more complex rendering online. This is particularly the case with such manuscripts where even the reading of different characters may be debatable. Although I do not discuss the text recognition of handwritten documents here in depth, it should



be pointed out that this is the type of document where some new facsimilé type of online publishing options would be the most needed. What I would most like to see is a set of tools that allow very easy and shallow rendering of XML and medium-quality images, instead of building the solution on top of heavy servers and byzantine architectures that require complex customization to fit the needs of specific projects. As I have discussed here, the needs of different projects may in reality be more similar than we think. The XML files could be even stored in public repositories such as GitHub, which would bring the work closer towards actually open and collaborative science.

In an ideal world, OCR model development could be done with open source tools for languages that have the following resources:

- A large corpus available with a permissive license
- A good collection of fonts that closely match the ones used in publications
- Frequency lists from possibly even a larger corpus
- Consistent spelling that does not vary across publications

It is noteworthy that for large languages such as English or Finnish, all these resources and conditions exist. However, for many of the smaller languages, these pose a large issue. Yet I'm not so sure if the distinction between bigger and smaller languages is always so clear. As the Komi Zyrian corpus is also now *over 30 million words* (<http://komicorpora.ru>), I see no problems in adapting almost any approach which works with larger languages to Komi as well.

Challenges for language technology

OCR'd texts pose many tasks for natural language processing. First of all, there is the need to transform idiosynchronic writing systems into the same, ideally phoneme-level representation. The text often needs some kind of a morphological analysis in order to be more usable for linguistic analysis. Applying different tools to this end is highly desirable. For example, with digitalized Komi-Zyrian and Saami materials, the GiellaTekno tools have already proven very useful, and maybe something similar could be applied later also to INEL corpora. It can also be mentioned that many texts are attractive targets for named entity recognition (NER) since they often repeatedly cover the same prominent historical events in different languages and repeat the same historical figures.

One particularly relevant tiny task I can foresee is the automatic matching of recognized text with the original pages. This can be done by automatically comparing the proofread text masses with the XML output using word coordinate information. The two texts are naturally different, as one variant is not proofread and contains many errors that were never present in the other version. However, I believe this is a minor obstacle, and detecting the most similar pages (especially knowing that the following page normally continues the text) should be very doable in the immediate future. I have not found any existing solutions to this,



which again may reflect the immaturity in the use of OCR'd research data in fields where high quality text output is necessary or a high priority.

Contribution to language documentation

In corpus linguistics and language technology, the texts are usually treated as the target of investigation in themselves. When applied to the endangered languages, there is often a need to personalize them further and add into metadata information about the writers and translators, among other relevant details. Instead of variables such as publishing time and place, there is the obvious need to know writer's date of birth and place of birth. These already tell quite a bit about the possible native dialect of the writer, which again can be one way to explain idiosyncronic features of individual texts. Often this kind of information is stored already in public repositories such as Wikipedia, so one additional question is how to employ these connections the most effective way.

In many endangered language communities, the pool of active and prominent speakers is small. It is highly likely that the same individuals have been writing to local newspapers and even published longer pieces of prose, and have also been speaking in public events that have been recorded. Alternatively, they may have ended up being recorded by different linguists or ethnographers, resulting in archive items that can be relatively easily associated with each other. Similarly, if the writers are deceased, it can be expected that members of the community will recognize and remember them.

Lastly, it could also be speculated that with many languages, the existing language documentation resources are the largest available bodies of texts in that language. In these cases, one could also imagine a scenario in which the language documentation data could be directly used to create an OCR model. This is somewhat counterintuitive, but text is text, and if the match between phoneme representation in written and spoken varieties is close enough, which it often is with languages with new orthographies, the chances of this succeeding are high.

Parallel texts also open up additional research possibilities. In the Fennougrica project in Helsinki, one of the main ideas has clearly been to collect resources that exist in several languages spoken in Russia. The result is that there are now publications which are translated to many Uralic languages. When the books are compared, the texts are immediately useful, since they share exactly the same structure, often on the sentence level with only small deviations.

Some resources are parallel in less obvious ways. There are historical events that have been covered in every newspaper in the Soviet Union and beyond. These are not parallel sentences as such, but still the texts are thematically linked in very intriguing ways. As mentioned, applying tools such as named entity recognition to this kind of data could be very interesting.

Even more abstract parallels can be drawn between spoken and written resources, since there are many narratives that have been recorded in different



retellings and published multiple times. A good example of this is the Komi folktale Zarnia Bözha Kañ, which is also connected to the Russian folktale Maša i Medved. It is published in two different variants which are both stored in Komi Nebögain collection in Syktyvkar, and has also been recorded by Erkki Itkonen, as told by Vasiliy Lytkin in 1957 (Kotus recording id: 1323_2az). It is still unknown how exactly this data can be used, but as far as I see, there are many open roads that can be explored.

Closing words

It is easy to talk about the differences in the data different projects deal with and produce, but there is a lot that is essentially the same in all OCR workflows. These similarities mean that most of the practices and solutions that work in one place will also work elsewhere.

There is some kind of a boundary between modern printed resources and early modern manuscripts and prints, and many tools are customized to work with one or another. I also outlined above that the needs for presentation are clearly different for the handwritten manuscripts and printed products. This said, the boundary is not that clearly cut. Researchers today continue to produce handwritten field notes, the handling of which does not essentially differ from what we have to do for the notes from Castrén's time. This is also obvious in the work we are doing in INEL. Kuzmina's and Donner's fieldnotes have more in common than not, although they are separated by more than a century. Similarly, early phonogram recordings can be dealt with conceptually in a rather similar way to modern multimedia recordings.

* * *

1. *Pääkkönen Tuula, Jukka Kervinen, Asko Nivala, Kimmo Kettunen, Eetu Mäkelä*. Exporting finnish digitized historical newspaper contents for offline use // D-Lib Magazine. – 2016. – № 22.

2. *Tafti, Ahmad P., Ahmadreza Baghaie, Mehdi assefi, Hamid R. Arabnia, Zeyun Yu, Peggy*. Peissig OCR as a service: An experimental evaluation of Google docs ocr, Tesseract, ABBYY Finereader, and Transym // In advances in visual computing: 12th international symposium (12–14 Deember, 2016, Las-Vegas). – 2016. – P.735–746.

3. *Wagner-Nagy, Beata, Hanna Hedeland, Timm Lehmberg, M. Rießler*. INEL: eine Infastruktur zur Dokumentation indigener nordeurasischer Sprachen // In Konferenz «Forschungsdaten in den Geisteswissenschaften (Forge 2015)» (18 September, 2015, Hamburg). – 2015.

4. *Mancinelli, Tizana*. Early printed edition and ocr techniques: What is the state-of-art? Strategies to be developed from the working-progress mambrino project work // Historias findigas. – 2016. – № 4. – P.255–260.



5. *Smith, Ray*. An overview of the Tesseract OCR engine // In proceedings of the 9th international conference on document analysis and recognition (23–26 September, 2007). – 2007. – P.629–633.

6. *Silfverberg, Miikka, Jack Rueter*. Can morphological analyzers improve the quality of optical character recognition? // In proceedings of the 1st international workshop in computational linguistics for Uralic languages (16 January, 2015, Tromsø). – Tromsø, 2015.

7. *Holley, Rose*. How good can it get? Analysing and improving OCR accuracy in large scale historic newspaper digitization programs // D-Lib Magazine. – 2009. – № 15.

8. *Arkhangelskiy T., Medvedeva M*. Developing morphologically annotated corpora for minority languages of Russia // in proceedings of corpus linguistics fest (6–10 June, 2016). – Bloomington, 2016.

9. *Kettunen Kimmo, Pääkkönen Tuula*. Measuring lexical quality of a historical Finnish newspaper collection – analysis of garbled OCR data with basic language technology tools and means // In proceeding of the 10th international conference on language resources and evaluation (23–28 May, 2016, Portož). – Portož, 2016.



T.A. Pirinen,

*Hamburger Zentrum für Sprachkorpora,
Universität Hamburg*

F.M. Tyers,

*Giela ja kultuurra instituhtta,
UiT Norgga arktalas universitehta,
Arvutiteaduse instituut, Tartu Ülikool*

T. Trosterud,

*Giela ja kultuurra instituhtta,
UiT Norgga arktalas universitehta*

V. Vincze,

MTA-SZTE, Szegedi tudományegyetem

M. Rießler,

*Freiburg Institute for Advanced Studies,
Albert-Ludwigs-Universität Freiburg*

ACL SIGUR – ASSOCIATION OF COMPUTATIONAL LINGUISTICS' SPECIAL INTEREST GROUP ON URALIC LANGUAGES

The Association of Computational Linguistics [1] (ACL) is an organisation organising major conferences in the field of computational linguistics. The special interest groups are sub-groups of ACL members and others that are interested in a specific sub-topic of computational linguistics. Typical activities of special interest groups include workshops, conferences and publications as well as providing best common practices and guidance, and a communication platform for interested scholars and language enthusiasts.

What we present in this article is an interest group of computational linguistics of Uralic languages – the ACL SIGUR [2] / The ACL SIGUR was initially founded at the First International Workshop for Computational Linguistics of Uralic Languages [3], where we in the discussions with our colleagues recognised that there was an urgent need to get researchers of Uralic computational linguistics organised and to better share the informations about resources and practices.

This article is organised as follows: in Section 1 we present how the SIG is organised and how to participate in SIG activities. In Section 2 we describe



our current regular workshop series and potential future meetings. In Section 3 we describe our catalogue of computational linguistics resources for Uralic languages. In Section 4 we describe the standard guidelines we are promoting for researchers of Uralic languages in order to get larger interoperability and participation in research community of computational linguistics at large as well. Finally in Section 5 we describe what we are hoping to achieve more in the future years of the SIG.

Organisation

The special interest group is free and open for anyone to join, the only thing required is filling in a Google form at our website [4]. The group has no membership fees and the participation does not have any requirements. In the sign-up form you have the option to subscribe to our low-volume mailing list; the mailing list is mainly used to announce the upcoming conferences and meetings and the average mail load is less than one message per month at the moment. The group is formed under common ACL SIG practices and the rules related to that can be found in the constitution [5]. What this means in practice is that we have a board and our regular yearly meetings, where we decide on future events and activities organised by the SIG.

