

ПРАВИТЕЛЬСТВО РЕСПУБЛИКИ БАШКОРТОСТАН  
МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БАШКОРТОСТАН  
МИНИСТЕРСТВО КУЛЬТУРЫ РЕСПУБЛИКИ БАШКОРТОСТАН  
АКАДЕМИЯ НАУК РЕСПУБЛИКИ БАШКОРТОСТАН  
БАШКИРСКАЯ ЭНЦИКЛОПЕДИЯ  
УФИМСКИЙ ФЕДЕРАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР РАН  
ОРДЕНА ЗНАК ПОЧЁТА ИНСТИТУТ ИСТОРИИ,  
ЯЗЫКА И ЛИТЕРАТУРЫ УФИЦ РАН  
ДОМ ДРУЖБЫ НАРОДОВ РЕСПУБЛИКИ КОМИ  
ЦЕНТР ИННОВАЦИОННЫХ ЯЗЫКОВЫХ ТЕХНОЛОГИЙ

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

Материалы II Международной научной конференции  
(Уфа, 11–12 декабря 2019 г.)

Уфа  
«Башкирская энциклопедия»  
2019

УДК 004:81-13:811  
ББК 81.1:73:81.2  
Э45

II Международная научная конференция «Электронная письменность народов Российской Федерации: опыт, проблемы и перспективы» проводится в рамках реализации Государственной программы «Сохранение и развитие государственных языков Республики Башкортостан и языков народов Республики Башкортостан» на 2019–2024 гг.

*Ответственный редактор:* Ахмадеева А. У.  
*Редакционная коллегия:* Сиразитдинов З. А., Бускунбаева Л. А., Ишмухаметова А. Ш., Шамсутдинова Г. Г.

Ответственность за правильность, точность и корректность цитирования, ссылок, достоверность информации и оригинальность представленных материалов несут их авторы. Мнение авторов может не совпадать с позицией редакционной коллегии.

Э45      **Электронная письменность народов Российской Федерации** : опыт, проблемы и перспективы : Материалы II Международной научной конференции (Уфа, 11–12 декабря 2019 г.) /под общ. ред. чл.-корр. АН РБ А. И.Акманова. – Уфа : Башк. энцикл. 2019. – 224 с.  
ISBN 978-5-88185-458-4

В издании представлены материалы II Международной научной конференции «Электронная письменность народов Российской Федерации: опыт, проблемы и перспективы», состоявшейся в г. Уфе. Приоритетными направлениями работы конференции явились актуальные проблемы корпусной лингвистики, разработки языковых программных продуктов, функционирования языка в сети Интернет, образовательных ресурсов и методики обучения национальным языкам на основе новейших IT-технологий.

Сборник адресован ученым, разработчикам лингвистических программных продуктов, преподавателям, аспирантам, студентам.

**УДК 004:81-13:811**  
**ББК 81.1:73:81.2**

ISBN 978-5-88185-458-4

© Ордена знак Почёта Институт истории,  
языка и литературы УФИЦ РАН, 2019  
© «Башкирская энциклопедия», 2019

## НАЦИОНАЛЬНЫЕ КОРПУСЫ ЯЗЫКОВ. ПРОБЛЕМЫ КОРПУСНОЙ ЛИНГВИСТИКИ

*Л. А. Бускунбаева  
г. Уфа, Россия*

### ПРИНЦИПЫ СБОРА ПОЛЕВЫХ МАТЕРИАЛОВ ДЛЯ ДИАЛЕКТНОГО КОРПУСА БАШКИРСКОГО ЯЗЫКА

В данной статье рассматриваются проблемы разработки диалектного корпуса башкирского языка, основные принципы сбора полевого материала.

**Ключевые слова:** диалект, диалектология, корпус, полевой материал.

Диалектный материал является бесценным источником для разработки вопросов современного состояния и истории формирования языка, позволяющий выявлять закономерности употребления или, напротив, исчезновения тех или иных языковых явлений (фонетических, лексических, морфологических либо синтаксических). Для всестороннего, синхронического и диахронического исследования языка говоров и диалектов нужен богатый полевой материал, охватывающий все имеющиеся говоры и диалекты башкирского языка. Именно для таких целей разрабатывается корпус диалектных текстов башкирского языка, который будет содержать обширный диалектный материал для изысканий в области грамматики народного языка [Сиразитдинов, Бускунбаева, Ишмухаметова, Шамсутдинова 2018, 154-160].

Каждый диалект является важнейшей частью башкирского языка. Однако сегодня влияние литературного языка на диалекты очень сильное, что, в первую очередь, проявляется через школьное обучение, процесс урбанизации, СМИ и т.д. Северо-западные говоры башкирского языка сильно подвержены влиянию татарского языка. Широкое распространение среди башкирского населения явления билингвизма (русский язык в РБ выступает в качестве основного языка-посредника, обслуживающего многонациональное население республики) также приводит к снижению ценности диалекта. К сожалению, под воздействием данных факторов происходит активация процесса исчезновения диалектных особенностей языка.

Носителями диалектных особенностей выступают, прежде всего, пожилые люди, проживающие в сельской местности. Именно в их речи сохраняются основные фонетические, лексические и грамматические диалектные черты, в то время как в речи молодого поколения диалектные особенности слабо выражены. Следовательно, одной из основных задач башкирских языковедов является сбор диалектных материалов с широким охватом носителей с учетом поло-возрастных и других социальных групп для дальнейшего всестороннего анализа.

Сотрудники Лаборатории лингвистики и информационных технологий отдела языкознания ИИЯЛ УФИЦ РАН работают над диалектным корпусом башкирского языка. На данном этапе осуществляются сбор и обработка речевого материала по восточному диалекту башкирского языка.

Уникальность данного корпуса в том, что помимо транскрибированного диалектного материала, литературной формы и перевода на русский язык, будет возможность прослушать нужный аудиофрагмент.

Языковеды не ставят задачу охватить все фонетические, лексические и грамматические особенности рассматриваемого диалекта, а представить диалектную речь в его естественном звучании. Данный подход позволит исследовать не только диалектные особенности, но и исследовать звучащую речь, учитывая все ее особенности и нюансы. Например, наличие аудиоматериала дает возможность изучения ритмико-интонационного строения спонтанной речи, куда относятся интонация, паузы, логические ударения, темп, тембр и ритм речи, паралингвистическая информация (смех, плач, кашель, вздох), которые в башкирской лингвистике либо мало исследованы, либо вовсе не исследованы.

Отличаются методы и принципы сбора материала для диалектного корпуса от анкетирования с помощью программы-вопросника, используемого башкирскими учеными в ходе многочисленных диалектологических экспедиций.

Основным инструментарием до недавнего времени, используемом при сборе полевого материала, были анкеты и вопросники, в которые включались вопросы, нацеленные на получение необходимой информации. Участники экспедиций придерживались заранее заготовленных вопросов. Например, для сбора диалектного материала для «Диалектологического атласа башкирского языка» была составлена специальная программа-вопросник [Башкорт теленең диалектологик атласы өсөн... 1973], в котором основной акцент был сделан на выявление диалектных особенностей, образующих

лингвистические ареалы. На основе полученных результатов были составлены карты изоглоссных явлений, интерпретация изоглосс. Атлас дает ценные материалы по характеру и эволюции явлений в проекции географического распространения.

Как утверждают сами авторы, Программа не лишена и своих недостатков. Например, не уделено достаточного внимания синтаксическим особенностям в диалектах и говорах. Раздел синтаксиса представлен лишь тремя вопросами: формы дательного падежа, изафетная конструкция и система выражения указательно-определяющей конструкции [Диалектологический словарь башкирского языка 2005, 36]. В то же время одним из строгих требований в сборе материала для атласа, наряду с заполнением вопросника, была магнитофонная запись региональной речи.

Принцип сбора материала для диалектного корпуса башкирского языка отличается от прежних. Прежде всего, составляется “паспорт” текста, на основе которого разрабатывается металингвистическая разметка, которой будет снабжен корпус. Перед началом записи собирается полная информация, включающая следующие параметры: имя, фамилия, отчество информанта, возраст, образование, язык обучения, язык общения в семье, национальность, место последнего долгого проживания до переезда в данное место (в случае переезда), время проживания до последнего места проживания (в случае переезда), время проживания в данном месте, дата осуществления записи и др. [подр. см.: Сиразитдинов 2018].

Запись осуществляется на цифровой диктофон в несжатом формате. Для получения качественного звука и последующей удачной расшифровки исключаются любые посторонние помехи (телевизор, радио и др.).

Для записи информантов определены 15 тем:

- свадьба, свадебные обычаи;
- блюда (повседневные и праздничные);
- домашние животные (какие держат и как содержат);
- система родства (дети и близкие родственники);
- приусадебное хозяйство (огород, сад);
- дом, постройки (когда и кем построен, какая крыша, рамы);
- топонимия в окрестностях поселения;
- история села, школы, рода;
- повседневная жизнь (работа, школа);
- времена года, погода;
- малые формы фольклора (частушки, пословицы, поговорки, сказки);
- поездка в райцентр (по каким делам, каким транспортом пользуются);
- игры детей;
- друзья (кто они, где они живут);
- животный мир около поселения (какие птицы и звери обитают).

По мнению разработчиков диалектного корпуса башкирского языка, вышеперечисленные темы охватывают практически все стороны жизни сельского населения: семья, работа (учеба), постройки, хозяйство, жизнь села, дорога, погода, животный и растительный мир и др., что позволяют выявить диалектную лексику и специфику местных говоров.

Перед началом записи информанту предлагается тема, на которую он должен дать достаточно развернутый ответ.

В качестве основной формы записи информанта выбран монолог.

В “Словаре лингвистических терминов” монологическая речь определяется как “форма речи, обращенной говорящим к самому себе, не рассчитанной на словесную реакцию другого лица. Монологическая речь, в отличие от диалогической, характеризуется своей развернутостью, что связано со стремлением широко охватить тематическое содержание высказывания, наличием распространенных конструкций, грамматической их оформленностью” [Розенталь 1985].

Монолог, как отмечает В. В. Виноградов, имеет следующие особенности:

- монолог есть продукт индивидуального построения;
- он характеризуется последовательным изложением мыслей, что основано на логико-смысловых связях между предложениями;
- монолог характеризуется самостоятельным расположением речевых единиц;
- отличительная черта монолога – его непрерывность;
- монологическая речь характеризуется относительной развернутостью;
- большей произвольностью;
- меньшей зависимостью от собеседника [Виноградов 1963, 96].

Выбор монолога в качестве основной формы для сбора диалектного материала обоснован тем,

что он дает возможность самостоятельного выбора языковых средств для выражения своих мыслей. Такая монологическая речь обладает признаками спонтанности и неподготовленности, непринужденности, что выражаются “на всех языковых уровнях, и наличии особого рода сбоев в ее гладком построении” [Зайдес 2017, 28]. У информанта нет времени на обдумывание и построение структуры рассказа. Активное же участие интервьюера в беседе может оказать определенное воздействие на речь информанта, поскольку он начинает использовать литературные средства выражения, “подражая” исследователю.

Как известно, разговорная речь диалогична. Как заметил Л. В. Щерба, “монолог является в значительной степени искусственной языковой формой, и что подлинное свое бытие язык обнаруживает лишь в диалоге” [Щерба 1915, 3]. Следовательно, при сборе материала для диалектного корпуса башкирского языка в качестве основной формы записи информанта должен выступать и диалог, что позволит интервьюеру получить более достоверную и обширную информацию о языковых особенностях того или иного говора, при этом направляя информанта в нужное русло.