Meetings and conferences

The group was founded when we began organising a workshop series in computational linguistics of Uralic languages, building on that, our annual meeting schedule so far has been based on organising a yearly workshop. The first workshop was organised in Tromsø, Norway in January 2015, followed by Szeged, Hungary in 2016 [6] and St. Petersburg, Russian Federation in 2017 [7].

The negotiations for the 2018 conference are still underway, but our aim is to organise it in a prominent University within a Uralic-speaking area with strong background in computational linguistics. Each of the workshop has had an open call for papers and the successful papers have been published in reputable free open-access repositories [8]. In 2015 and 2016 we invited the best papers to be extended for a full-length journal publication in Northern European Journal of Language Technology (NEJLT) journal and Acta Linguistica Hungarica (ALH) respectively.

Resources

One of the leading ideas of founding the SIG was that we need to be better aware of the resources that are developed in the various research sites and communities of Uralic computational linguistics, in particular we identified a problem of overlapping and duplicated work efforts, which are not greatly desirable in such scarcely resourced field as Uralic studies in computational linguistics are. In the European context, larger projects like CLARIN and META-NET have made extensive studies of the state of the national computational



linguistics, e.g. the META-NET white paper series [9], sadly however, this does not include much detail in Uralic languages apart from the three national Uralic languages of EU: Finnish, Hungarian and Estonian. While we have taken inspiration of these publications when designing our resource catalogue, our aim is much more modest and down-to-earth usable resource collection for researchers of Uralic languages, we practically just intend to maintain a resource listing for the practical use of finding resources to use in research: when performing tests, comparisons etc. The current view on the catalogue is available on our web site in resource matrix form [10].

Current best practices

One of the goals of the ACL SIGUR is to work towards consensus on some of the best common practices in working with computational linguistics' tasks for Uralic languages. What this means in practice is that we want to provide instructions on best experiences we had with building the computational linguistics infrastructure, e.g. creating new standardised keyboard layouts for minority languages, building spell checkers and correctors, designing morphological analysers including tag sets etc.

In terms of morpho-syntactic analysis, we are participating in the *Universal Dependencies* project*. Universal dependencies project is a standard for morphological analysis and dependency syntax for all of the languages, but it has a language specific component for extensions. One of our aims in writing

* Joakim Nivre, Zeljko Agic, Lars Ahrenberg, Maria Jesus Aranzabe, Masayuki Asa-hara, Aitziber Atutxa, Miguel Ballesteros, John Bauer, Kepa Bengoetxea, Yevgeni Berzak, Riyaz Ahmad Bhat, Eckhard Bick, Carl Borstell, Cristina Bosco, Gosse Bouma, Sam Bowman, Gülşen Cebiroglu Eryigit, Giuseppe G. A. Celano, Fabrizio Chalub, ^agri ^oltekin, Miriam Connor, Elizabeth Davidson, Marie-Catherine de Marneffe, Arantza Diaz de Ilaraza, Kaja Dobrovoljc, Timothy Dozat, Kira Droganova, Puneet Dwivedi, Marhaba Eli, Tomaz Erjavec, Richárd Farkas, Jennifer Foster, Claudia Freitas, Katarina Gajdosová, Daniel Galbraith, Marcos Garcia, Moa Gardenfors, Sebastian Garza, Filip Ginter, Iakes Goenaga, Koldo Gojenola, Mem-duh Gokirmak, Yoav Goldberg, Xavier Gómez Guinovart, Berta González Saavedra, Matias Grioni, Normunds Grázltis, Bruno Guillaume, Jan Hajic, Linh Ha MD, Dag Haug, Barbora Hladká, Radu Ion, Elena Irimia, Anders Johannsen, Fredrik Jørgensen, Hüner Kaşıkara, Hiroshi Kanayama, Jenna Kanerva, Boris Katz, Jessica Kenney, Natalia Kotsyba, Simon Krek, Veronika Laippala, Lucia Lam, PhDng Lê HDng, Alessandro Lenci, Nikola Ljubesic, Olga Lyashevskaya, Teresa Lynn, Aibek Makazhanov, Christopher Manning, Catalina Mărănduc, David Marecek, Héctor Martínez Alonso, André Martins, Jan Masek, Yuji Matsumoto, Ryan McDonald, Anna Missilä, Verginica Mititelu, Yusuke Miyao, Simonetta Montemagni, Keiko Sophie Mori, Shunsuke Mori, Bohdan Moskalevskyi, Kadri Muischnek, Nina Mustafina, Kaili Müürisep, LDDng NguyDn ThD, HuyDn NguyDn ThD Minh, Vitaly Nikolaev, Hanna Nurm, Petya Osenova, Robert Östling, Lilja Övrelid, Valeria Paiva, Elena Pascual, Marco Passarotti, Cene-



the Uralic guidelines [11] is to minimise the differences between the same language-specific extensions as well as ensure that the same Uralic features use the same encoding throughout. This is crucial for example for things like participial morpho-syntax that is rich in Uralic languages but traditions of encoding it in linguistic descriptions vary within Uralic area and all of Uralic descriptions are rather different from what is common in the non-Uralic languages.

Future goals

While the majority of Uralic languages (although not the majority of speakers) are spoken within the borders of the Russian Federation, membership of the special interest group has so far been restricted to a couple of sites in Moscow and Syktывkar. It is part of our future goals and ongoing efforts to increase membership throughout the Uralic-speaking areas of the Russian Federation.

* * *

1. URL: <http://aclweb.org>
2. URL: <https://acl-sigur.github.io>
3. URL: <http://gtweb.uit.no/iwclul2015>
4. URL: <https://docs.google.com/forms/d/19jasNykCI5449BFNtm5jJ94CFIfQYXNEt4uaFuOkADw/viewfo:rm>
5. URL: <https://acl-sigur.github.io/sigur-constitution.html>
6. URL: <http://rgai.inf.u-szeged.hu/iwclul2016>
7. URL: <https://acl-sigur.github.io/iwclul2017.html>
8. *Tommi A. Pirinen, Francis M. Tyers, and Trond Trosterud*. Proceedings of the First International Workshop for Computational Linguistics of Uralic Languages. – Tromsø, 2015; *Tommi A. Pirinen, Francis M. Tyers, and Trond Trosterud*. Proceedings of the Second International Workshop for Computational Linguistics of Uralic Languages. – Szeged, 2016; *Tommi A. Pirinen, Francis M.*

Augusto Perez, Slav Petrov, Jussi Piitulainen, Barbara Plank, Martin Popel, Lauma Pretkalinina, Prokopis Proko- pidis, Tiina Puolakainen, Sampo Pyysalo, Alexandre Rademaker, Loganathan Ra- masamy, Livy Real, Laura Rituma, Rudolf Rosa, Shadi Saleh, Baiba Saulite, Sebastian Schuster, Wolfgang Seeker, Mojgan Seraji, Lena Shakurova, Mo Shen, Natalia Silveira, Maria Simi, Radu Simionescu, Katalin Simkó, Mária Simková, Kiril Simov, Aaron Smith, Carolyn Spadine, Alane Suhr, Umut Sulubacak, Zsolt Szántó, Takaaki Tanaka, Reut Tsarfaty, Francis Tyers, Sumire Uematsu, Larraitz Uria, Gertjan van Noord, Viktor Varga, Veronika Vincze, Lars Wallin, Jing Xian Wang, Jonathan North Washington, Mats Wirén, Zdenek Zabokrtsky, Amir Zeldes, Daniel Zeman, and Hanzhi Zhu. Universal dependencies 1.4, 2016. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics, Charles University in Prague.



Tyers, and Trond Trosterud. Proceedings of the Third International Workshop for Computational Linguistics of Uralic Languages. – St. Petersburg, 2017 [to appear].

9. *George Rehm Hans Uszkoreit*. Meta-net white paper series: Europe's language in the digital age. – URL: www.meta-net.eu/whitepapers

10. URL: www.acl-sigur.github.io/matrix.html

11. URL: <http://universaldependencies.org/uralic>



M. Rießler,
*University of Freiburg,
Fellow at the Freiburg
Institute of Advanced Studies*

LANGUAGE DOCUMENTATION MEETS LANGUAGE TECHNOLOGY*

Credits

This is the written version of my speech at the international scientific conference on «Electronic Writing of RF Peoples». It builds on my ongoing projects, which are carried out in very close collaboration with several colleagues¹ and the preliminary results of which are published (or are prepared to be published) in [1].

Last but not least, important ideas were also developed while writing the application for the INEL project² by Beáta Wagner-Nagy together with The Hamburg Center for Language Corpora.

Introduction

Language documentation (aka documentary linguistics) is an emerging sub-field of applied linguistics. Research in language documentation aims at the provision of long lasting, comprehensive, multi-faceted and multi-purpose records of linguistic practices characteristic of a given speech community, often in conditions where these languages are under threat of disappearing (cf. [2]). Although it evolved out of traditional fieldwork methodology used primarily by descriptive linguists and language anthropologists, language documentation is no longer merely a method, as it has its own primary aims and methodologies. One of the most important purposes of language documentation is making data available for further research on and for endangered languages, for both further theoretical and applied research, as well as for direct use by the relevant speaker communities.

© Rießler M., 2017

* Печатается в авторской редакции.

¹ Those are especially Rogier Blokland (Uppsala), Ciprian Gerstenberger (Tromsø), Marina Fedina (Syktyvkar), Niko Partanen (Hamburg), and Joshua Wilbur (Freiburg).

² Grammars, Corpora, and Language Technology for Indigenous Northern European Languages (INEL). – URL: <http://inel.corpora.uni-hamburg.de>; first applicant and project leader: Beáta Wagner-Nagy, second applicant: Michael Rießler, planned duration 2016–2033, funded in the Academies' Programme by the Union of the German Academies of Sciences and Humanities.



Ideally, the data pool provided by the language documenter includes a comprehensive, deeply annotated and easily accessible corpus of primary language recordings, representing a wide variety of texts in terms of chronology (e.g. age of recorded speakers), geography (e.g. dialects), and other sociolinguistic variables (e.g. gender and educational background of speakers, registers, genres, etc.). In addition to annotations, cataloging metadata are crucial for the intellectual accessibility of the documented data and concern both the *content* of the recorded speech sample (typically represented as phonological transcriptions, morphosyntactic tagging, and translations) as well as the *context* (such as actors, places, speech events, but even meta-documentation about the project itself; on «meta-documentation», cf. in particular [3]).