Одним из основных требований при сборе диалектного материала является качество записи, которое зависит от многих факторов. Большую роль играет коммуникабельность информанта. Если один информант на конкретную тему может выдать свыше пятидесяти фраз, то другой на ту же тему – только десять. Немаловажным является и расположенность информанта к разговору в конкретной ситуации (смог ли интервьюер расположить к себе информанта, от настроения информанта в день записи и т.д.).

Если в начале записи информант чувствует себя зажато и выдает односложные ответы и краткие фразы (*Нимә һөйләйем икән?* ‘Не знаю что рассказать’, *Мин бер намә лә белмәйем* ‘Я ничего не знаю’), то, включившись в процесс записи, выдает текст, разнообразный по своей структуре и языковым средствам выражения.

Надо признать, что под воздействием различных факторов диалекты постепенно утрачивают свои характерные черты. И создание диалектного корпуса башкирского языка, включающий в себя богатый диалектный материал, станет бесценным источником для последующих изысканий в области фонетики, лексики и грамматики башкирских говоров и диалектов башкирского языка.

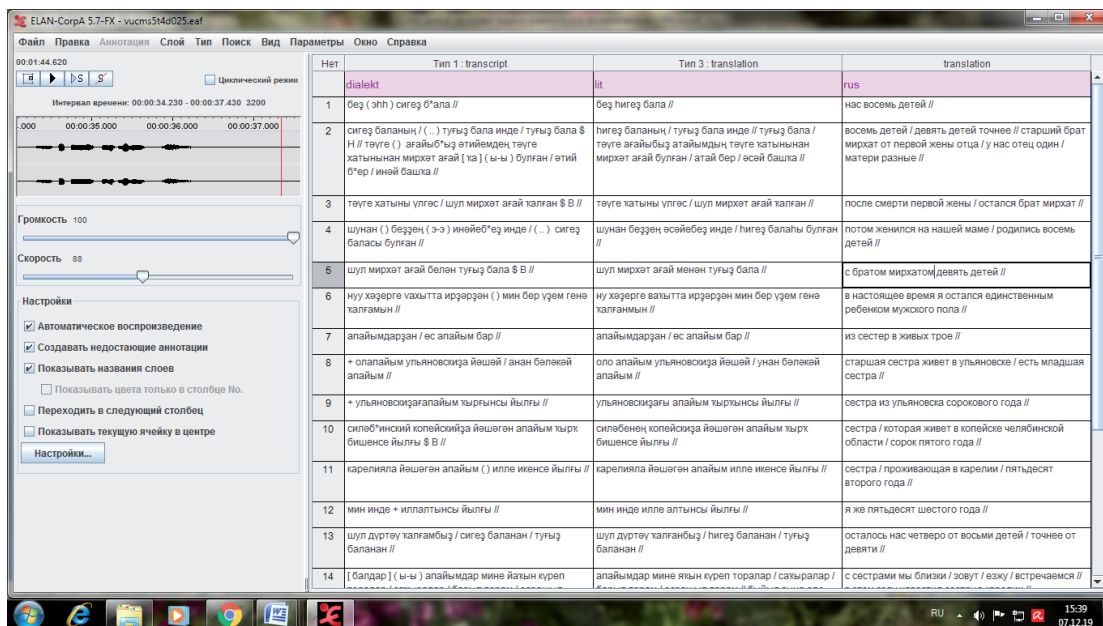


Рис. 1. Обработка аудиоматериала в программе ELAN

Уникальность данного проекта в том, что он пополняемый, т.е. диалектные материалы могут быть внесены в базу данных по мере их записи и транскрибирования, что позволит проследить изменения и во временном отрезке.

Наличие аудиоматериала в корпусе исключает неосознанное домысливание материала расшифровывающим в соответствии с собственными гипотезами или в силу неразборчивости речи информанта. Если данное явление и наблюдается, у исследователей есть возможность прослушать звуковую дорожку с нужными примерами вновь и вновь.

### Литература

- Анкета. Башкорт теле һөйләштәренән материалдар йыйыу өсөн. – Өфө, 1961. – 56 б.
- Башкорт теленен диалектологик атласы өсөн мәғлүмәт йыйыу программаһы. – Өфө, 1973. – 32 б.
- Виноградов В. В. Стилистика. Теория поэтической речи. Поэтика. – М.: Изд-во Акад. наук СССР, 1963. – 256 с.
- Диалектологический словарь башкирского языка. – Уфа: Гилем, 2005. – 234 с.
- Зайдес К. Д. Метакоммуникация в русской устной спонтанной речи. – СПб, 2017.
- Розенталь Д. Э. и др. Словарь лингвистических терминов. – М., 1985: [Электронный ресурс]. URL: <http://rus-yaz.niv.ru/doc/linguistic-terms/fc/slovar-203.htm> (дата обращения: 15.11.2019).
- Сиразитдинов З. А., Бускунбаева Л. А., Ишмухаметова А. Ш., Шамсутдинова Г. Г. Проблемы разработки диалектного корпуса башкирского языка // Вопросы диалектологии. 2018. № 1-2. – С. 97–106.
- Щерба Л. В. Восточно-лужицкое наречие, т. 1. – Пг., 1915. – С. 3 и 4 приложения.
-

## РАЗРЕШЕНИЕ МОРФОЛОГИЧЕСКОЙ МНОГОЗНАЧНОСТИ В КОРПУСЕ ТАТАРСКОГО ЯЗЫКА НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ

В данной статье описаны результаты экспериментов по разрешению морфологической многозначности в Национальном корпусе татарского языка «Туган тел» на основе машинного обучения с использованием различных инструментов на основе вероятностных и нейросетевых моделей. Представлены предварительные оценки точности разрешения в зависимости от различных параметров.

**Ключевые слова:** разрешение многозначности, морфологическая многозначность, корпус языка, татарский язык, вероятностные методы, нейронная сеть.

Татарский национальный корпус «Туган тел» [<http://tugantel.tatar/>] является информационным ресурсом современного татарского литературного языка объемом более 180 млн словоупотреблений. Помимо основного корпуса, включающего тексты различных стилей и жанров, в состав ресурса входят общественно-политический подкорпус, подкорпус текстов IT-тематики и параллельный русско-татарский подкорпус. Тексты в корпусе снабжены морфологической разметкой, выполняемой автоматически с использованием собственного модуля морфологического анализа [Gilmullin and Gataullin 2017]. На данном этапе одной из основных задач является разрешение морфологической многозначности и автоматизация этого процесса. Помимо применения непосредственно в самом корпусе, разрабатываемые методы разрешения многозначности могут быть использованы в других задачах автоматической обработки языка.

В настоящее время используются методы автоматического разрешения многозначности, основанные на правилах, методы машинного обучения, использующие вероятностные модели, а также гибридные методы [Гатауллин 2016; Gataullin 2017]. Примерами эффективных инструментов являются PurePos 2.0 [Orosz and Novák 2013], реализующий гибридную модель на основе скрытых Марковских моделей, а также нейросетевая модель на основе рекуррентных нейросетей с долгой краткосрочной памятью LSTM [Qinlan Shen et al 2016]. Инструмент PurePos разрабатывался для языков со сложной морфологией, в том числе для агглютинативных языков, а также для языков с малой ресурсной базой. Основные эксперименты проводились для венгерского языка [Orosz and Novák 2013]. Авторами было выдвинуто предположение, что модель способна работать и для других языков, в том числе для языков тюркской группы.

Эксперименты с нейросетевой моделью на основе рекуррентной нейронной сети с долгой краткосрочной памятью (long short-term memory, LSTM) описаны в работе [Qinlan Shen et al 2016]. Для обучения модели требуются размеченные тексты со снятой многозначностью. Авторы также отмечают, что модель языконезависима. В качестве демонстрации этого утверждения модель была применена для разрешения морфологической многозначности в турецком, русском и арабском языках. Результаты, полученные авторами данного исследования с применением LSTM в целом не уступают, а в некоторых случаях и превосходят результаты других методов разрешения.

Нами были проведены собственные эксперименты с PurePos 2.0 и LSTM на материале Национального корпуса татарского языка «Туган тел». На начальном этапе работы из выборки текстов корпуса объемом более 21 млн словоупотреблений были получены статистические данные о частотности словоформ с наличием альтернативных (омонимичных) разборов, приведенные в таблице 1. Доля словоформ с омонимичными разборами составила 25,75%.

Таблица 1.

**Распределение вариантов морфологического разбора**

Варианты разборов	Количество	Доля в корпусе
Всего словоформ с омонимией	5650820	25,75%
2 разбора	4282108	19,51%
3 разбора	1045392	4,76%
4 разбора	296547	1,35%
5 и более разборов	26773	0,12%
Всего в выборке	21940452	100%

Для проведения экспериментов с использованием машинного обучения необходимо создание обучающего корпуса со снятой морфологической многозначностью. Нами в качестве данных для обучения использовался фрагмент морфологически размеченного общественно-политического подкорпуса снятой вручную морфологической многозначностью. Статистика подкорпуса приведена в таблице 2.

Таблица 2.

**Статистика обучающей и тестовой выборки**

	Обучающая выборка	Тестовая выборка
Количество контекстов (предложений)	54580	944
Количество токенов (включая пунктуацию)	600480	11655
Количество многозначных разборов	125480 (21%)	2527 (21%)
Количество уникальных словоформ	29953	2788
Количество уникальных лемм	7117	1226
Количество уникальных морфологических форм	1898	346

Ручное снятие морфологической многозначности общественно-политического корпуса выполнялось лингвистами с помощью Веб-инструментария для снятия морфологической многозначности в текстовом корпусе татарского языка [Гатауллин 2014]. По результатам проведенной работы было подготовлено 56524 предложения со снятой морфологической многозначностью.

Модели на основе PurePos и LSTM обучались только на обучающей выборке, тестовая выборка использовалась только для тестирования. Каждая модель обучалась на одной и той же обучающей выборке и проходила валидацию на одной и той же тестовой выборке. В таблице 3 приведены оценки точности распознавания леммы и аффиксальной цепочки с морфологической разметкой и без нее.

Таблица 3.

**Показатели точности распознавания лемм и аффиксальной цепочки**

Показатели	PurePos 2.0	LSTM
Точность распознавания леммы, % (без морфологической разметки)	95,77	не поддерживается
Точность распознавания аффиксальной цепочки, % (без морфологической разметки)	93,42	не поддерживается
Точность распознавания леммы, % (с морфологической разметкой)	97,8	96,94
Точность распознавания аффиксальной цепочки, % (с морфологической разметкой)	96,46	95,46

В таблице 4 представлены результаты оценки точности разрешения морфологической многозначности в тестовой корпусной выборке в зависимости от числа вариантов морфологического разбора омонимичной словоформы.