Along with methodologies and best practices related to fieldwork and archiving (including questions of research ethics, protection of copyrights, resource discoverability, data standards and long-term data safety), the usefulness of the actual product of language documentation for linguistic research hinges on the quality and quantity of *annotations* as the basis for further empirical analyses and data derivations.

The use of language documentations for corpus-based investigations on endangered and less-known languages and the role of computational linguistics for the field has been a driving topic in recent years. In principle, with respect to the data types involved, endangered language documentation generally seems similar to corpus linguistics (i.e. «corpus building») of non-endangered languages. Both provide primary data for secondary (synchronic or diachronic) data derivations and analyses (cf. [4], for data types in language documentation, and [5], for a comparison between corpus linguistics and language documentation). The main difference is that traditional corpus (and computational) linguistics deals predominantly with larger non-endangered languages, for which huge amounts of mainly written corpus data are available. The documentation of endangered languages, on the other hand, typically results in rather small corpora of spoken genres.

Identifying Methodological Gaps and Problems

Regarding collaborative tools and user interfaces for transcribing, archiving and browsing multimedia recordings, language documentation has made huge technological progress. However, the development of digital tools is only one possible application of *computational methods* in language documentation. Paradoxically, the field has only rarely considered applying such methods to annotate data automatically and hence more efficiently – both with respect to quality and to quantity – in order to create a solid foundation for new and better corpus-based linguistic research on smaller languages.



Note that such computational methods for automated corpus annotation are well-known from corpus linguistic research on majority written languages as well as spoken varieties (e.g. dialects and sociolects) of these languages. Even smaller written languages are becoming more and more the focus of computational linguistic research, and relevant language-technological methods and tools are completely functional even for endangered written languages such as Northern Saami (cf. [6]).

Compared to such true computational approaches to corpus building, researchers working in the framework of language documentation, i.e. fieldwork-based documentation, preservation, and description of endangered languages, create their annotations either manually or semi-manually with the help of software like Field Linguist's Toolbox³, FieldWorks Language Explorer⁴, or similar tools (cf. also [7] for a description of a similar approach during an early stage in my own Kola Saami Documentation Project). Common morphosyntactic annotations created with the help of these tools include glossed texts with morpheme-by-morpheme interlinearization. Whereas these annotations are qualitatively rich, including the time alignment of annotation layers to the original audio/video recordings, the resulting corpora are relatively small and rarely reach 100,000 word tokens. The main reason for the limited size of such annotated language documentation corpora is that (semi-)manual glossing is an extremely time consuming task.

The mismatch between the application of computational and corpus linguistic methodology in endangered language projects vs. projects working on majority languages is also reflected in the use of «corpus-based» as a fashionable label for a new, innovative approach in descriptive linguistics of small and under-described languages. In its actual sense, corpus-based analysis and description can only be understood as a truly quantitative approach. A corpus can be used to find specific examples, but to base one's analysis and description on the whole corpus as such and to ground analyses on statistically significant generalizations are not only very different tasks, but also require a different focus in order to achieve.

Another main problem I identify especially in the documentation of small Northern Eurasian languages is that projects sometimes ignore the existence of orthographies and prefer phonemic or even detailed phonetic transcription. Note that most of the languages of Northern Eurasia (or at least their main variants) have established written standards as the result of institutionalized and/or community-driven language planning and revitalization efforts. For some of

³ Toolbox, for short, <http://www-01.sil.org/computing/toolbox>

⁴ FLEx, for short, <http://fieldworks.sil.org/flex>



these languages, e.g. Komi-Permiak, Komi-Zyrian, Northern Khanty, Northern Selkup, and Tundra Nenets, a significant amount of printed texts can be found in books and newspapers, and several of these languages are also used digitally on the Internet today. Many contemporary and historical printed texts written in these languages are also already digitized⁵, which would make it easily possible in principle to combine spoken and written data in one and the same corpus. Compared to common practice in corpus building for non-endangered languages or dialects (cf., e.g., the discussion by [8] for Russian), the use of phonemic transcriptions in the mentioned language documentation projects seems highly anachronistic.

The last problems I want to address concern archiving practices and workflows dealing with the multimodal character of language documentation data. Archiving practices have been a driving issue in the field of endangered language documentation. As one result, new archiving institutions, such as The Language Archive (Nijmegen)⁶ and the Endangered Languages Archive (London)⁷ as well as a set of standards have been established. However, there are several gaps in integrating these standards into the existing infrastructures in a way that suits the needs of projects relying on truly computational methods in corpus annotation (like the projects I work with). Typically, fieldwork-based language documentation projects consist of teams working with complex multimedia recordings linked to linguistic annotations and other metadata. Since a language documentation – or a linguistic corpus for that matter – is never ultimately finished, the relevant workflows for processing and publishing data need to build on collaborative, versioned working platforms to ensure multi-cited teamwork and scientific referenceability. Though present as an ideal, this barely corresponds to the reality, which is mainly because the current digital archiving platforms are too static to handle continuously refined corpus data.

Some language documentation projects ignore this problem or postpone archiving for an indefinite period into the future⁸. Often, projects rely on local

⁵ For printed sources from the Soviet Union and earlier, the Fenno-Ugrica Collection is especially relevant, <http://fennougrica.kansalliskirjasto.fi>; for both Komi languages, see <http://komikyv.ru>

⁶ TLA, for short, <https://tla.mpi.nl>

⁷ ELAR, for short, <https://www.soas.ac.uk/elar>

⁸ My own work with the Kola Saami Documentation Project provides an example for such a case, which becomes obviously clear when one looks at the unfinished state of its archive [Rießler, M., in collaboration with Afanasyeva, A., Behnke, A., Danilova, S., Dubovtsev, A., Ershtadt, A., Jackermeier, D., Karvovskaya, E., Kotcheva, K., Kusmenko, J., Litvak, M., Nikolaev, S., Olyzko, K., Partanen, N., Scheller, E., Sharshina, N., Vinogradova, G., Wilbur, J., Zhivotova, E. & Zolotuchina, N. Kola Saami Documentation Project. Linguistic and ethnographic documentation of the en-



working platforms or webpages for preliminary solutions of data publication and sharing, sometimes resulting in differing, but unlinked and/or non-continuing versions⁹. Last but not least, language documentation projects do not always follow best practice in keeping primary data persistently interlinked with all respective data derivations. On the contrary, textual annotations are sometimes de-linked from the original multimedia recordings once the task of transcribing of fieldwork recordings is finished¹⁰. In this case, all additional future annotations (typically this concerns morphosyntactic tagging created with FLE_x or similar tools) rely exclusively on the quality of the preliminary transcriptions. Without the original recording as a reference, neither the transcription or the tertiary derived linguistic annotations can later be reconsidered or corrected. Furthermore, the resulting corpus data can only be searched and further analyzed in a predefined way (e.g. morphosyntactically). Adding completely new types of annotations, like phonetic transcriptions or transcriptions of the intonation patterns of the recorded sound, is impossible. This practice is therefore against one of the main pillars of contemporary language documentation, namely collecting, annotating, and making available of *multi-functional data*.

Language Documentation meets Language Technology

The language documentation projects I work with are concerned with the building of multimodal language corpora, including at least spoken and written (i.e. transcribed spoken) data and applying innovative methodology at the interface between endangered language documentation and endangered language technology. I understand language technology as the functional application of computational linguistics as it is aimed at analyzing and generating natural lan-

dangered Kola Saami languages // The Language Archive (TLA). DoBeS archive (Digital language archive). 2005–2017 – Nijmegen. URL: <https://corpus1.mpi.nl/ds/asv/?2&openhandle=hdl:1839/00-0000-0000-0005-8A34-E>].

⁹ Examples are the not completely overlapping Nganasan corpora presented in [Wagner-Nagy, B. & Szeverényi, S. Linguistically annotated spoken Nganasan corpus // *Tomsk Journal of Linguistics and Anthropology* 2. 2015 – 25–33] and at <http://www.iling-ran.ru/gusev/Nganasan>, or the corpus of Forest Enets archived as [Comrie, B. Shluinsky, A. & Khanina, O. Documentation of Enets. Digitization and analysis of legacy field materials and fieldwork with last speakers // *Endangered Languages Archive (ELAR)*. Digital language archive. 2005–2017 – London: url: <https://elar.soas.ac.uk/Collection/MPI950079>] and also published at <http://larkpie.net/siberian-languages/forest-enets>.

¹⁰ Examples are the annotated corpus of spoken Beserman Udmurt, <http://urn.fi/urn:nbn:fi:lb-2015081401> (which otherwise is one of the very few positive examples of easily accessible and easily searchable Uralic spoken corpora!) and the annotated corpora of spoken Eastern Khanty and Southern Selkup, <https://elar.soas.ac.uk/Collection/MPI43298>.



guage in various ways and for a variety of purposes. Language technology can create tools which analyze spoken language corpora in a much more effective way, and thus allow one to create better linguistic annotations for and descriptions of the endangered languages in question. A morphosyntactic tagger applied in corpus building is but one example of such a practical application. Using automated tagging, rather than the non-automated methods described in the previous section, allows for directing more resources towards transcription and for working with larger data sets because slow manual annotation no longer necessarily forms a bottleneck in a project's data management workflow.

At present, I am mostly working with Kildin and Skolt Saami as well as with Komi-Zyrian. Another language investigated in a project affiliated with The Freiburg Research Group in Saami Studies¹¹ is Pite Saami. We are developing a common framework for our different language projects in order to systematically apply methods from Natural Language Processing (NLP) to enrich the respective corpora with linguistic annotations. In order to do so, we rely on the following two main principles, which we have begun implementing consistently in our own documentation projects:

- use an orthography-based transcription system,
- apply computer-based methods as much as possible in creating automated higher-level annotations (tagging) of the compiled corpus data.

While keeping track of versions and persistently linking original recordings with annotations and other metadata, our corpora under construction (and under permanent refinement) are re-distributed both through a true language archive (for long-term persistency and safety) as well as through different other open community portals (for user-friendly and easy redistribution to various thematic user groups).

Whereas our local working platforms – building on ownCloud, GitHub, and SVN – are used for daily project collaboration, we use automated (script-based) routines to convert the data for the various outlets. This ensures both persistency, sustainability, and ease of use. Our long-term goal is to combine our internal project workflows with regular data releases to the digital archives we work with (i.e. ELAR and TLA), regular updates of the Internet-based presentations for the communities¹² as well as other potential outlets in the future, e.g. the digital research infrastructure of the Hamburg Center for Language Corpora (HZSK)¹³ and the linguistic corpus interface at Giellatekno in Tromsø¹⁴.

¹¹ <http://saami.uni-freiburg.de>

¹² One example is <http://videocorpora.ru>

¹³ <https://corpora.uni-hamburg.de/hzsk>

¹⁴ <http://gtweb.uit.no/korpus>



Orthographic Transcription

For us, using orthographic transcriptions not only allows for quicker and more efficient transcription of field recordings but makes it also possible to easily integrate all available (digitized or digital) non-spoken texts into the corpus. In addition, any available (digitized or digital) lexical resources can be integrated into the annotation tools under creation as well, rather than building new dictionaries from scratch via interlinearization (which is the typical approach by projects using FLE_x or similar tools).

Note, however, that not all digitized texts or dictionaries are written using standard orthography. In particular, the samples of spoken language collected by linguists and printed as transcriptions (and translations) in books typically use non-orthographic writing systems, such as the Finno-Ugric transcription system. The same is true for descriptive dictionaries targeting a scientific audience. While we keep the original transcriptions after digitization whenever feasible, normalizing these into standard orthography is the most effective way to integrate all different spoken texts into a single writing system and then to combine these texts with available written corpora, which of course use standard orthography.

Automated Corpus Annotation using NLP

Although the dominating paradigm within NLP today is based on statistical methods we use a rule-based («grammar-based») approach. The approach chosen by our projects is a combination of using finite-state transducer (FST) technology for the morphological analysis, and Constraint Grammar (CG) for the syntactic analysis. This has been tested with several written languages, for which it routinely provides highly robust analyses for unconstrained text input. We adapt the open preprocessing and analysis toolkit provided by the Gielatekno¹⁵ project (cf. also [9]).

The annotation process works on three levels: 1) Preprocessor-Tokenizer; 2) Morphological analyzer, programmed as an FST for modeling free and bound morphemes as well as the linear and non-linear rules according to which morphemes combine in word formation and inflection; 3) Syntactic disambiguator, written as a CG.

Note that the lack of a higher-level analysis often leads to cases of ambiguity concerning the morphological analysis, i.e., multiple homonymous analyses for one and the same word form. For the syntactic disambiguation of these homonyms, we use CG. It takes the morphologically analyzed text as its input, and returns only one reading, ideally the appropriate one. CG is a language-independent formalism for morphological disambiguation and syntactic analysis of text corpora developed by [10]. The CG analysis can be enriched with syntactic functions and dependency relations if all underlying grammatical rules are

¹⁵ <http://giellatekno.uit.no>



described sufficiently. Since the output of a CG analysis is a dependency structure for a particular sentence, the output may also be converted into phrase structure representations.

Our work with the CG description is only at the beginning stage. So far, we have only started with Komi. However, the experience of other Giellatekno projects working with CG shows that some months of concentrated work can already result in a CG description that can be implemented in a preliminary tagger useful for lexicographic work as well as for several other purposes. This means that the approach is readily adjustable to language documentation projects with limited resources.

Summary

While the rule-based morphosyntactic modeling is not state of the art in contemporary computational linguistics, it does have significant advantages specifically in endangered language documentation: 1) the results of the automatic tagging are *exceptionally precise* and cannot normally be reached with statistical methods applied on the relatively small corpora we work with; 2) while incrementally formulating rules and testing them on the corpus data, we are not only creating a tool but producing a *full-fledged grammatical description based on broad empirical evidence* at the same time; and last but not least 3) our work will eventually also help develop new ICALL technology, i.e. *computer-assisted language learning* systems for the languages we work on. One relatively simple example is the direct implementation of our description in spell-checkers, which is a easily done with the Giellatekno toolkit.

Last but not least, since the official support and language planning activities are significant at present for some of the languages we are working on, they are increasingly used in spoken and written form. Better adaptation of computational technology by language documenters will eventually be necessary in order to more efficiently annotate and make use of the increasing amount of data available.

Ultimately, our work will therefore also contribute to future language planning, language development, and language vitalization.

* * *

1. Blokland R., Gerstenberger C., Fedina M., Partanen N., Rießler M., Wilbur J. Language Documentation meets language technology// First International Workshop on Computation Linguists for Uralic Languages (16 January, 2015, Tromsø). – Tromsø, 2015. – P.8–18; Gerstenberger C., Partanen N., Rießler M., Wilbur J. Utilizing language technology in the documentation of endangered Uralic languages // Northern European Journal of Language Technology. – 2017; Gerstenberger C., Partanen N., Rießler M., Wilbur J. Instant annotations Applying NLP methods to the annotation of spoken language docu-



mentation corpora // Proceedings of the Third International Workshop on Computational Linguistics for Uralic languages (23 January, 2017, St. Petersburg) – St. Petersburg, 2017; *Gerstenberger C., Partanen N., Rießler M.* Instantnt annotation in ELAN corpora of spoken and written Komi, an endangered language of the Barents Sea region // Proceedings of the Second workshop on the Use of Computational Methods in the Study of Endangered Languages (6–7 March, 2017, Hawaii) – 2017.

2. *Himmelman N.* Language documentation. What is it and what is it good for? // Essentials of Language Documentation / ed. J. Gippert. – Berlin, 2006. – P.1–30; *Woodbury A.C.* Language documentation // The Cambridge handbook of endangered languages/ ed. PK Austin. – ambrige, 2011. – P.159–186; *Austin P.K.* Language documentation in the 21st century // Journal LIPP. – 2014. – P.57–71. (URL: www.lipp.up.lmu.de/article/download/190/83).

3. *Austin P.K.* Language documentation and meta-documentation // Keeping languages alive. Documentation, pedagogy and revitalisation / ed. M. Jones. – Cambridge, 2013. – P.3–15.

4. *Himmelman N.* Linguistic data types and the interface between language documentation and description // Language Documentation & Conservation: Vol.6. – Manoa, 2012. – P.187–207.

5. *Cox C.* Corpus linguistics and language documentation. Challenges for collaboration // Corpus-based Studies in Language Use, Language Learning, and Language Documentation / ed. J. Newman. – Amterdam, 2011. – P.239–264.

6. *Trosterud T.* Grammar-based language technology for the Sámi Languages // Lesser used Languages & Computer Linguistics (13 – 14 November, 2006, Bozen). – Bozen, 2006. – P.133–148.

7. *Rießler M., Wilbur J.* Documenting the endangered Kola Saami languages // Språk og språkforhold i Sápmi / ed. T. Bull. – Berlin, 2007. – P.39–82.

8. *von Waldenfels R., Daniel M., Dobrushina N.* Why standard orthography? Building the Ustya River Basin Corpus, an online corpus of a Russian dialect // Computational linguistics and intellectual technologies. Proceedings of the Annual International Conference «Dialogue» (4–8 June, 2014, Bekasovo). – M., 2014. – P.720–729.

9. *Moshagen S., Pirinen T.A., Trosterud T.* Building an open-source development infrastructure for language technology projects // Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALODA 2013) (22–24 May, 2013, Oslo). – Linköping, 2013. – P.343–352.

10. *Karlsson F.* Constraint Grammar as a framework for parsing unrestricted text // Proceedings of the 13th International Conference of Computational Linguistics (August 20–25, 1990, Helsinki). – Helsinki, 1990. – P.168–173; *Karlsson F., Voutilainen A., Heikkilä J., Anttila A.* Constraint Grammar. A language-independent system for parsing unrestricted text. – Berlin, 1995.



J. Rueter,

Ph.D., University of Helsinki

FIELDWORK AND EARLY LITERARY TEXTS (FELT)*

Fieldwork and early literary texts are oftentimes the sources of initial linguistic research. These sources, however, are not limited to the broadening of our understanding of a language, but usually serve to develop our comprehension of the people and their society in alignment with cultural and historical perspectives. Making fieldwork and early literary texts available provides not only for a scientific audience but for the interested layman, as well. It helps a society to intensify its internal networks by offering a mutual point of venue for the researcher and the native speaker.

Electronic accessibility is seen as a guarantee and protector of shared knowledge that can be developed. By making research discoverable, accessible and queryable, places such as the University of Helsinki, in Finland, are protecting their academic work against plagiarism, on the one hand, and providing a lower threshold for quotations, on the other. How do they do it? Where possible theses, in part or in their entirety are made freely available on the web through an «e-thesis» service: <https://ethesis.helsinki.fi/>. This enhanced presence of academic materials helps these writings and their authors to be seen, and therefore the chances of their being cited is much greater than is the case of singular publications only found on a shelf. The key concepts here are: being discoverable, accessible and queryable.

Literature from the 1920s, 1930s and early 1940s can be found in the libraries and archives of the Russian Federation that is still not discoverable. Fortunately, the extensive digitization program and project efforts described by the Russian Federal Archiving Agency <http://www.rusarchives.ru/> have found contributors and collaborators both domestic and international. From July 2012 until October 2013, a Kindred Language Digitization Pilot funded by Kone Foundation, in Finland, and actively directed by Jussi-Pekka Hakkarainen, brought a new dimension to digitization collaboration between the National Libraries of Finland (Helsinki) and Russia (Saint-Petersburg). A push was made to proffer minority-language non-central publications a higher priority in the digitization process. The idea was to preserve publications from crowded archives

© Rueter J., 2017

* Печатается в авторской редакции.



for posterity; documents representative of a time in history when indigenous literature was beginning to flourish; literature which in subsequent years has been delimited to narrower spheres. The fruit of this endeavor can be found by searching the sites of these initial collaborators: (Saint-Petersburg) http://www.nlr.ru/rlin/ruslbr_v3_e.php, and (Helsinki) <https://fennougrica.kansalliskirjasto.fi/>. Documents are available primarily as searchable pdf documents, but for some of the materials there are also data packages available with further variants: Alto XML, CSV, TXT and TIFF.