**Точность разрешения в зависимости от числа  
вариантов морфологического разбора**

Количество вариантов	PurePos 2.0	LSTM
n=2	1613 / 1826 = 88,33%	1545 / 1826 = 84,61 %
n=3	337 / 424 = 79,48%	268 / 424 = 63,21 %
n=4	125 / 192 = 65,1%	141 / 192 = 73,44 %
n=5	7 / 9 = 77,78 %	7 / 9 = 77,78 %
n=6	50 / 72 = 67,57 %	37 / 72 = 51,39 %
n=7	0 / 2 = 0 %	0 / 2 = 0 %
n=8	0 / 1 = 0 %	0 / 1 = 0 %
Общее	2132 / 2527 = 84,36%	1999 / 2527 = 79,1%

Учитывая ограниченный набор корпусных данных для обучения, результаты экспериментов с использованием машинного обучения с помощью инструментов PurePos 2.0 и LSTM показали достаточно хороший уровень точности разрешения морфологической многозначности, который составил 84,36% и 79,1% соответственно. Разрешение с помощью инструментария PurePos 2.0 показало более высокую точность разрешения практически для всех вариантов разбора, кроме n=4. Также PurePos может быть использован для распознавания нераспознанных морфологическим анализатором словоформ с точностью разметки леммы, экспериментально оцениваемой в 79%. По нашему мнению, более низкие показатели точности нейросетевой модели прежде всего связаны с объемом обучаемых данных.

В рамках продолжения данного исследования, машинное обучение планируется использовать в процессе снятия морфологической многозначности в Национальном корпусе татарского языка «Туган тел» путем комбинации предварительного автоматического разрешения и ручного постредактирования его результатов, что позволяет повысить производительность этого процесса и постоянно улучшать качество и увеличивать объем обучающих данных.

### Литература

Gataullin R., Khakimov B., Suleymanov D., Gilmullin R. (2017) Context-Based Rules for Grammatical Disambiguation in the Tatar Language. // N.T. Nguen et al. (Eds). ICCCI 2017, Part II, LNAI 10449, 2017, P. 529-537.

Gilmullin R., Gataullin R. Morphological Analysis System of the Tatar Language // N.T. Nguen et al. (Eds). ICCCI 2017, Part II, LNAI 10449, 2017. P. 519-528.

Orosz G. and Novák A. PurePos 2.0: a hybrid tool for morphological disambiguation. Proceedings of Recent Advances in Natural Language Processing, pages 539–545, Hissar, Bulgaria, 7-13 September 2013.

Qinlan Shen, Daniel Clothiaux, Emily Tagtow, Patrick Littell, Chris Dyer. The Role of Context in Neural Morphological Disambiguation // Proceedings of COLING 2016, Osaka, Japan, December 11-17, 2016. P. 181–191,

Гатауллин Р.Р. Аналитический обзор методов разрешения морфологической многозначности // Российский научный электронный журнал (Электронные библиотеки). Том 19, № 2 (2016). – С. 98-114.

Гатауллин Р.Р. Веб-инструментарий для снятия морфологической многозначности в текстовом корпусе татарского языка // Сохранение и развитие родных языков в условиях многонационального государства: проблемы и перспективы: материалы V Международной научно-практической конференции (Казань, 19-22 ноября 2014 г.). – Казань: Отечество, - С. 71-73.

Татарский национальный корпус “Туган тел”. Электронный ресурс: <http://tugantel.tatar/> (дата обращения: 10.10.2019).

**ПРОБЛЕМА РАЗГРАНИЧЕНИЯ ЧАСТЕРЕЧНОЙ ПРИРОДЫ  
СЛОВА «ЯКЫН» В ТАТАРСКОМ ЯЗЫКЕ  
(на материале корпуса татарского языка)**

В татарском языке вопрос о разграничении наречий и имён прилагательных была и остается одним из актуальных вопросов. Несмотря на большой арсенал работ, посвященных исследованию данной проблематики, некоторые её аспекты по сей день остаются открытыми. В рамках данной статьи на материале корпуса татарского языка исследуется частеречная природа слова “якын”.

**Ключевые слова:** морфология, грамматика, наречие, имя прилагательное, часть речи.

В. Р. Гафарова, Г. Ф. Гайнуллина,  
Казан ш., Русиә

**ТАТАР ТЕЛЕНДӘ «ЯКЫН» СҮЗЕНЕҢ СҮЗ ТӨРКЕМЕ  
ТАБИГАТЕН БИЛГЕЛӘУ МӘСЬӘЛӘСЕ  
(татар теле корпусы материаллары нигезендә)**

*Татар телендә рәвеш һәм сыйфат сүз төркеменә караган сүзләрнең мөнәсәбәтен, чиген, күләмен билгеләү хәзерге көндә дә актуаль. Бу өлкәдә күпсанлы эзләнүләр булуга карамастан, сорау тудырган фикерләр булуы бәхәссез. Бу мәкалә кысаларында татар теленең корпусы материаллары нигезендә “якын” сүзенең табигате ачыклана.*

XVIII гасырның ахырында чыга башлаган беренче татар грамматикаларында ук сүзләр төркемнәргә берләштерелеп өйрәнелә. Сүзләрне төркемләнүнең күп гасырлык тарихы булса да, бу өлкәдә бүгенге көндә дә бәхәсле һәм хәл ителмәгән мәсьәләләр бар. Бу мәсьәләгә үз чорында В. Хангилдин, Д. Тумашева, Ф. Хисамова, Ф. Ганиев, Ч. Харисова игътибар итәләр. В. Хангилдин фикеренчә, аз, күп, якын, ерак кебек сүзләрнең статусын билгеләү «аларның сыйфатлык яки рәвешлекләре шул ике сүз төркеменә хас төп грамматик билгеләрнең кайсы аларда көчлерәк булуын, күбрәк табылуын чамалау аша» хәл ителергә тиеш [Хангилдин 1954, 84–85]. Д. Тумашева грамматик билгеләре белән охшаш булган бу сүз төркемнәрен аеруда лексик-семантик якка өстенлек бирергә кирәк дип саный. «Әгәр дә сүзнең семантикасында билге төсмере көчле булса, аларны, аергыч яки хәл булып килүләренә карамастан, сыйфат дип карарга кирәк: *сойкемле кыз, сойкемле сөйләшә, авыр мәсьәлә, авыр чиш елә* һ.б. Мондый очракларны сыйфатларның рәвеш ролендә килүе һәм, киресенчә, (*күп кеше, татарча китап* һ.б.) рәвешнең сыйфат ролендә йөрүе дип бәяләү мәсьәләне хәл итүдән ерак тора...» [Тумашева 1964, 83]. Ф. Хисамова да шушы фикердә тора.

Ф. Ә. Ганиев мәсьәләне башка фәнни юнәлештә чишүгә омтылыш ясы. Рәвеш һәм сыйфатның охшаш кулланылышын галим конверсия ысулы белән сүз ясалышы буларак билгели [Татар грамматикасы 2016, 16].

Озак еллар татар телен укыту һәм сүзлекләр белән эшләрү дәвамында кайбер сүзләр безнең игътибарыбызны жәлеп иттеләр, үзләренә кызыксыну уяттылар. Бу хезмәтебездә “якын” сүзенең мөнәсәбәтен ачыклау максаты куелды. Анализга күчкәнчә, бу сүзнең лексикографик чыганаclarда бирелешен күзәтик.

1981, 2005 елларда басылган “Татар теленең аңлатмалы сүзлек”ләрендә “якын” сүзе сыйфат сүз төркеменә карый, аңа түбәндәге аңлатма бирелә:

**Якын** сф. 1. Еракта урнашмаган, еракта булмаган, шушы тирәдәге, тиз генә барып житәрлек; киресе: ерак... // рәв. мәгъ. Монда гына, янәшәдә, терәлеп үк. 2. Тиздән булырга тиешле, тиз арада булачак. [Татар теленең аңлатмалы сүзлеге 1981, 628]

**Якын** сф. 1. Арасы еракта булмаган. *Я.урман*. 2. Тиздән булырга тиешле, тиз арада булачак. *Я.елда...* // 2. рәв. мәгъ. Янәшәдә генә. *Якында гына кемдер мыгырдый...* 3. ис. мәгъ. Дус, иптәш яки кардәш, туган кеше турында. *Эти, энинең якыннары*. 4. хәб. функ. Алар бик якын. 5. Бәйл функ. *Бер чиләккә якын сөт*. [Татар теленең аңлатмалы сүзлеге 2005, 727]

Шул ук сүзлекләрдә “якын” сүзенең антонимы “ерак” рәвеш сүз төркеменә карый, аңа түбәндәге аңлатма бирелә: Ерак рәв. 1. Зур аралык турында; киресе: якын. // с.мәгъ. Зур ара белән аерылган, ерактагы.// с.мәгъ.Зур аралыклы, озакка сузылган 2. Озак, озак вакыт белән аерылган. // с.мәгъ. Күптән үткән, күп вакыт элек булган. [Татар теленең аңлатмалы сүзлеге 1977, 332]

**Ерак** рэв. 1. 1) Зур ара турында; киресе: якын 2) сөйл. Озак, озак вакыт белән аерылган. (Киләчәктә булачак эш, хәл турында) *Экзаменга ерак калмаган*; 3.) күч. Күзгә ташланмый торган, ышанычлы урынга *Акчаны е. яшерерү...* 3. с. мәгъ. 1) Зур ара белән аерылган, ерактагы, озын күч. 2) Берәр нәрсәгә бөтенләй кагылышы яки бәйләнеше булмаган. *фәннән е. кеше*. 2. хәб. функ. Күп нәрсә житми, күп нәрә кирәк. *Аның кебек булырга сиңа ерак эле*. 3. сыйф. мәгъ. Зур ара белән белән аерылган, ерактагы. *Ерак авыл*. [Татар теленең аңлатмалы сүзлегә 2005, 158]

**Ерак** рэв. 1. 1) Зур ара белән; киресе: якын. Сагынсаң да юк чаралар, Ерак булгач аралар. 2) күч. Кеше күрмәслек, белмәслек итеп, ышанычлы урынга качырып. *Мин, ераграк тыгып куйган акчамны чыгарып, вакыт үтсен өчен, буфетка кереп утырдым, ашарга-эчәргә китерттем*. А.Шамов. 2. с. мәгъ. 1) *Ничек итсә иткән, [командирны] ерак кыр госпиталенә тапшырган* [Татар теленең аңлатмалы сүзлегә 2016, 239–240].

Татарча-русча сүзлекләрдә “якын” һәм “ерак” сүзләренең бирелеше аңлатмалы сүзлекләрдәге традицияне дәвам итә:

**Якын** (прил.) 1. прям., перен. близкий, ближний, ближайший, недалёкий, недалний; ~ урманга гына бару ходить только в ближний лес; ~ юл близкая дорога; ~ киләчәктә в недалёком будущем; ~ өметләр ближайшие надежды 2. близкий, ближайший, свой, родной, недалёкий, недалний (о родне, друзьях); ~ туган близкий родственник; иң ~ дус ближайший друг; ~ кешем юк минем нет у меня родного человека 3. в знач. (сущ.) мн. в притяж. ф. якыннары 1) близкие, родня, родственники; ана ягыннан ~нары родственники по материнской линии 2) приближённые; директорның ~нары приближённые директора [Татарча-русча сүзлек онлайн <https://intertat.tatar/syzlek/>]

**Ерак** (нареч.) прям., перен. далеко; алар бездән ~ торалар они живут далеко от нас; эле көзгә бик ~ до осени ещё очень далеко; миңа аның кебек булырга ~ шул мне ещё далеко до него; экрен барсаң, ~ китәрсәң (погов.) тише едешь, дальше будешь || (прил.) 1. в разн. знач. далёкий, дальний, отдалённый [Татарча-русча сүзлек онлайн <https://intertat.tatar/syzlek/>].

**Якын** пр 1. близкий // близко 2. близкий, родной // родственник, близкий 3. *посл* вроде, почти как, похоже 4. *Посл*. почти, около, с.

**Ерак** нар далеко, далёко // далёкий, дальний, отдалённый; длинный (дистанция) [Татарско-русский большой словарь <http://tatar.com.ru/dict/ya.php>].

Морфология буенча дәрәслекләрдә “якын” рәвеш сүз төркеменә кертелә. Ф. М. Хисамованың югары уку йортлары өчен язылган “Татар теле морфологиясе” дәрәслегендә тамыр рәвешләргә мисал итеп “ерак”, “якын” сүзләре китерелә [Хисамова 2015, 285]. Ч. М. Харисованың “Татарский язык: справочник” хезмәтендә урын рәвешләренә мисал итеп “ерак”, “якын” сүзләре күрсәтелә [Харисова 2014, 78].

Тел белемендә сүзләренә төркемләү, мәгълүм булганча, өч аспектка нигезләнә: лексик-семантик, морфологик һәм синтаксик. Тикшерү объекты булган “якын” сүзгә лексик-семантик яктан предметның һәм процессның билгесен белдереп килә: *Якын Көнчыгышта күзәтелүче вазгыять чыннан да демократиягә тартылу белән бәйлеме? (Татар-информ) Иң якын мәктәп - Зәйдә, анда 20 километр барасы. (Я. Шәфыйков) Әмма аның тән сакчылары мине якын да җибәрмәделәр. (Я. Шәфыйков) Мәктәп Бауман урамына якын гына урнашкан иде. (Ә. Хәсәнов)*

Морфологик яктан бу сүз дәрәжә кушымачалары ала, исемләшә: *Мин, аптырап, көзгегә якын-рак килдем һәм шунда гына, эремәгә салган фотокагазды сурәт ачыла барган кебек, көзгедә миңа бик таныш йөз пәйда булды ... (А. Тимергалин) Чөнки ул чагында тылдан килеп торган боеприпасларны монда бушатмыйча, алгарақ, фронтка якынак җиргә куачаклар. (Ә. Еники) Тантананы алып баручы Мөхәмәт хәзрәт галләмов каһарманның иң якын якыннарына сүз бирде. (Р. Биглов) Бу хат, канатлы кош кебек, ерак-еракларга – түземсезлек белән көткән якыннарына «оча». (И. Габдрахманова)*

Синтаксик яктан исә «якын» сүзгә исем сүз төркеме белән дә бәйләнешкә керә ала һәм жөмләдә аергыч функциясен башкара: *Беркөнне бер якын иптәше ярдәме белән Тукай икенче мөсафирханәгә күчәнә. (А. Зиятов) Фигыльләр белән бәйләнешкә кереп, хәл функциясендә дә килә: Шуннан соң абыйлары берничә тапкыр өйрәтеп карарга тырышсалар да, суга якын да килмәде. (Р. Галиуллин)*

Сыйфат һәм рәвешләргә хас булганча, “якын” сүзгә жөмләдә бәйсез позициядә хәбәр булып килә: *Ул татарстанлылыларга шуның белән икеләтә якын (Татар-информ). Чөнки бу тарих безгә бик якын. (Л. Әбүдәрова)*

Бу хезмәтебездә без «якын» сүзенң статусын билгеләү өчен татар теленең “Туган тел” корпусына нигезләнеп статистик мәгълүматлар тупладык. «Якын» сүзгә кергән барлыгы 1200 жөмлә тикшерелде: 244 жөмләдә «якын» сүзгә эш-хәлнең билгесен белдереп, фигыльне ачыкый; 399 жөмләдә – предметның билгесен белдереп, исемне ачыкый; 481 жөмләдә «якын» сүзгә бәйлек функциясендә кулланыла һәм 76 очракта – хәбәр функциясендә.

Фигыльне ачыкмый	Исемне ачыкмый	Бэйлек функциясендә	Предикат	Барлыгы
244	399	481	76	1200

Статистик мәгълүматлардан күренгәнчә, татар телендә «якын» сүзе исем сүз төркеме белән ешрак бәйләнешкә керә һәм жөмлэдә аергыч функциясен башкара, димәк, сыйфат билгеләре көчлерәк чагылыш таба. Бер антонимик рәткә кергән сүзләрнең төрле сүз төркеменә керү ихтималы татар теле өчен ят күренеш түгел дип саныйбыз.

#### Әдәбият

Татар грамматикасы: өч томда / проект жит. М.З. Зәкиев. –Тулыландырылган 2 нче басма. – Казан: ТӘҺСИ, 2016 – Т. II. – 432 б.

Татар теленең аңлатмалы сүзлеге. – Казан: “Матбугат йорты” нәшрияты, 2005. – 848 б.

Татар теленең аңлатмалы сүзлеге: I том: А – Й. – Казан: Татар. кит. нәшр., 1977. – 476 б.

Татар теленең аңлатмалы сүзлеге: II том: Г – Й. – Казан: ТӘҺСИ, 2016. – 748 б.

Татар теленең аңлатмалы сүзлеге: III том: Т – Ы. – Казан: Татар. кит. нәшр., 1981. – 832 б.

Татарско-русский большой словарь: [Электронный ресурс]. URL: <http://tatar.com.ru/dict/ya.php> (дата обращения: 02.10.2019).

Татарча-русча сүзлек онлайн: [Электронный ресурс]. URL: <https://intertat.tatar/syzlek/> (дата обращения: 02.10.2019).

Тумашева Д. Г. Хәзерге татар әдәби теле морфологиясе. – Казан: КДУ нәшр., 1964. – 300 б.

Хангильдин В. Н. Татар теле грамматикасы: морфология буенча очерклар. – Казан: Татар. кит. нәшр., 1954. – 335 б.

Харисова Ч. М. Татарский язык: справочник – Казан: Татар. кн. изд-во 2014, – 191 с.

Хисамова Ф. М. Татар теле морфологиясе: югары уку йортлары өчен д-лек. – 2 нче басма. – Казан: Тат. кит. нәшр., 2015 – 335 б.

## **ОБ ОСОБЕННОСТЯХ СОСТАВЛЕНИЯ КОРПУСА РАЗГОВОРНОЙ РЕЧИ КАЗАХСКОГО ЯЗЫКА**

В статье автор ставит задачу создания корпуса устной речи казахского языка, приводится обзор работ и возможных подходов при создании аналогичных корпусов по другим языкам. Автором выдвигается своя концепция создания корпуса устной речи, определяются возможные области использования такого корпуса, в частности в образовательном процессе, в сфере фундаментальных научных исследований, в разработке компьютерных приложений. Особо подчеркивается значение корпуса устной речи в создании системы автоматического распознавания единиц дискретной устной речи национального языка.

**Ключевые слова:** Корпусная лингвистика, корпус устной речи, компьютерная программа, речевая база, корпусная база, корпусное моделирование, речевой сигнал, языковой ресурс, акустическая вариативность, искусственный интеллект, автоматическое распознавание, фрейм, синтез речи, дискретная речь, непрерывная речь, фонемная/фонетическая/просодическая транскрипция, автоматическое синтезирование, эмоциональное содержание, технология речи.

*А. Қ. Жұбанов  
Алматы қ., Қазақстан*

## **ҚАЗАҚ ТІЛІНІҢ СӨЙЛЕУ КОРПУСЫН ҚҰРУДЫҢ ӨЗЕКТІЛІГІ**

Мақалада соңғы онжылдықтағы әлемдік тілдердің «Сөйлеу тілі корпустарын құру» мәселесіне қысқаша шолу жасалып, қолданылып келе жатқан әдістердің корпусдық модельдеу мен сөйлеу тілін автоматты синтездеу салаларына қарай ауыса бастағандығы жайлы сөз болады. Бұл жағдайдың сөйлеу тілінің просодикалық сипаттамасын, оның эмоционалды мазмұнын модельдеу үшін аса маңызды екені айтылған.

Сонымен бірге, мақалада, болашақта құрастырылатын қазақ тілінің сөйлеу корпусын тек ғылыми зерттеу мақсатында пайдаланумен бірге дискретті сөйлеу тілінің бірліктерін автоматты түрде тану жүйесін құру мәселесінің де шешімі табылатыны жайлы сөз болады.

**Тірек сөздер:** Корпустық лингвистика, компьютерлік бағдарлама, корпусар базасы, корпустық модельдеу, сөйлеу тілі корпусы, сөйлеу қоры, сөйлеу сигналы, тілдік ресурс, акустикалық вариативтілік, жасанды интеллект, автоматты тану, дыбыстық сигнал.

Қазақ тілін зерттеуде корпустық лингвистика саласының әлемдік дәрежедегі теориялық және практикалық жақтарын зерттеу қажеттігі туындайды. Әлем бойынша мәтін корпустарын құрастыру мен олардың қызметіне қатысты жалпы және нақты мәселелерге арналған мақалалар жарық көрген ғылыми журналдардың арнайы басылымдары да шыға бастады [Научно-техническая информация 2003].

Бірақ әлде де қазақ тіл білімі үшін корпустық лингвистикаға қатысты көптеген мәселелер арнайы зерттеуді қажет ететіні белгілі. Оған жататындар: корпустық лингвистика мен оның негізгі ұғымдарының анықтамалары, корпустық лингвистиканың тіл білімі құрылымында алатын орны, әдіс-тәсілдері және т.б. Сонымен бірге жаңа бағыттың теориялық негізін ұғыну мәселесі корпусарды нақты зерттеулерде пайдалануға қарағанда белгілі дәрежеде қалып келе жатқаны да байқалады.

Корпустық лингвистика пәнін осы саланың мамандары тілдік корпусарды құру мен оны пайдалану жағдайын зерттейтін тіл білімінің бір саласы ретінде ғана қарастырып келді. Кейбір ғалымдар ол пәннің түсінігін тар шеңберде қарастырып, оны тек компьютерлік лингвистика саласының аясында ғана түсіндіреді: «Корпусная лингвистика – раздел компьютерной лингвистики, занимающийся разработкой общих принципов построения и использования лингвистических корпусов (корпусов текстов) с использованием компьютерных технологий» – дейді [Захаров, Богданова 2011, 102].

Ал компьютерлік лингвистика ұғымын, әдетте, компьютерлік құралдарды пайдаланудың кең мүмкіндігі ретінде түсіндіруге болатыны белгілі. Бұл жердегі «компьютерлік құралдар» деп отырғанымыз – компьютерлік бағдарламалар, тілдік деректерді өңдеу мен компьютерлік технологияны орынды ұйымдастыру жұмыстары және т.б. [Баранов 2003, 59].

Ал корпустық лингвистика компьютерлерді тек «құрал» ретінде пайдалананады. Міне, сондықтан да корпустық лингвистика өзіне жүктелген міндетті ондай құралсыз атқара да алмас еді.

Бірақ компьютер мұндай рөлді қазіргі білім саласының барлық түрлерінде де атқаратынын ескерсек, онда олардың бәрін бірдей компьютерлік лингвистика саласына жатқыза беруге де болмайды.