Tangible extensions to this work can be found in the subsequent years of collaboration and further development of strategies. In addition to the original partners in Saint-Petersburg and Helsinki, partnerships have also been made with other libraries and archives: The Komi National Library, National Library of Yamal Nenets, National Library of the Udmurt Republic, National Library of the Karelian Republic, Murmansk State Regional Universal Scientific Library, Tver Regional Universal Research Library named after A.M. Gorky, Göttingen State and University Library, University of Tartu Library, just to name a few. The initial target languages from the Finno-Ugric/Uralic family have been joined by nearly a score of others. An open-source pdf revizing tool has also been developed.

Scientific societies in Finland, such as the Finno-Ugrian Society, in Helsinki, have seized the opportunity to make their early publications more discoverable and accessible on line (see <https://www.kansalliskirjasto.fi/fi/uutiset/tieteellisten-seurojen-julkaisu-avoimeen-kayttoon> and specifically: <http://fennougrica.kansalliskirjasto.fi/handle/10024/89724>). In the spring of 2016, the Finno-Ugrian Society, in Helsinki <http://www.sgr.fi/>, opted to push for the digitization of published research materials ahead of what might have otherwise been a twenty years wait for the scheduled digitization of all Finnish scientific series. The Society has seen digitization as an opportunity to return research knowledge recorded in the fieldwork of the nineteenth century to the individual language communities from which they had originally been acquired.

It was also in the spring of 2016 that I myself began the FELT pilot project dealing with a portion of the publications made available by the Finno-Ugrian Society. The aim of my project was to make materials specific to the Mordvin collections of Heikki Paasonen more discoverable, accessible and queryable in modern corpus infrastructure, such as that available publicly in the auspices of the Finnish Language Bank (Kielipankki: <https://www.kielipankki.fi/>) at <https://korp.csc.fi> or the analogous http://gtweb.uit.no/u_korp/ on the Giella infrastructure, in Tromsø, Norway.

The Pilot consisted of establishing practices, steps and infrastructure for an efficient presentation of transcribed fieldwork materials on line. Here is an enumeration:



- (1.1) The materials are digitized.
- (1.2) Proofreading is practiced at various points in the process.
- (1.3) An xml schema is adopted that is compatible not only with old and new fieldwork, but can be used with early literary texts, as well.
 - (1.3.1) The material is divided into individual poems, songs and stories to allow for interview-specific metadata, e.g. the interviewees, interviewers.
 - (1.3.2) The digitized transcriptions of fieldwork texts are aligned with their translations (German and possibly Russian), thus providing for the rudiments of parallel corpora work in the future.
 - (1.3.3) The transcribed fieldwork texts are the basis of normalized variants of the language. This serves the needs of queryability and semi-automatic annotation of these texts, and is a practice that can be applied to early literary texts, as well.

(1.4) The normalized texts are analyzed semi-automatically, where possible, and the resulting analysis is stored in xml format.

Digitization can be both time consuming and expensive depending on the approach. In light of what had been achieved in the Kindred Language Pilot at the National Library of Finland, the FELT pilot opted to have all eight volumes of Mordvin folk poetry (SUST 77, 81, 84, 99, 161, 162, 176 and 178) digitized in collaboration with the Finno-Ugrian Society and the National Library of Finland. The digitization covered the entire content of the Erzya and Moksha folk poetry series, including introductions and notes beyond the fieldwork materials and their translations; over 4000 pages representing the Erzya and Moksha traditions.

Proofreading focuses on the presentation of the language materials that lack automated proofing strategies. The first transcribed Erzya materials were proofread by a native researcher, Olga Erina, Ph.D., who was able to make extensive notes on word usage and mythological terminology during the process. The German-language materials were initially checked only for blatant misspellings, with additional proofing confined to what a German spell checker revealed. An analogous strategy was applied to the Russian. After normalization of the Erzya texts, a copy of them was spell checked with an open-source finite-state spell checker available presently for the open-source LibreOffice from the Giella infrastructure in Tromsø, Norway at http://divvun.no/static_files/oxts/myv_LO-voikko-5.0.oxt. (Open-source spell checkers for many other languages can also be found at <http://divvun.org/proofing/proofing.html> with frequent updates.)

Establishing an xml schema that will be compatible with fieldwork and literary materials alike requires some insight. One must select a medium of presentation to be targeted and an outline of properties important to a larger user group. Familiarity with the corpora concordance afforded by korp, developed at the University of Gothenburg <http://www.gu.se/english/research/publication?p>



ublicationId=156080, and awareness of its being open-source were fundamental in its selection as a target presentation platform for fieldwork texts, even those with little or no annotation.

A preliminary test was made with short transcribed fieldwork texts for Erzya, Moksha and Komi-Zyrian. In addition to metadata in the head containing interview, publication, proofreading and license specifics, and paragraph and sentence element reference to position, three text sets were established for each sentence element. Two of the text sets were rendered as attributes, containing the original transcriptions and translations, whereas the third, a generated normalized text, was presented as text value, with one word form or punctuation mark per line. In (2.1-2.3) only the original transcription and translation data is present.

(2.1) (SUST LXXVII)

<sent pgNo="0005" pgLi="4" type="li" orig="mastor langks karmí
šuro šačoma" tr1="und auf Erden beginnt Getreide zu wachsen,"/>

(2.2) (SUST 178)

<sent pgNo="0258" pgLi="30" type="li" orig="pidek tejne keli
pačat" tr1="Webacke uns breite Pfannkuchen,".>

(2.3) (SUST 221)

<sent pgNo="0008" pgLi="12" orig="voasni turj ordę." tr1="Sie kommen
bein Kranich an."/>

In (3.1-3.3) the text content of the element has been derived by normalizing the transcribed fieldwork text. This is the first version of data made available in the Finno-Ugrian Society corpora: SUS-Fieldwork. All three text sets can be utilized in queries.

(3.1)

<sent pgNo="0005" pgLi="4" type="li" orig="mastor langks karmí
šuro šačoma" tr1="und auf Erden beginnt Getreide zu wachsen,">

мастор

лангс

карми

сюро

чакомо

</sent>

(3.2)

<sent pgNo="0258" pgLi="30" type="li" orig="pidek tejne keli
pačat" tr1="Webacke uns breite Pfannkuchen,">

пидек

тейнек

кели

пачат

</sent>



(3.3)

<sent pgNo=»0008» pgLi=»12» orig=»voasnı turı ordę.” tr1=»Sie kommen bein Kranich an.”>

воасны

тури

ордö

.

</sent>

The idea of a normalized text field for queries allows for larger integrated corpora. It provides for the comparison of older texts with modern ones, and thus it provides for a better understanding of multi-dimensions in language development, research and revitalization. The Erzya and Moksha materials of Paasonen are representative of 200 distinct collection points: When considering the five basic dialect types hypothesized for each of the two languages, the categorization of these locales might require a research project of its own. Normalization scripts are written several tasks: (a) the transformation of the Finno-Ugrian transcription to the Cyrillic alphabet – language specific; (b) normalization of local phonetic development – dialect or locale specific, and (c) restoration of hyper-normalized lexical forms to their actual standard orthographic representation. Where certain solutions are already in place, it's better to use them, e.g. the phonetic rendering of Komi-Zyrian might best be transformed into the Molodtsov alphabet, which already has a converter at FU-Lab in Syktyvkar: <http://fu-lab.ru/convertermolodcov>.

The Moksha sentence variants in (4.1-4.2) represent semi-automatically annotated xml data and a slightly more reader-friendly form for transformation into korp data. The individual w=word elements include the following attributes: wf= word form; l= lemma; pos= part of speech; mds= morpheme analysis; synt= syntactic analysis, and dep= dependency.

(4.1)

<sent pgNo=»0258” pgLi=»30” type=»li” orig=»**pidęk tejńek keli pačat**” tr1=» Webacke uns breite Pfannkuchen,”>

<w wf=»пидек” l=»пидемс” pos=»V” mds=»V+IV+Imprt+ScSg2” synt=»FMV” dep=»1->0”/>

<w wf=»тейнек” l=»минь” pos=»Pron” mds=»Pron+Pers+Pl1+Dat” synt=»<ADVL” dep=»2->1”/>

<w wf=»кели” l=»кели” pos=»A” mds=»A+Sg+Nom+Indef” synt=»>N” dep=»3->4”/>

<w wf=»пачат” l=»пача” pos=»N” mds=»N+Pl+Nom+Indef” synt=»<OBJ” dep=»4->1”/>

</sent>



(4.2)

<sent pgNo=»0258» pgLi=»30» type=»li» orig=»**pidek tejnek keli**
pačat” tr1=” Webacke uns breite Pfannkuchen,”>

“<пидек>” “пидемс” V IV Imprt ScSg2 @FMV 1->0

“<тейнек>” “минь” Pron+Pers+Pl1+Dat @<ADVL 2->1

“<кели>” “кели” A+Sg+Nom+Indef @>N 3->4

“<пачат>” “пача” N+Pl+Nom+Indef @<OBJ 4->1

</sent>

‘Make us large pancakes’

This last representation (4) would be an ideal achievement, but there is still much syntactic analysis to be automated at present. And our project is still in its first half with many things still to be done. What we can offer are some pointers to resources, tools and infrastructural work that is open-source and in a place you can find it.

More places to notice and things to try out.

An introductory of what is to come of the Finno-Ugrian Society fieldwork in korp:

https://korp.csc.fi/?mode=other_languages#?lang=eng&prequery_within=sentence&cqp=%5B%5D&corpus=sust_myv,sust_kpv,sust_mdf

Collections of open publications with a high Uralic language content:

<https://fennougrica.kansalliskirjasto.fi/>

<http://www.doria.fi/bitstream/handle/10024/130799/Liite%201.pdf?sequence=3>

Site containing specifics for Permic languages, many spheres:

<http://fu-lab.ru/>

Open-source infrastructure for the research and development of language technology for Sami, Circum-polar, minority languages etc.