Жоғарыда сөз болған корпустық лингвистика пәнінің теориялық және тәжірибелік жақтары қазақ тілі мәтіндері бойынша компьютерлік корпустар базасын құру жағдайында да ескерілуі қажет. Корпустық лингвистика қазақ тіл білімінің ерекше саласы ретінде қалыптасатын болса, қазақ тілі мамандарына көлемді тәжірибелік материалдарды пайдалануға, қажетті деген тілдік деректерді тауып алуға және оларға тиісті деген өңдеулер жүргізуге мүмкіндік туындатады. Осының бәрі қазақ тіліне қатысты зерттеулердің шынайылыққа (ақиқаттыққа) жетудің эмпирикалық тәсілдеріне жаңаша көзқараспен қарауға және ғылыми айналым аясына аса маңызды тілдік материалдарды енгізуге жағдай жасайды.

Қазіргі кезде әлемдік корпустық лингвистиканың даму сипаты – ұлттық толық мәтіндерді арнайы зерттеу нысаны етіп алу. Сондықтан автоматтанған қазақ тіліндегі мәтіндер корпусының компьютерлік базасы (теориялық және практикалық жағынан қарастырғанда) жақын болашақта жүзеге асатын «Қазақ тілінің ұлттық корпусының» аса маңызды бастамасы болары сөзсіз. Мұндай зерттеулердің нәтижелері қазақ мәтіндерінің стильдік, құрылымдық, мағыналық, функционалдық және т.б. сипаттарын анықтауда да өзекті мәселелердің бірі болып саналады.

Енді қазақ тіл білімінде әлі қарастырыла қоймаған «**Сөйлеу тілі корпусы**» жайлы осы мақалада қысқаша сөз етпекпіз.

Дыбыстама сөйлеу корпустарын деректердің сөйлеу қоры деп те атайды және оны тілдік ресурстардың маңызды бір түрі ретінде де санайды. Корпустың құрамына компьютерлік бағдарламаларды да қоса есептеу жиі кездеседі, оның себебі ондай бағдарламалар тілдік, оның ішінде фонетикалық ресурстарды құру, жинау, ұйымдастыру мен басқару әрекеттерін қамтамасыз ететін құрал. Сөйлеу корпустарын құруға қызығушылық бастама болған, негізінен, сөйлеу тілін автоматты түрде тануға қатысты жүргізілген зерттеулер аясы деуге болады. Себебі, бұл салада зерттеушілер тілдің дыбыстық бірліктерінің көптеген акустикалық вариативтілігімен жиі кездесіп отырады. Ал мұндай вариативтілік алуан түрлі дереккөздерде кездесетіні белгілі. Мысалы, оларға жататындар – сөйлеушінің немесе сөйлеу материалын жазуға арналған микрофонның сипаттамасына қатысты психофизиологиялық күйіне дейінгі жүйелік контекстік вариативтілік. Заманауи сөйлеу тілін тану жүйелері, әдетте, таспаға жазылып алынған көптеген дикторлардың (100-ден көп) аса үлкен дыбыстама сөйлеу ауқымдары (массив) арқылы үйретіледі. **Сөйлеу тілі корпусы** дегеніміз – сөйлеу тілінің құрылымданған бөліктерінің жиынтығы. Мұндай мәліметтермен корпус бойынша әрекет ету арнайы жазылған компьютерлік бағдарламалар арқылы қамтамасыз етіледі. Ал **сөйлеу тілінің бөліктерін** базалық бірлік ретіндегі сөйлеу сигналының цифрланған бөлігі деп және ассоцияланған сипаттағы ақпараттың бір түрі ретінде қабылдау қажет.

Қазіргі кезде сөйлеу корпусын көлемді, көп түрлі және ақпаратты иемдену жағынан бай (көп салалы), сонымен бірге құрастыру мен пайдалану жақтарын ұтымды ету іргелі фонетикалық зерттеулер үшін аса өзекті болуда.

Сол сияқты «Жасанды интеллект» саласындағы зерттеулер де қазіргі кезде әлем бойынша көпшілік ғалымдарға зор қызығушылық тудыруда. Атап айтқанда, жасанды интеллект саласы машиналық (компьютерлік) оқытумен тығыз қатынаста. Аталған сала, яғни жасанды интеллект саласы ғылыми тәжірибеде кең қолданыс табуда және тілдік бейнелерді автоматты түрде тануда көптеген мәселелердің шешімін табуда (Pattern Recognition). Тілдік бейнелерді автоматты тану дегеніміз – ол тілдік бейнелерді бірнеше категориялар немесе кластар бойынша топтастырумен айналысатын ғылыми пән. Мәселен, фонетика ғылым саласы бойынша сөйлеу тілін ойдағыдай тану үшін, әдетте, дыбыстық сигналдың фрейм деп аталатын бірнеше миллисекунд аралығындағы бейнесін ғана қарастырады екен.

Болашақта, сапалы түрдегі автоматты тану мен сөйлеу тілін синтездеу әдістері көлемді түрдегі оқыту саласына арналған компьютерлік ойындарға негіз болуы да мүмкін.

**Сөйлеу корпустарын топтастыру жайында.** Сөйлеу корпустарын құрастыру мен оларды пайдалануға қатысты жинақталған тәжірибе бірнеше белгілерді бөліп алуға мүмкіндік тудыра отырып, сөйлеу деректер қорын топтастыруға негіз болады және ол деректер жаңа сөйлеу корпусын жобалау кезінде ескеріледі. Енді ондай топтастырудың ең маңызды сипаттамаларына тоқталайық [Богданов, Кривнова, Подрабинович, Фарсобина 1998].

**Сөйлеу корпусын мақсатты пайдалану:**

- мамандандырылған, жалпы (репрезентативті), оқыту және иллюстрациялық;
- сөйлеу материалының түрі: дискретті сөйлеу, үзіліссіз сөйлеу және оқу, өздігінен сөйлеу (спонтанная речь), арнайы диалогтар;

– мәтіндік материалдар түрі: сөздер/буындар тізімдері, жеке сөйлемдердің жиынтығы, өзара байланыстағы мәтіндер; көптақырыптық (монотематикалық) немесе көпфункционалды;

**Сөйлеу сигналының түрі:** зертханалық сөйлеу, кеңсе сөзі, көпшілік алдында сөйлеу, телефон арқылы сөйлесу (әдетті немесе ұялы телефон арқылы, радио, теледидарлық сөйлеу).

– **дыбыстық сигналмен байланысты ақпарат түрі (аннотациялар):** орфографиялық жазба, фонемалық/фонетикалық транскрипция, просодикалық транскрипция, сигналдың акустикалық-фонетикалық белгіленімдер: «оқиғалық», сегменттік, просодикалық, лингвистикалық аннотациялар мен пікірлердің басқа түрлерінің болуы, мысалы, жеке ерекшеліктер туралы сөз иесінің сөйлеуі немесе сөйлеу бөліктерінің эмоционалдық бояуы;

– **тілдің дыбыстық бірліктерінің статистикалық теңдестіру түрі:** табиғи, біркелкі, репрезентативті (өкілдік), арнайы статистикалық сызбаға сәйкес;

– сөйлеу корпусының жадына енгізілген **қосымша сигналдық ақпараттың болуы және олардың түрі:** дыбыстық сигналмен қатар, қарапайым, мультимодальды және арнайы құрастырылған сөйлеу корпустары.

Әдетте, сөйлеу дерекқорлары көптілділік сипатта болып келеді. Сөйлеу корпустары тек барлық технологиялық маңызды тілдерге (американдық ағылшын, неміс, жапон, қытай және т.б.) ғана емес, сондай-ақ Еуропалық Одақтың ресми тілдерінің көпшілігінде: ағылшын, голланд, дат, швед, неміс, француз, итальян, испан және басқа да бірнеше тілдік корпусар үшін құрастырылған деуге болады.

Сөйлеу корпустары үшін жазылған *Corpnicus ELRA* корпусық бағдарламасының жүзеге асырылуының нәтижесінде Шығыс Еуропа тілдері (поляк, болгар, эстон, румын және венгер) үшін сөйлеу корпустары пайдаланыла бастады. Интернеттегі Еуропа Қауымдастығының веб-сайтынан орыс тіліне қатысты сөйлеу корпусын және оның мүмкіндіктерімен де танысуға болады. Мысалы, одан сөйлеу жағдайларын табуға мүмкіндік бар. Аталған орыс тілінің сөйлеу корпусын жүзеге асыру әрекетіне Санкт-Петербургтің “Одитек” компаниясы да өз үлесін қосты деуге болады.

**Орыс тілінің ISABASE сөйлеу корпусы** жайында қысқаша ақпарат.

90-шы жылдардың соңында Ресейдің ғылым академиясының жүйелік талдау институтында тек қана ғылыми мақсаттарға ғана емес, сонымен қатар, Мәскеу мемлекеттік университетінің филология факультетінің сөйлеу корпусымен айналысатын ғылыми топтарының қатысуымен, орыс тіліне қатысты мәтіндерге сөйлеу фрагменттерінің дыбыстық бірліктеріне шартты түрдегі белгіленімдер енгізу арқылы алғашқы орыс тілінің сөйлеу корпусы құрастырылып, қолданысқа ұсынылды. Бұл корпус тек ғылыми зерттеу мақсатына ғана арналған емес, сонымен бірге дискретті сөйлеу тілінің бірліктерін автоматты түрде тану жүйесін құру мәселесімен де айналысуға арналған еді [Богданов, Кривнова, Подрабинович, Фарсобина 1998].

**RuSpeech атты сөйлеу корпусын** құрастыру жобасы 2000–2001 жж. ИСА РАН тапсырысы бойынша *Intel* корпорациясының ғалымдарының күшімен жүзеге асқан болатын. Қазіргі кезде **RuSpeech** атты орыс тілінің сөйлеу корпусы ең өкілді деп саналады және басқа тілдердің сөйлеу корпустарын құрастыруға үлгі ретінде пайдалануға бағыт-бағдар беретін аса ыңғайлы корпус. **RuSpeech** жобаның ең маңызды нәтижесі деп, сөйлеу корпусын құрудың сыннан өткен технологиясы және осы технологияны қамтамасыз ететін бағдарламалық (компьютерлік) құралдар жиынтығы. Бұл жоба орыс тілін автоматты тану жүйесін құруға қатысты ғылыми-зерттеу жұмыстарын және т.б. қажетті деген әрекеттерді жүзеге асыру шарасында да пайдалануға болады [Богданов, Брухтий, Кривнова, Подрабинович, Строкин 2003]. Сонымен бірге, корпусқа қатысты келесі компьютерлік бағдарламаларды да атап кетуге болады:

– орыс тілінің транскрипторларын автоматтандыруға қатысты бағдарламаны дұрыстау (отладка программы);

– қажетті фонетикалық және статистикалық сипаттамаларына қатысты мәтіндік мәліметтерін жинақтайтын бағдарламаларды құру;

– эксперт-фонетистердің автоматтанған жұмыс орнын құру бағдарламасы;

– диктор сөздерін пакеттік жазу бағдарламасы;

– зерттеу жұмысының негізгі кезеңдерін верификациялайтын (анықтайтын, тексеретін) бірнеше бағдарламаны құру [Arlazarov, Bogdanov, Krivnova, Podrabinovitch 2004].

Соңғы онжылдықта сөйлеу тілін тануға қатысты байқайтынымыз, ол – «қол» ережелері мен алгоритмдеу әдістерінен корпусық модельдеу мен сөйлеу тілін автоматты синтездеу салаларына қарай ауыса бастағандығы. Бұл, әсіресе, сөйлеу тілінің просодикалық сипаттамасын, оның эмоционалды мазмұнын модельдеу үшін аса маңызды, сонымен бірге сөйлеуші дауысының жекелік ерекшелігіндегі еліктеушілігін модельдеу де аса құнды. Сөйлеу корпустары дербес тұрып-ақ ғылыми қызығушылық тудырады, ал әртүрлі тілдердегі дыбыстама сөйлеуді талдау (анализдеу) мен сипаттауға қатысты қажеттілік көптеген ғылыми мәселелерде туындайтыны мәлім.

Әлем бойынша алғанда, сөйлеу технологияларының екпінді дамуы компьютерлік бағдарлама мен іргелес ғылыми салалардан хабары бар кең түрдегі филолог мамандарын және мақсаттық бағыттағы фонетика мамандарын дайындауға қатаң талап қою қажеттігі. Осыған байланысты Қазақстан Республикасындағы жоғары оқу орындарындағы қазақ тілінің фонетика саласы кафедраларында «Қолданбалы лингвистика (сөйлеу тілінің технологиялары)» атты қосымша мамандар дайындау қажет.

Қорыта айтқанда, компьютерлік лингвистика мамандарының тұжырымдауынша, компьютерлік тілдік қор дегеніміз – ғылым адамының өз зерттеу нысанына жаңаша тұрғыда көз салу мүмкіндігі. Мұндай тілдік қор неғұрлым қомақты болса, солғұрлым тіл құрылысының сыры тереңірек ашылады, сөйтіп, зерттелетін нысан жөніндегі түсініктердің аумағы кеңиді, адамның білім өрісіндегі «ақтаңдақтардың» бедер-бейнесі айқындала түседі. Сол сияқты, зерттеуші адамның қалып-қабілеті әлденеше есе артады, шығармашылық қуат көздері ашыла түседі, сөйтіп, бұл жаңа мүмкіндіктер қазақ тілінің жүйелілік қасиеттерін жетілдіруге және тіл жүйесін мұқият тануға жұмсалатыны сөзсіз.

#### Әдебиет

Arlazarov V. L., Bogdanov D. S. Krivnova O. F., Podrabinovitch A. Ya. Creation of Russian Speech Databases: Design, Processing, Development Tools // International Conference SPECOM'2004. Proceedings. S-Pb. Russia, 2004. – Pp: 650–656.

Баранов А. Н. Компьютерная лингвистика // Баранов А.Н. Введение в прикладную лингвистику: Учебное пособие. – М.: Едиториал УРСС, 2003. – С. 13–38.

Богданов Д. С., Кривнова О. Ф., Подрабинович А. Я., Фарсобина В. В. База речевых фрагментов русского языка ISABASE // Сб. «Интеллектуальные технологии ввода и обработки информации». М., Едиториал УРСС, 1998.

Богданов Д. С., Брухтий А. В., Кривнова О. Ф., Подрабинович А. Я., Строкин Г. С. Технология формирования речевых баз данных // Сб. «Организационное управление и искусственный интеллект». М., Едиториал УРСС, 2003.

Захаров В. П., Богданова С. Ю. Корпусная лингвистика. Иркутск: ИГЛУ, 2011. – 161 с.

Научно-техническая информация. Серия 2. Информационные процессы и системы. 2003. №6, №10.



## ИЗ ОПЫТА СОЗДАНИЯ ГАЗЕТНОГО ПОДКОРПУСА КОМИ ЯЗЫКА

В данной статье пошагово описывается процесс создания газетного подкорпуса коми языка, процесс обработки текстов от исходных газет до включения его в базу данных письменного корпуса коми языка.

**Ключевые слова:** корпусная лингвистика, корпус языка, газетный подкорпус, обработка текстов.

Национальный корпус коми языка — это электронная справочно-информационная система на основе аннотированного корпуса текстов литературного коми языка с возможностью поиска языковых единиц по различным параметрам. Он состоит из систематизированной базы текстов на коми языке и специальных помет к нему, необходимых для филологического анализа этих текстов. Национальный корпус представляет коми язык на определенном этапе (или этапах) его существования и во всём многообразии жанров, стилей и т. п. В числе прочих в корпусе коми языка большое место занимает газетный подкорпус, куда включены тексты из комиязычных средств массовой информации (газеты «Коми му», «Йӧлӧга», «Парма гор», «Вьль туйӧд», «Звезда», Радио «Юрган»).

Для того, чтобы загрузить текст в корпус, нужно его специально обработать. На примере газеты «Коми му» (далее КМ) мы расскажем процесс работы и последовательность действий с текстом.

Газету нам направляют в формате PDF, поэтому для начала текст нужно извлечь оттуда. Есть два варианта извлечения текста из PDF файла. Первый способ — более долгий — с помощью программы FineReader, где нужно распознать текст, а затем вычитать его, исправляя возможные при распознавании ошибки. Второй способ более лёгкий и быстрый. Именно его мы используем в работе. Он предполагает извлечение текста в операционной системе Linux в программе Gedit. Почему именно Linux? Он позволяет извлекать текст без погрешностей. Если мы будем извлекать текст в Windows, то возникают проблемы. В некоторых словах текста появляются пробелы, то есть слово разделяется на два. Программа проверки орфографии не видит этих ошибок, поэтому текст, извлечённый в Windows, придётся читать, что займёт много дополнительного времени.

Первая строчка текста должна содержать выходные данные газеты: название газеты, дату и номер выпуска. Всю эту информацию можно найти на первой странице издания и оформлять нужно следующим образом:

==Коми му. Четверг, 2018 вося октябрь 11 лун, 42 (18347) №

Газета КМ имеет рубрики, то есть некие темы, подзаголовки, в соответствии с которыми выстроены статьи в газете. Если кому-то в дальнейшем нужно будет найти тексты определённой рубрики, это будет легко сделать. Для этого перед названиями рубрик ставится знак ==. Например:

==Веськыд сёрни

==Лун бӧрся лун

==Ас овмӧс и т. д.

Для того, чтобы после извлечения не было сплошного текста, перед каждой статьёй, заметкой и любым другим жанром газеты ставятся знаки {}. Это заготовка для заполнения выходными данными.

Многие статьи (но не все) имеют автора, и чаще всего их имена пишутся после статьи. Для выделения от текста и более удобного поиска в будущем, мы выделяем автора знаками []. Помимо фамилии и имени автора, нередко указывается его должность или род деятельности, также можно увидеть фамилию фотографа имеющих к статье снимков. Фамилия фотографа может быть написана и без имени автора статьи. Например:

[Григорий ХАТАНЗЕЙСКИЙ.]

[Гижӧдъяссӧ дасьтис Анастасия ЯРОВАЯ.]

[Галина РОЧЕВА, Мокчойса культура керкаын художествоа веськӧдлысь.]

[Павел КОРОЛЁВ. Снимокъясыс авторлӧн да ӧтуввезйысь.]

[Ангелика ЕЛФИМОВА. Снимокъясыс Инга УСАЧЁВАЛӦН архивысь.]

[Снимокъясыс ӧтуввезйысь.]

Как можно видеть, варианты могут быть разные.

Нужно отметить, что при извлечении текста возникают некоторые затруднения. При выделении текста захватываются сразу два столбика. Это происходит потому, что в КМ очень часто расположение

абзацев таково, что под широким столбцом располагаются два узких (Рис. 1). Выход из этой ситуации есть. Текст, который не получается извлечь, мы отмечаем знаками [], и извлекаем позже уже в Windows.

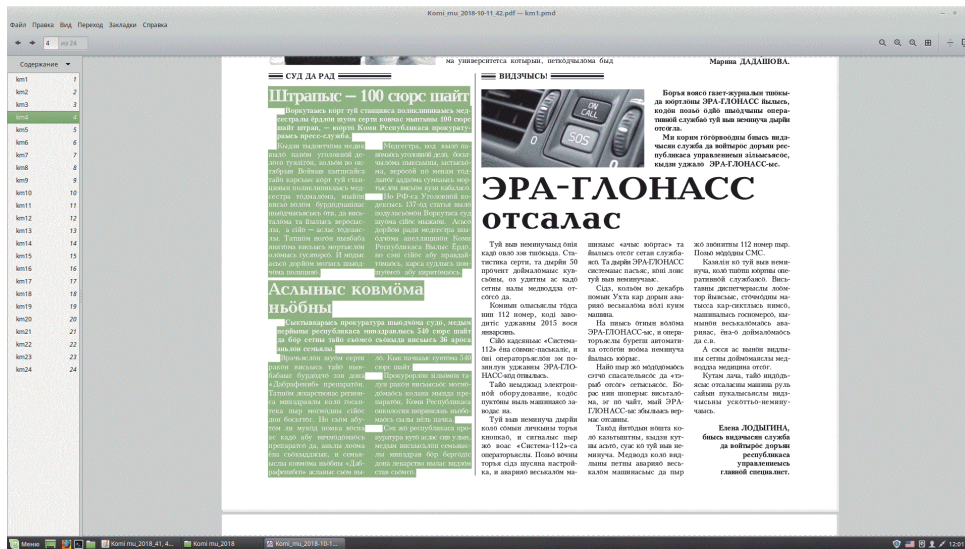


Рис. 1. Пример расположения двух узких столбцов под широким столбцом

После извлечения текста, сохраняем его и переходим в Windows. Основную работу над текстом мы проводим на базе двух программ: LibreOffice и Gedit.

С помощью программы проверки орфографии HunSpell, которая может работать самостоятельно либо на базе каких-либо программ (в нашем случае это LibreOffice и Gedit), мы без труда выявляем новые, не внесённые до сих пор в программу слова, варианты и формы слов; красной волнистой линией подчёркиваются некими слова и ошибки и опечатки в тексте. Это очень удобно, потому что не нужно прочитывать весь текст.

Мы вкратце попытались описать процесс обработки, который проходит текст, прежде чем его разместить в корпусе. Работа задействует две операционные системы, каждая из которых имеет недостатки и достоинства, и две программы. Эта работа требует внимательности и сосредоточенности.

## АРХИТЕКТУРА КОРПУСНОГО МЕНЕДЖЕРА И РАЗМЕТКА ТЕКСТОВ КОРПУСА ВЕПКАР<sup>1</sup>

Открытый корпус вепсского и карельского языков (ВепКар) представляет интерес не только для лингвистов, но и для диалектологов, поскольку диалектные тексты даны с подробной паспортизацией (указано имя информанта, год и место его рождения, место записи, диалект и так далее). В статье рассмотрен наименее проработанный людиковский подкорпус. Кратко представлена архитектура разработанного корпусного менеджера `dictorpus`, обслуживающего корпус ВепКар. Описана разметка текстов, семантическая и грамматическая связь между словами в текстах корпуса и словарными статьями. В планы развития корпуса входит семантическая категоризация значений лемм на основе сопоставительно-ономасиологического словаря и разработка морфологического анализатора на основе конечных автоматов.

**Ключевые слова:** вепсский язык, карельский язык, корпусная лингвистика, разрешение лексической многозначности, разметка текста.