<http://giellatekno.uit.no/index.eng.html>

Open-source tools for minority languages:

<http://divvun.org/>

Open-source work in rule-based translation:

http://wiki.apertium.org/wiki/Main_Page



T. Trosterud,

*dr.art., Professor; Giellatekno,
The Arctic University of Norway*

LANGUAGE TECHNOLOGY IN RUSSIA*

Introduction

Russia has a 100-year history of active language planning, with approximately 50 languages having official status on republic or autonomous area level. The languages belong to the Uralic, Mongolic, Tungusic, Turkic, Caucasian or Paleo-Siberian language families, thus they are all languages with a very complex morphology. Written with the Cyrillic alphabet, almost all of them have letters not found in the Russian alphabet, several of them also with diacritical symbols. This has posed challenges for text publishing in these languages from the very beginning of the literary languages, and does so even more today, after the introduction of digital means of text production.

The present article discusses some of the challenges, and presents a language-independent open source infrastructure for morphology-rich languages, the Giella infrastructure.

Challenges

Language use, and especially the use of several languages in parallel, is increasingly dependent upon advanced language technology. We want to produce and process text in such a way that the content can be presented to us, in multiple languages if needed, and that linguistic content may be classified, retrieved, and translated by the computer in a reliable way. For a small number of the world's languages this development is well underway, the challenge for language planners is to bridge what has been called the digital gap, and make also other languages able to function in the same way.

The non-Indoeuropean languages of Eurasia differ from the languages that dominate the technological development in several ways. First, almost all of them have alphabets with letters added to the basic alphabets, letters that must be accessed in order to use the language at all. Second, they have a rich morphological structure, with every lexeme occurring in literally hundreds of different forms, making the type/token ratio high. Any advanced text processing algorithm will have to find a way to link the inflected word form to its underlying lexeme. Third, several of these languages have a limited access to text corpora. For an overview, see Тростерюд 2012 [1].

© Trosterud T., 2017

* Печатается в авторской редакции.



Letters and keyboards

The major obstacle for representing the world's languages digitally was overcome by the creation of Unicode. Now, all letters of all living languages can be found in the Universal Character set, and with the exception of some minor obstacles for a handful of languages, they can also be represented in text-processing software. In order to actually write the languages, the users still need good and standardised keyboard layouts, working in all programs and on all digital devices, from computers to tablets, smartphones and more.

Many languages have keyboard layouts from the age of the typewriter, layouts that need to be analysed and adjusted to digital keyboards. In Russia, virtually all languages use a slightly different set of letters, and the frequency of the individual letters will vary from language to language. Each language thus needs its own keyboard layout, as far as possible identical to previously existing typewriter layouts.

Morphological structure

Northern minority languages are dominantly agglutinative, but with a varying degree of morphophonological processes, where the shape of the stems and various affixes are dependent upon their environment. The fastest and most efficient way of analysing these complex word forms are by bidirectional finite-state transducers, capable of both analysing and generating word forms. The agglutinative morphology may be treated in one transducer, giving stem + suffix strings as output to lemma + analysis, or vice versa. Shape alternations of stem and affixes are then treated in a separate transducer, and the final analyser is made as a combination of these two.

This analyser will then act as the basis for all other applications for the language in question.

Syntactic structure

In many contexts, a morphological analyser is not enough. Whereas a generator for Komi may generate *из* as the imperative or con-negative form of *изны* 'to grind', we would often like to know whether the form *из* when encountered in a text really represents the verb and not a noun *из* 'stone'. In order to find that out, we need to disambiguate the text, to resolve the ambiguities offered by the language. This may be done in a way similar to human processing, that is, the program may look at the context in which the word form occurs. Thus, in order to decide what kind of *из* we have encountered, we may, in a given context, inspect whether it is preceded by a negative verb, like *оз* 'not', or by an adjective, like *гырысь* 'big'. In the first case, the verbal reading will be the most likely one, whereas the latter calls for the noun analysis. Equipped by such a program we may e.g. conduct a survey finding the most common lexemes in any given language, as opposed to only finding the most common word forms.



Turning grammatical analysers into applications

Grammatical analysers may be turned into working applications in numerous ways. The present article will present one way of doing it, a language-independent, open source infrastructure originally made by members of the research groups Divvun and Giellatekno, for a planned project on Tundra Nenets, Komi and Kildin Saami, the Giella infrastructure at UiT The Arctic University of Norway («giella» is North Saami for ‘language’).

The basic principle is to distinguish between language-specific and language independent source code. The language independent part is twofold: The first part is a setup of files made to compile the language-specific files into grammatical analysers, in the form of binary program files. These language independent files consists of compilers, different filters, and files to manage the different dependencies between source files and compilation processes. The second language-independent part consists of a similar setup, integrating the binary analysers into a wide range of practical application programs, ranging from morphologically enhanced electronic dictionaries via spellcheckers to machine translation programs.

An international network of linguists and language technologists have contributed to putting this infrastructure in work for the following 21 languages of Russia (for an overview, see Moshagen et. al. 2013 [2]):

- Bashkir, Buriat, Eastern Mari, Erzya, Evenki, Ingrian, Khanty, Kildin Sami, Komi-Permyak, Komi-Zyrian, Livvi, Mansi, Moksha, Nenets, Nganasan, Russian, Tatar, Udmurt, Veps, Votic, Western Mari.

The source code, i.e. both the linguistic and technical resources, are available as open source code. The linguistic coverage of each of the analysers varies considerably, but for about two thirds of the languages they are good enough to be utilised in one or several practical applications. The following sections will present some of the application types.

E-dictionaries

The Giella infrastructure offers a setup where the user looks up a (possibly inflected) word form. The web-based dictionary sends the word form to the morphological analyser, which finds the base form (the lemma) and forwards it to the dictionary. The dictionary finds the relevant lemma article, and presents it to the user. The user is then offered the following possibilities of generating a morphological paradigm (basic paradigm or all attested forms), or of looking up the word form or any of the forms in the paradigm, in a text corpus containing example sentences for the language in question. The dictionary may also be integrated into the web browser of the user, offering a click-in-text facility when reading web pages. An example is the Permian dictionary, found at <http://kyv.oahpa.no> («oahpa» being the name of the E-learning program for which the



dictionary setup was originally made). The setup behind these dictionaries is presented in Johnson et al. 2013 [3].

Spellcheckers

As pointed out above, the morphological analysers are transducers, with an upper side containing lexeme + grammatical analysis, and a lower side containing the set of word forms. By extracting the lower side and integrating it into word processors like LibreOffice or Microsoft Word, the analysers are turned into spellcheckers. The Giella infrastructure contains a pipeline for doing exactly that, and developers may actually build their own spellcheckers, and put them up for download for the user community. With access to the grammatical analysis of each word form, the spellchecker may also be augmented by grammatical information that can be used to give priority to certain suggestions over others (e.g. common grammatical categories over rare ones).

The Giella infrastructure contains several spellcheckers for free download or online use, but the main point in this context is that it offers a possibility for to make spellcheckers for any morphologically complex language, languages that due to their structure are excluded from simple list-based technologies (cf. Moshagen et al. 2014 [4]).

Grammar checkers

With access to the grammatical properties of the words in the sentence, it is also possible to make grammar checkers, programs that tells the user that a given word or phrase, although correctly spelled in itself, does not constitute a grammatical sentence, be it for reasons of person agreement, case declension, word order, or some other grammatical structure. For reasons of robustness, grammar checkers may only correct a subset of all grammatical errors. In addition to capturing grammatical errors, such a program will also capture typos disguised as real-word errors.

Within the Giella infrastructure, the work on grammar checkers is still in an initial phase, and confined to work on North Saami only.

Machine translation

The dominant paradigm within machine translation is statistically based translation, where one builds models for translation based upon the likelihood that a given string of word may represent some string in the target language. The empirical fundament for the model is a large bilingual corpus, aligned on a sentence level, as well as models of the target language, based upon large monolingual text corpora.

Northern Eurasian languages do not lend themselves readily to such an approach. First, the amount of parallel texts with other languages is small, and second, the morphological structure gives a high type/token ratio, where the chance of having seen a given word form earlier in the training material is low.



The Giella alternative to this is rule-based machine translation, analysing the source language with the same tools as are used for the other applications, accessing a bilingual dictionary and a component for grammatical transfer between source and target language, as well as a morphological generator (inverse analyser) for generating the target language sentences. Compared to statistical models, rule-based machine translation typically gives less idiomatic output, but adheres more closely to the actual message of the source language. The Giella infrastructure offers an integration with the Apertium rule-based machine translation platform. Apertium also contains several language pairs containing languages spoken in Russia, especially Turkic and Slavic language, whereas especially Uralic languages pair have been built as a combination of Giella and Apertium resources.

Grammatically annotated corpora for linguistic research

Using an interface made at the University of Gothenburg, the Giella infrastructure also offers morphologically and syntactically annotated text corpora for grammatical research. So far the text corpora for languages of Russia are small and contain no grammatical analysis, but the infrastructure is ready for it.

Conclusion

Language technology tools may be build in different ways, on different platforms. As long as the languages in question are morphologically rich, of the type found in Russia, it is not easy to come up with an alternative to the approach presented here. What the Giella infrastructure offers is an advanced, still open platform to work on, thereby cutting development time by years.

* * *

1. *Тростерюд Т.* Роль языковой технологии в сохранении и ревитализации языка // Саамская идентичность: проблемы сохранения языка и культуры на севере: материалы Международной науч. конф. (13 марта 2011 г., г. Мурманск). – Мурманск, 2011. – С.3–11.

2. *Moshagen Sjur N., Tommi A. Pirinen, Trosterud T.* Building on open-source development infrastructure for language technology projects // Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA 2013) (22–24 May, 2013, Oslo). – Oslo, 2013. – P.34–352.

3. *Johnson R., Antonsen L., Trosterud T.* Using finite state transducers for making efficient reading comprehension dictionaries. // Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA 2013) (22–24 May, 2013, Oslo). – Oslo, 2013. – P.59–71.

4. *Moshagen S., Rueter J., Pirinen T., Trosterud T., Tyers F.M.* Open-Source Infrastructures for Collaborative Work on Under-Resourced Languages // Workshop: Collaboration and Computing for Under-Resourced Languages in the Linked Open Data Era (28–30 May, 2014, Reykjavik). – Reykjavik, 2014. – P.71–77.