### *История создания*

Проект «Открытый корпус вепсского и карельского языков» (ВепКар)<sup>2</sup> ведет свою историю с 2009 года, когда в Институте языка, литературы и истории Карельского научного центра РАН (ИЯЛИ КарНЦ РАН) под руководством д.ф.н. Нины Григорьевны Зайцевой началась работа по созданию Корпуса вепсского языка. В свое время «Корпус вепсского языка» помог в подготовке «Орфографического словаря вепсского языка» [ОСВЯ, 2012], поскольку обеспечил возможность поиска в диалектном материале при отборе форм для создания правил орфографии вепсского языка. В 2017 году проекту «Корпус Вепсского языка» была присуждена премия ФАДН «Ключевое слово» в номинации «За лучший научный проект».

Представителями вепсского и карельского народов решаются идентичные проблемы по возрождению языков и культур. Это требует значительных языковых ресурсов, которые позволят более надежно создавать и развивать правила также и карельской орфографии, введут в научный и общественный оборот значительное количество материалов, востребованных авторами учебников по языку и истории родного края, учителями, писателями.

В целях сохранения, развития и популяризации не только вепсского, но и карельского языка в 2016 году сотрудники ИЯЛИ и Института прикладных математических исследований КарНЦ РАН приступили к созданию многоязычного корпуса ВепКар на базе Вепсского корпуса. Карельский язык представлен в корпусе тремя наречиями (собственно карельским, ливвиковским и людиковским).

### *Архитектура корпуса и объем данных*

Корпусным менеджером в проекте ВепКар является разрабатываемая нами компьютерная программа `dictorpus`. Это программа с открытым исходным кодом, доступна для разработчиков онлайн (<https://github.com/comronavt/dictorpus>). Система поддержки корпуса разработана на платформе Laravel, язык программирования PHP. Данные хранятся в базе данных MySQL.

Архитектура корпусного менеджера `dictorpus` (рис. 1) включает корпус текстов и связанный с ним словарь. На ноябрь 2019 года в корпусе было 2600 текстов на 26 диалектах вепсского и карельского языков, свыше 900 тысяч словоупотреблений<sup>3</sup>. У некоторых текстов дан параллельный перевод на русский язык, десяток текстов сопровождается видеодорожкой. Словарь корпуса содержит свыше 30 тысяч словарных статей со словоформами на 18 диалектах. Толкования слов в основном даны на русском и английском языках, хотя есть возможность давать толкования на вепсском, наречиях карельского и финском языках.

<sup>1</sup> Статья подготовлена в рамках проекта РФФИ 18-012-00117 «Проблемы создания корпусов языков малочисленных народов России на примере Открытого корпуса вепсского и карельского языков»

<sup>2</sup> <http://dictorpus.krc.karelia.ru/>

<sup>3</sup> См. статистику корпуса: <http://dictorpus.krc.karelia.ru/ru/stats>



Рис. 1. Архитектура системы ВепКар, объём корпуса и словаря на ноябрь 2019 года

Вместе с текстом хранятся метаданные: информация о записи (информанты, собиратели, год и место записи), информация о месте хранения материала (обычно в Научном архиве КарНЦ РАН) и выходные данные публикации в печати.

Совместными усилиями идет разметка текста. Тексты разбиты на предложения. Если у текста есть параллельный перевод, то выполнено выравнивание текстов по предложениям (рис. 2). Предложения разбиты на слова (токенизация). Устанавливаются семантические связи (между токенами-словами в тексте и значениями лемм) и грамматические (между словами текста и тройками: (1) словоформа, (2) набор грамматических признаков, (3) диалект).

Система ВепКар при разметке текста автоматически ищет слова из текста среди словоформ из словаря и устанавливает связи между ними. Задача специалиста — проверить эти связи и подтвердить их корректность, либо выбрать верную связь из нескольких возможных, либо добавить новую связь.

Новую связь требуется добавлять в том случае, когда нет подходящей словоформы, либо нет словарной статьи (леммы), либо нет нужного значения в словаре корпуса ВепКар. В этом случае редактор создаёт соответственно новую словоформу, новую словарную статью или добавляет отсутствующее значение леммы.

Для автоматизации процесса пополнения словаря ВепКар словоформами были построены таблицы словоизменения для вепсского и собственно карельского наречия, то есть сформулированы флективные правила, согласно которым по заданной лемме и набору окончаний формируются все возможные основы. Затем по списку основ строятся словоформы. Вепсский язык отличается большим разнообразием диалектных форм и этому уделялось особое внимание.

Таким образом, вепсская и собственно карельская части корпуса ВепКар существенно выросли, пополнившись громадным количеством словоформ. Например, в 2019 году в словарь корпуса ВепКар было добавлено вручную более 5000 вепсских лемм, с ними автоматически были созданы системой ВепКар около 320 тысяч вепсских словоформ по флективным правилам. Для карельского языка автоматически (с помощью разработанного парсера словаря) было добавлено 3000 собственно карельских лемм, и соответственно около 130 тысяч собственно карельских словоформ.

Теперь редактор, введя одну строку текста, получает таблицу со 115-117 словоформами для глагола, и до 46 словоформ для именных частей речи. Для ливвиковского и людиковского наречия такая работа ещё предстоит, будут построены таблицы словоизменения для этих наречий с учетом диалектных особенностей.

Любые исследователи, заинтересованные в данных ВепКар, могут обратиться к разработчикам системы и получить данные корпуса в различных форматах. Например, тексты корпуса ВепКар в формате CONLL [Klyachko, 2019] участвовали в организации дорожек для соревнования по морфологическому анализу малоресурсных языков России. Словоформы с грамматической атрибуцией были-

предоставлены для международного проекта UniMorph (вепсский<sup>4</sup>, собственно карельское наречие<sup>5</sup>, ливвиковский<sup>6</sup> и людиковский<sup>7</sup>).

## Akal oli maks

### Корпус: подкорпус вепсских сказок

Северновепсский

Информант(ы): Арестова Клавдия Алексеевна, 1907, Матвеева Сельга (Matvejan selg),

Прионежский р-н, Республика Карелия

г. записи: 1957

записали: Лонин Рюрик Петрович

Источник: Вепсские народные сказки, (1996), с. 103-105

НА КарНЦ, кол.58, ед. хр. 68

Akal oli maks  
(вепсский)

Ende eletihe ukk akanke.

Lapsid hiil ii olnu.

Ukk käveskel' kaikušti mecha radole haugod  
čapmaha, a akk oli kodiš, emagjičihe kodiradol.

A mužikad ičeze ii g **kodi** (1) дом, строение,  
жилище, корпус, кров) +

**koda** (1) кабина) +

Oli maks. **koda** (2) киоск, конура,  
будка) +

Kut, nägettjk, mužik **koda** (3) курятник) +

hānen edes ozuteli **koda** (4) скворечник) +

Kerdan läks' mužik **kodi** (2) родной дом) +

Tuli maks besedale, **kodi** (2) родной дом) +

paniškat'he.

Неверная жена  
(русский)

Жили-были муж с женой.

Детей у них не было.

Старик всегда ходил в лес дрова рубить, а  
старуха была дома, занималась домашней  
работой.

Мужа своего она не любила.

Был у нее любовник.

Поскольку мужа своего она не любила, то  
всегда прикидывалась перед ним больной.

Однажды пошел старик в лес рубить дрова.

Пришел к ней любовник на беседу, и стали

Рис. 2. Пример размеченного текста на вепсском языке. Желтым цветом выделено параллельное предложение на русском языке. В вепсском тексте синим цветом выделены слова, для которых система нашла одно значение, красным – более одного значения, зеленым — слово проверено специалистом. По клику на красное слово kodiš открылась панель со всевозможными значениями леммы. Специалист должен выбрать правильное значение

### Людиковские тексты в ВепКар

В карельском языке выделяются 3 крупных наречия, два из которых (ливвиковское и людиковское) возникли под влиянием языка вепсов, что сделало языки еще более близкими. В нашей статье рассмотрим людиковский подкорпус, поскольку по сравнению с остальными он наименее проработан.

Выбирая за основу тот или иной говор, исследователь помимо прочего анализирует его представленность и частотность в размещаемых материалах. Абсолютно все говоры, представленные в классификации П. М. Зайкова [Зайков 2000, 27], отразить в корпусе невозможно, так как часть из них либо практически исчезла, либо находится на грани исчезновения. Речь идет, прежде всего, о людиковских говорах, носителей которых осталось, по неофициальным данным, не более 300 [Родионова 2017, 23].

Людиковское наречие в корпусе ВепКар представлено святозерским и михайловским диалектами. Святозерский диалект обладает наибольшим количеством “людиковских черт” по сравнению с михайловским диалектом, на который в большей степени оказал влияние вепсский язык.

<sup>4</sup> См. <https://github.com/unimorph/vep>

<sup>5</sup> См. <https://github.com/unimorph/kr1>

<sup>6</sup> См. <https://github.com/unimorph/olo>

<sup>7</sup> См. <https://github.com/unimorph/lud>

В корпусе сейчас 100 людиковских текстов. Для людиковского подкорпуса используются тексты из опубликованных образцов речи [ОЛР 1978; НКК 1994], подготовленных А. П. Баранцевым<sup>8</sup>. Все тексты представляют собой расшифровки магнитофонных записей и передают естественную речь информанта со всеми ее индивидуальными и общеязыковыми особенностями. Исследователь представил идиолект, то есть продемонстрировал особенности речи отдельного лица – уроженки деревни Пелдожа Святозерского сельсовета Пряжинского района Карельской АССР. Материал собирался в течение 10 лет: с 1963 по 1973 годы.

А. П. Баранцев планировал издать серию образцов людиковской речи, но, в итоге, был издан только один том. В этот том вошли материалы с сюжетами на три темы:

описание деревни Пелдожа и ее окрестностей,  
рассказы о жителях деревни Пелдожа,  
повествования о пережитых информантом событиях, свидетелем которых она являлась.

А. П. Баранцев включил в этот том более ста текстов. Они будут размещены в корпусе.

Кроме образцов святозерского диалекта А. П. Баранцев собирал материал у северных людиков в деревне Тивдия в 1980-х годах и в селе Михайловское в 1983 году, материал также был расшифрован и опубликован в первом томе сборника «Образцов карельской речи» [НКК 1994, 1]. Все тексты, опубликованные в образцах речи, обладают достаточно тонко представленной транскрипцией, что позволяют глубже прочувствовать фонетическую структуру людиковской речи, в чем А. П. Баранцев был и знатоком, и глубоким специалистом.

Тексты содержатся в параллельных подкорпусах с переводом на русский язык, что удобно для тех, кто не владеет карельским языком (в данном случае людиковским наречием). Диалектные тексты даны с подробной паспортизацией, нужной для работы диалектологов. В паспорте обычно указаны:

подкорпус, к которому отнесен текст,  
диалект,  
информанты (имя, год рождения, место рождения),  
год и место записи,  
научный сотрудник, записавший текст,  
опубликованный и/или архивный источник текста.

### *Будущие работы*

Планируется пополнять корпус текстами как на младописьменных формах карельского языка, это – поэтические и прозаические произведения, в том числе духовного содержания, статьи из периодической печати и пр., так и диалектными материалами, а именно изданными образцами речи и лично собранными полевыми материалами. Размещаемые тексты будут сопровождаться переводами на русский.

### **Семантическая категоризация**

В сопоставительно-ономасиологическом словаре [СОС 2007] предложена проработанная схема семантической категоризации диалектов карельского, вепского, саамского языков.