Dr. Joshua Wilbur,
Linguistics Researcher
Freiburg Research Group
in Saami Studies
Skandinavisches Seminar
University of Freiburg

THE PITE SAAMI LEXICOGRAPHIC BACKBONE FROM A FILEMAKER PRO DATABASE TO PUBLISHED DIGITAL RESULTS*

1. Introduction and Background

Pite Saami is a critically endangered Uralic language in the Samic branch of Finno-Ugric. It is spoken in Swedish Lapland, mainly in and around the municipality of Arjeplog. As a result mainly of on-going language shift to the dominant Indo-European languages Swedish and Norwegian over the last century, there are currently only a handful of individuals in Sweden who speak the language, while the last speaker on the Norwegian side died several generations ago [1].

While the size of the Pite Saami community has never been particularly large, the language has been the object of linguistics research on a fairly regular basis since the late 1800s¹. No pedagogical materials have been published for Pite Saami, although a few speakers have put together some basic teaching resources for their own use. None of the publications from before the turn of the millennium are in digital form, and, until recently, no standard orthography existed either. However, starting in 2008, a local lexicographic project and my own language documentation projects created a foundation for digital language technology resources for the Pite Saami, as outlined below.

Specifically, a group of Pite Saami language activists carried out a wordlist project titled *Insamling av pitesamiska ord* 'Collection of Pite Saami words' [2]. Roughly simultaneously, I began my documentation and descriptive work on the language. This proved to be a serendipitous opportunity, as the project's

© Wilbur Joshua, 2017

* Печатается в авторской редакции.

¹ Cf. [Bengtsson, Nils-Henrik, Marianne Eriksson, Inger Fjällås, Eva-Karin Rosenberg, Gry Helen Sivertsen, Valborg Sjaggo, Dagny Skaile, Peter Steggo & Joshua Wilbur (2016). «Pitesamisk ordbok». In: Pitesamisk ordbok samt stavningsregler. Ed. by Joshua Wilbur. Samica 2. Freiburg: Albert-Ludwigs-Universität Freiburg, pages 8–10] for a discussion of the previous literature on Pite Saami.



participants, with their native-speaker competency, combined with my formal linguistic training ultimately resulted in the publication of the first Pite Saami dictionary [3] and orthographic standard [4], together in a single volume [5]. As a further result of this cooperative work, additional digital linguistic resources are now available for Pite Saami. A searchable version of the ever-growing wordlist is available on-line. Language technology tools are under development for Pite Saami in cooperation with Giellatekno at the University of Tromsø² and based on those resources (specifically a Finite State Transducer and a Constraint Grammar, which will for instance enable the creation of spell checker). In the following sections, the infrastructure and workflow behind these various projects and their digital and analog products will be described, with a focus on how FileMaker Pro database software is being utilized with this aim in mind. This discussion includes a critical evaluation of the advantages and disadvantages of using FileMaker Pro in linguistics work, which can hopefully help inform other similar projects considering how to develop their own digital data infrastructure.

2. A lexicographic backbone in FileMaker Pro

As mentioned above, current lexicographic resources for Pite Saami consist of:

- a paper dictionary and orthography standard [6];
- a searchable online version of the wordlist: <http://saami.uni-freiburg.de/psdp/pite-lex>;
- language technology tools: <http://giellatekno.uit.no/cgi/index.sje.eng.html>

In the name of efficiency, I use the application FileMaker Pro to manage the data and automatically export these into formats which feed the three resources listed above. FileMaker Pro is proprietary commercial database software³ typically used for business solutions, but also very adequate for linguistic databases. It provides a typical graphical user interface (GUI) to manage relational databases on Mac OSX and Windows, and includes a particularly useful scripting feature as well as the ability to import from and export into other common formats. My original impetus to use FileMaker Pro originated in instances of FileMaker databases used by colleagues to manage typological feature data and historical-comparative lexical data. Currently, I have FileMaker databases to collect and manage data for inflectional paradigms, phonemes, and metadata for language recordings, among others.

[FIGURE 1 (image: FMPscreenshot-BandW.png) SOMEWHERE AROUND HERE]

² Cf. <http://giellatekno.uit.no/cgi/index.sje.eng.html>

³ <http://www.filemaker.com>



The screenshot displays the FileMaker Pro interface for the 'sje ortografi' database. The record shown is for the word 'mähtseârroj' (ID 3999). Key fields include:

- Identification:** 3999, mähtseârroj, Nr.
- Morphology:** M (ur spr.), EN (hermit), SV (eremit), NO (eremitt), syll# 4.
- Analysis:** st.-växl., omijud, ordklass (subst), variant (+), Samica status (not checked, check again, finalized).
- Export/Action:** MySQL export, Samica export, only Samica, not finalized, do not export, override no export.
- Lexical Data:** LexC table (N, V, Adj, Adv, PP), compound (mähtse+ârroj), pseudocompound (LexC).
- Notes/History:** my notes, orig. wordlist, Lule-dic, Lule cognate, Latin, added on 13.04.2010, last modified 24.01.2017.

Figure 1: Screenshot of the Pite Saami lexical database in FileMaker Pro, with numbered references (in circles) to the feature inventory provided in § 2.

However, most importantly, my collection of lexical data for Pite Saami is in a FileMaker Pro database; a screenshot of the main window can be seen in Figure 1 on page ?? . To provide an impression for the array of types of data that can be collected in this application, some of the data fields in my database that are particularly worth mentioning include:

1. unique identification number
2. lexical item
3. link to the lemma entry when the lexical item is a wordform
4. homonym number
5. translations



6. morphological patterns
7. part of speech
8. cognate in Lule Saami
9. syllable count
10. variant forms
11. audio example
12. compound boundaries
13. deviant stress pattern in loan words
14. cataloging data
15. backwards spelling
16. prosodic segment slots
17. inflectional class categorization
18. notes and comments

The screenshot in Figure 1 illustrates what the main database window looks like, and includes references to the types of data listed above. Note that many data must be entered by hand or via import from a compatible file (see below for more), but data is calculated automatically whenever possible using the FileMaker scripting feature. This includes for instance syllable count, inflectional class categorization, reverse spelling, phonological segment slots and even linking to the relevant audio file. Not every field is relevant or contains data for each entry. While the Pite Saami lexical database is not perfect⁴, it works quite well for my purposes.

The lexical item (2) is the central data slot of the database, although in fact – from a technical database standpoint – each record's unique identification number (1) is the piece of information that distinctly defines each record. Each lexical item is a specific wordform, but the majority of entries simultaneously represent a specific lemma; the latter is always the case when no link to any relevant head word is provided (3). Cases of homophony are differentiated using a homonym number in (4). English and Swedish (and sometimes Norwegian and, when relevant, Latin scientific nomenclature) translations are provided (5), particularly when the lexical item represents the lemma. The same is true for relevant morphological stem alternations such as consonant gradation, umlaut and stem extension patterns (6), and of course part of speech (7). Similarly, any Lule Saami cognates are listed in (8). The number of syllables an entry has (9)

⁴ One potentially sensible improvement to the design of the database, particularly from a lexicographic point of view, would be to separate wordforms and lemmas into two different tables. The current workaround of specifically marking entries containing a wordform that does not correspond to that wordform's lemma is a practical solution borne out of the fact that the original wordlist import did not distinguish the two either. However, this solution is not ideal.



is calculated automatically based on vowel patterns in the orthographic form. Any non-standard variants are listed as well (10). When available, a representative audio clip of the lexical item is linked into the database (11) and can be played directly in the application. If an entry is a compound, the morphological boundary between the compound's components is indicated in (12). The initial stressed syllable for loan words with a stress pattern deviating from the standard trochaic system is indicated in (13). A variety of cataloging data such as added-on and modified-on timestamps, the source of an item, and the publication status are included in a variety of places (14). Mainly because FileMaker Pro has a rather quirky and restricted built-in ability to search using regular expressions (see § 3.2), entries are spelled backwards in (15) in order to facilitate backwards sorting and right-to-left searches (particularly useful when filtering for specific suffixes). A rather complex set of FileMaker scripts analyses a given entry to extrapolate both consonants and vowels into the respective phonological segment slots in (16), as well as the inflectional class category used in the Pite Saami Finite State Transducer in (17). Finally, relevant notes consisting of free text are included in (18), in this case there is one field for notes from the original wordlist project and one field for my own notes.

2.1. Transformability between FileMaker Pro and other formats

One essential feature of FileMaker Pro as a database application for linguistics is its ability to be compatible with more common formats. The program can import from and export into the following established file types: *xml* (extensible markup language) and *csv* (comma separated values⁵) as well as the commonly used *xlsx*-format (Microsoft Excel spreadsheet format)⁶.

From the earliest stages of the database, this was a vital aspect because the data collected by the wordlist project mentioned above (cf. [7]) was in Microsoft Excel spreadsheet format. After receiving a digital copy of the wordlist as well as permission to use it in 2012, it was possible to import all entries (nearly 7,000) from the wordlist into my lexical database. This core set of lexical items formed the basis for the refined, edited versions that now exist online and in print.

Exporting allows the data to be presented in other ways. In the case of my Pite Saami lexical data, there are currently three main publication formats, as mentioned above and listed here again with the relevant data format indicated for each case:

- an online searchable lexical database; format: MySQL;

⁵ Tab separated values are also accepted.

⁶ In addition, FileMaker Pro also allows for importing and exporting using dbf (dBase III and IV DBF), merge (for instance with Microsoft Word Mail Merge) and exporting as an html-table.



- language technology analytical and generative tools; format: lexc;
- a print version; format: LaTeX.

As is clear from the beginning of this section, FileMaker Pro does not natively export into any of these formats, so an intermediate format is necessary to ultimately transform the data into these formats. The obvious solution is to choose xml export, then apply an xslt style sheet transformation to achieve the desired end format. During FileMaker export, it is possible to apply a style sheet directly, although even the most recent version of FileMaker Pro (released in December 2016) only allows style sheets using the quite outdated version 1.0 of xslt.

By utilizing the built-in FileMaker scripting language and the ability to include buttons in the GUI to trigger actions such as scripts, updating these various output formats when any changes or additions have been made to the lexical database is quite simple. For instance, when new lexical items are added, morphological information is modified or translations are improved in the FileMaker database, exporting the modified data into the necessary formats can be done at the click of a button, and the various output files are updated automatically. Specifically concerning the three main digital output formats mentioned above, this happens in the following ways. For the online searchable database, a MySQL file is created via xslt and then uploaded to the server in a separate action. The language technology files (in lexc, using the Giellatekno model) are created via xslt and then updated on the host server using svn (subversion control). The LaTeX source file is automatically created in the relevant local subdirectory and loaded without any further action while typesetting⁷.

A further useful, but indirect application of this workflow concerning the language technology files is worth mentioning briefly because it shows how this database can be used extensively in corpus annotation. As described in more detail in [8] and [9], updated lexical data can be used to automatically update corpus annotations using the language technology analytical tools via a script that applies the tools to ELAN⁸ annotation files. This way, changes made in the

⁷ The paper version of the Pite Saami dictionary was published in October 2016 as [Gerstenberger, Ciprian, Niko Partanen, Michael Rießler & Joshua Wilbur (2017). «Instant annotations. Applying NLP methods to the annotation of spoken language documentation corpora». In: Third International Workshop on Computational Linguistics for Uralic languages, 23rd January, Saint Petersburg/Russia. Proceedings of the workshop. Ed. by Tommi A. Pirinen, Michael Rießler, Trond Trosterud & Francis M. Tyers. ACL anthology, SIG proceedings for SIGUR. Association for Computational Linguistics (ACL). In press.], but new pdf-versions (the output of LaTeX) can be created this way and distributed as digital documents.

⁸ ELAN is free software used to annotate multimedia recordings; see <https://tla.mpi.nl/tools/tla-tools/elan/>. The Pite Saami corpus, along with Kola Saami and Izhva-



FileMaker lexical database can be applied to an entire corpus as well, thus easily keeping corpus annotations up to date.

3. FileMaker Pro as a linguistic database

As should be clear in the discussion above, using FileMaker Pro as a database solution for lexicography or other linguistic data can be a very useful tool for data management and linguistics research. Aside from the specific advantages outlined above, a number of points can be considered advantages of using FileMaker Pro, as discussed in § 3.1 below. At the same time, there are some potentially significant disadvantages that should be considered when trying to decide whether FileMaker Pro is the best database solution for a specific project; these are described in § 3.2 below.

3.1. FileMaker Pro: Advantages

Particularly for users unfamiliar with command line tools, the FileMaker graphic user interface (GUI) is a common format including features such drop down menus, data entry fields accessible via mouse click or tabbing, and many common keyboard shortcuts (such as copying and pasting text using cmd+C and cmd+V). While the layout of the GUI for any given database is highly customizable, and can thus be tweaked to meet an individuals precise needs, a number of formatting templates are also available for beginners or anyone who does not want any graphical customization. FileMaker Pro is particularly good at linking relational databases. Multiple specific relationships between data fields in tables within a single database or even between separate data sources can be defined quite precisely. The flexibility of both the GUI and the relational database links is reflected overall in the ability of FileMaker Pro to be adapted to any user's specific needs. Any individual project can, and indeed must define its own tables, data fields and the relationships between datasets.

Fields can be set to only accept certain data types or data formats, thus reducing the likelihood of inconsistencies or errors during data entry. Once a database exists, searches in FileMaker can be powerful and specific, potentially searching multiple fields simultaneously. Scripting is available for many functions, such as for abstracting data from other data, for further restricting a found set, or for exporting into other formats, just to name a few. Third-party plug-ins (such as functions going beyond the standard set shipped with the application) are also available, although I have not come across any that were particularly useful for my own linguistics projects.

Finally, there are also mobile apps available for iOS (iPhone and iPad) which allow databases to be portable. When using cloud hosting or a FileMaker

Komi corpora at the Freiburg Research Group in Saami Studies, consist of ELAN files.



server application, databases can be accessed by multiple users from multiple locations, which can be particularly useful for team projects.

3.2. FileMaker Pro: Disadvantages

While the discussion so far has focused on one example of how FileMaker Pro can be implemented for lexicography as well as some of the general advantages of the application, FileMaker Pro also has some potentially significant disadvantages that deserve mention as well. To begin with, it is proprietary, closed-source software. This has several implications. First of all, it will only be available as long as the company that produces it continues to support and develop it, while no community-driven solution exists independent of any ultimately profit-driven motivation. Indeed, because FileMaker database files are in a proprietary format, no community-driven solutions may ever exist, and compatibility with older versions or with other software is thus ultimately dependent on the will of the proprietors. Secondly, it is rather expensive⁹, and users without a significant software budget at their disposal may be prevented from buying it at all.

While the GUI is quite flexible, as mentioned above, *everything* in FileMaker Pro is only accessible via the GUI (this includes scripting), emphasizing the fact that it is designed for end users who do not write code of any kind. Anyone who is familiar with other scripting languages is likely used to – and even prefers – writing scripts in plain text format. FileMaker scripts are only accessible via the GUI, in other words, using the mouse and clicking on the appropriate fields or scripts in a drop-down menu¹⁰. In fact, essentially everything is GUI-based (as opposed for instance to the common plain text interface of MySQL databases), including defining tables and fields, field behavior, and relationships between tables, fields and databases.

Another serious fault and potential reason not to choose FileMaker Pro is the surprisingly poor set of regular expressions available for database searches. The regular expressions that do exist are very limited in scope and unique to FileMaker Pro (and thus take some getting used to for users used to other common regular expression standards). For instance, there is no way to search for particular character classes other than digits (such as all non-word characters), specific character sets (such as either the character *a* or the character *á*, or three or more occurrences of the character *t*)¹¹.

⁹ For instance, at the time of writing, a single user license in the German on-line store costs € 415, although an academic license is significantly cheaper at € 250.

¹⁰ It should be noted however that it is possible to trigger scripts written in AppleScript, although these are only meaningful for communication with applications or systems external to FileMaker Pro.

¹¹ In other frameworks, such searches can easily be performed for instance using Java regular expressions such as: `\w`, `[aá]` and `t{3,}`, respectively.



One final aspect worth mentioning is the steep learning curve involved in becoming familiar with FileMaker Pro. Because of the complexities of relational databases and the unique GUI implementation of FileMaker Pro, it is safe to predict that an average user without previous experience with FileMaker will require a significant time investment in understanding how to set up a FileMaker database for lexicography, other linguistic data, or indeed any database at all.

4. Summary

The application FileMaker Pro makes it possible not only to collect and manage lexical data for Pite Saami in a customized structure, but allows for the (semi-)automatic exportation of updated lexical data into a variety of other digital formats on a regular basis. This includes an on-line searchable lexical database (in MySQL), language technology tools for wordform analysis and generation (in lexc), as well as updated print versions of the database (via LaTeX to pdf).

While FileMaker Pro is not ideal, it has a number of clear benefits, the most obvious being precisely this ability to update changes into multiple output formats and infrastructures. However, a number of disadvantages should be considered as potentially prohibitive as well, including FileMaker's high price, its poor implementation of rather quirky regular expressions for searching, and its steep learning curve.

In summary, my project has benefitted greatly and continues to utilize this solution not only for lexical data but for other types of relevant data as well. However, due to the high cost, the potential risk of a discontinuation of this proprietary format and the reliance on the GUI without exception, in hindsight I would have seriously considered another solution, such as MySQL, particularly for lexical data management. It is my hope that other linguists can benefit from this description and discussion in planning their own solutions for managing project data, both for lexicography and other aspects of their work.

* * *

1. Bengtsson, Nils-Henric, Marianne Eriksson, Inger Fjällås, Eva-Karin Rosenberg, Gry Helen Sivertsen, Valborg Sjaggo, Dagny Skaile and Peter Steggo «Insamling av pitesamiska ord». (2008–2012).

2. Bengtsson, Nils-Henric, Marianne Eriksson, Inger Fjällås, Eva-Karin Rosenberg, Gry Helen Sivertsen, Valborg Sjaggo, Dagny Skaile, Peter Steggo, Joshua Wilbur. Pitesamisc ordbok // Pitesamisk ordbok samt stavningsregler / ed. J. Wilbur. – Freiburg, 2016. – P.13–121.

3. Blockland, Rogier, Ciprian Gerstenberger, Marina Fedina, Niko Paranen, Michael Rießler and Joshua Wilbur. Language documentation meets



language technology // First International Workshop on Computational Linguistics for Uralic Languages (16 January, 2015, Tromsø). – Tromsø, 2015. – P.8–18.

4. *Gerstenberger, Ciprian, Niko Partanen, Michael Rießler, and Joshua Wilbur.* Instant annotations. Applying NLP methods to the annotation of spoken language documentation corpora // Third international Workshop of Computational Linguistics for Uralic Languages (23 January, 2017, St. Petersburg). – St. Petersburg, 2017.

5. *Valijärvi, Riitta-Liisa, and Joshua Wilbur.* The past, present and future of the Pite Saami language. Sociological factors and revitalization efforts // Nordic Journal of Linguistics. – 2011. – № 34. – P.295–329.

6. *Wilbur, Joshua.* A grammar of Pite Saami. Studies in Diversity Linguistics. – Berlin, 2014. – 280 p.

7. *Wilbur, Joshua.* Pitesamisk ordbok samt stavningsregler. – Berlin, 2016. – Volume 2.

8. *Blockland, Rogier, Ciprian Gerstenberger, Marina Fedina, Niko Partanen, Michael Rießler and Joshua Wilbur.* Language documentation meets language technology // First International Workshop on Computational Linguistics for Uralic Languages (16 January, 2015, Tromsø). – Tromsø, 2015. – P.8–18.

9. *Wilbur, Joshua.* Stavningsregler // Pitesamisk ordbok samt stavningsregler. – Freiburg, 2016. – Samica 2. – P.123–197.

Научное издание

ЭЛЕКТРОННАЯ ПИСЬМЕННОСТЬ
НАРОДОВ РОССИЙСКОЙ ФЕДЕРАЦИИ:
ОПЫТ, ПРОБЛЕМЫ И ПЕРСПЕКТИВЫ

Сборник материалов
Международной научной конференции

16–17 марта 2017 г.
Сыктывкар

Редактор *М.С. Жилина*
Корректор *И.А. Тюкавина*
Компьютерная вёрстка *И.М. Титовой*

Подписано в печать 06.03.2017. Формат 60х84 1/16.
Бумага «Снегурочка». Гарнитура «Times». Печать – ризография.
Усл. п. л. 18,0. Заказ № 2.
Тираж 150 экз.

Адрес издательства:
Коммунистическая ул., д. 11, Сыктывкар, 167982

Отпечатано в ГОУ ВО КРАГСиУ