Мы планируем использовать эту схему для создания семантических категорий в словаре ВепКар. При редактировании словарных статей для каждого из значений слов можно будет указать привязку к какой-либо из 1500 категорий из сопоставительно-ономасиологического словаря [СОС 2007].

В первую очередь в словарь ВепКар будут добавлены слова из этого словаря [СОС 2007], поскольку для них категория уже известна. Затем будет решаться задача соотнесения других слов семантическим категориям. Неизбежно схема семантической категоризации будет расширяться и пополняться.

Наличие такой категоризации позволит увидеть список слов, имеющих близкое значение, относящихся к одной категории. Схема категорий представляет собой дерево.

Но задача редакторов словаря усложнится, поскольку им нужно будет хорошо ориентироваться в схеме категорий, чтобы присваивать значениям новых слов верные категории, или создавать новые категории, если их не хватает для покрытия значений слов.

---

<sup>8</sup> На сайте ИЯЛИ КарНЦ РАН в разделе «Имена и даты» размещена краткая информация о научных трудах А.П. Баранцева <http://illhportal.krc.karelia.ru/section.php?id=2663>

## Морфологический анализ

Задача морфологического синтеза заключается в том, чтобы по лемме и морфологическим признакам сгенерировать словоформу. И обратная задача — вычисление по словоформе морфологических признаков и леммы — это задача морфологического анализа.

С работой морфологических анализаторов, лемматизаторов и стеммеров связана задача моделирования недетерминированных конечных автоматов по регулярному выражению. Один из способов анализа и генерации словоформ основан на работе конечных преобразователей, то есть разновидности конечных автоматов. Конечные преобразователи, разработанные несколько десятилетий назад, применяются в решении задач частеречной разметки, фонологического анализа и морфологического анализа.

Мы планируем использовать программное обеспечение для построения морфологических анализаторов (библиотека OpenFst, Helsinki Finite-State Transducer toolkit, система SMOR) для кодирования морфологических преобразований для вепсского и карельского языков.

## Литература

Зайков П. М. Глагол в карельском языке (грамматические категории лица-числа, времени и наклона). Петрозаводск: Издательство Петрозаводского государственного университета, 2000. – 294 с.

[ОЛР, 1978] Баранцев А. П. Образцы людиковской речи. Петрозаводск: «Карелия», 1978. – 287 с.

[ОСВЯ, 2012] Орфографический словарь вепсского языка = Vepsän kelen orfografine vajehnik / Н. Г. Зайцева, Е. Е. Харитоновна, О. Ю. Жукова. Петрозаводск: Периодика, 2012. – 432 с.

Родионова А. П., Нагурная С. В., Чикина Н. В. Людики: вопросы сохранения языка и культуры. Исследования и материалы. Петрозаводск: Карельский научный центр РАН. 2017. – 169 с.

[СОС, 2007] Сопоставительно-ономазиологический словарь диалектов карельского, вепсского, саамского языков // Под общей редакцией Ю. С. Елисеева и Н. Г. Зайцевой. Петрозаводск, 2007. – 346 с.

[Klyachko, 2019] Klyachko E. L., Sorokin A. A., Krizhanovskaya N. B., Krizhanovsky A. A., Ryazanskaya G. M. LowResourceEval-2019: a shared task on morphological analysis for low-resource languages // Computational Linguistics and Intellectual Technologies: papers from the Annual conference “Dialogue” (Moscow, May 29 – June 1, 2019) Issue 18 (25). 2019. [Электронный ресурс]. URL: [http://www.dialog21.ru/media/4870/\\_dialog2019scopusvolplus.pdf#page=52](http://www.dialog21.ru/media/4870/_dialog2019scopusvolplus.pdf#page=52) (дата обращения: 12.09.2019).

[НКК, 1994] Näytteitä karjalan kielestä I (Образцы карельской речи). Joensuu-Петрозаводск, 1994. – 455 с.

## УЧАСТИЕ СТУДЕНТОВ В СОЗДАНИИ ЛИНГВИСТИЧЕСКОГО КОРПУСА КАК ОБРАЗОВАТЕЛЬНАЯ ТЕХНОЛОГИЯ

В статье представлена образовательная технология вузовского обучения, предполагающая постоянное, на всех ступенях обучения, участие студентов в крупном научном проекте. Описаны конкретные виды выполняемых студентами работ при создании мультимедийного диалектологического корпуса.

**Ключевые слова:** образовательные технологии, проектная деятельность, корпусная лингвистика.

Проектирование лингвистических ресурсов является важнейшей составляющей образовательного процесса в обучении студентов по направлению «Фундаментальная и прикладная лингвистика». Особенно эффективной становится эта образовательная технология, если работа по созданию одного крупного ресурса носит продолженный характер, включена в целый ряд учебных дисциплин. Хорошие возможности для такой организации учебного процесса дает участие студентов в создании лингвистического корпуса, требующего фундаментальных знаний и прикладных навыков, относящихся к разным областям лингвистики.

Студенты, обучающиеся по направлению бакалавриата «Фундаментальная и прикладная лингвистика» в Саратовском государственном университете имени Н. Г. Чернышевского, с самого начала обучения приобщаются к работе по созданию мультимедийного ресурса «Саратовский диалектологический корпус» – СарДК [о корпусе см., напр.: Крючкова, Гольдин 2012]. Эта работа ведется с 1-го по 4-й курсы на базе целого ряда дисциплин («Общая фонетика», «Общая морфология», «Общая семантика и лексикология», «Основы прикладной лингвистики», «Технологии обработки текста и звучащей речи», «Технологии корпусной лингвистики», «Общая и компьютерная лексикография», «Русская диалектология», «Социолингвистика», производственная практика «Разработка лингвистических ресурсов»), в которых корпусная тематика и корпусная работа получает свои акценты, углубленно отрабатываются разные аспекты создания корпуса.

В результате студенты теоретически и практически осваивают все этапы лингвистической работы по созданию корпуса – от его проектирования до подготовки готовых для включения в корпус модулей.

В начале работы над СарДК обсуждаются существующие значительно различающиеся концепции создания русского диалектного корпуса: [Летучий 2005, 2009; Качинская 2011; Крючкова, Гольдин 2011]. Цель такого обсуждения – осмыслить значимость этапа проектирования корпуса. Достигается понимание того, что условием дальнейшей успешной работы является определение целей создания корпуса, решение вопросов о содержании понятия «репрезентативность» в отношении данного корпуса, о способах достижения его репрезентативности.

Очерчивается круг вопросов, требующих решения при проектировании диалектного корпуса:

1. о принципах отбора диалектного материала и критериях репрезентативности диалектного корпуса;
2. о принципах выделения единицы хранения в корпусе;
3. о параметрах выдачи текстовых фрагментов;
4. о формах представления диалектных текстов в корпусе;
5. о видах и правилах аннотирования текстовой базы корпуса;
6. о параметрах метаразметки диалектных текстов;
7. о представлении в диалектном корпусе нелингвистической информации;
8. об оптимальных для диалектологических исследований поисковых возможностях корпуса [Крючкова 2018].

Определяется, что цели создания СарДК (а) как лингво-культурологического корпуса, (б) как модели диалектной коммуникации, (в) как полноценного научно-образовательного ресурса (используемого не только в ознакомительных, иллюстративных целях или для презентации ненормативных в сравнении с литературным языком явлений) обуславливают специфические параметры корпуса. Выясняется, что от других диалектных корпусов, создаваемых с позиций иных концептуальных решений, СарДК отличается и принципами отбора текстового материала, и характером и формами единиц хранения в корпусе, и параметрами единиц выдачи по поисковому запросу, и типами и содержанием разметки [см., напр.: Крючкова, Гольдин 2015; 2017].



Следующий этап обучения – практическая работа по подготовке текстовых материалов (работа ведется небольшими группами – по 2-3 человека). На этом этапе студенты получают разнообразные навыки подготовки материалов для корпуса. Большим учебно-дидактическим плюсом проекта является устный и мультимедийный тип создаваемого корпуса, поскольку разработка корпусов устной речи представляет собой более трудоемкий процесс в сравнении с письменными корпусами, ставит дополнительные вопросы перед разработчиками и требует дополнительных умений.

Так, первостепенную значимость при создании корпусов устной речи приобретает вопрос о способах представления в корпусе речевого материала, характере его символической передачи: в виде орфографизированной записи, следующей общему стандарту письма, или в виде транскрипции, которая может иметь различные степени глубины (подробности). То или иное решение должно быть мотивировано задачами корпуса, особенностями его эмпирической базы [см.: Крючкова, Гольдин 2015]. На материале небольшого текстового фрагмента студентами апробируются разные типы расшифровки звучащей речи, разбираются особенности принятой в СарДК символической. После практического анализа проблем символической передачи звучащей речи и обоснования избранной в СарДК стратегии проводится расшифровка аудиозаписей диалектной речи в соответствии с разработанными для корпуса правилами (инструкцией). Результаты сделанных студентами расшифровок, обоснованность принятых в трудных случаях решений обсуждаются на занятиях.

В процессе расшифровки записей диалектной речи проводится также лингвистическая работа по фонетическому и словарному комментированию текстов.

На следующем этапе осуществляется корпусная обработка текстовых и звуковых модулей. В установленной последовательности студентами выполняются следующие работы:

1. Структурная сегментация текста: деление текста на абзацы (единицы выдачи по поисковому запросу), выделение и разметка неоднословных единиц.

Выделение неоднословных единиц в текстовом корпусе и организация соответствующего поиска являются, как известно, трудным местом корпусной лингвистики. Процедура выделения и разметки единиц данного типа проводится с опорой на инструкцию, характеризующую виды тегов, используемых в зависимости от типов неоднословных единиц (фразеологизмы, коллокации, раздельнооформленные словоформы, раздельнооформленные структурные единицы (предлоги, союзы, частицы и др.), составные имена собственные, тавтологические выражения типа *жизнью жить, старые старики*, аналитические выражения, семантически эквивалентные одному слову, типа *насмешку давать = насмехаться, крест делать = креститься*) и контактного или дистантного расположения их компонентов.

2. Жанрово-тематическая разметка текста. Опирается на подробную инструкцию, описывающую теги данного вида разметки и варианты их комбинирования в зависимости от позиции (начало, конец размечаемого фрагмента) и особенностей содержания текста (возможности пересечения тем и жанров в одном текстовом фрагменте). Инструкция также содержит перечень тем и жанров, используемых в жанрово-тематической разметке корпуса.

3. Метаразметка текста. Метаразметка включает сведения об информантах, о времени и месте записи, описание ситуации общения, отраженной аудио- или видеоматериалом (в доме, на улице и т.д., занятие информанта во время общения: вяжет, работает в огороде и т.д.), сведения об адресатах речи, характеристику (идентификацию) упоминаемых конкретных лиц (например, родители диктора, родственники, односельчане), обозначение времени описываемых событий (до революции, революция, гражданская война и т.д.), перечень выделенных в тексте тем и жанров.

4. Составление биографического текста. На основе обрабатываемой расшифровки записи речи диалектоносителя составляется биографический текст. Если такой текст был составлен ранее на основе других записей от того же информанта, то он может быть дополнен новыми сведениями.

5. Нарезка аудио/видеозаписи. Осуществляется автоматически с целью синхронизации символических и звуковых фрагментов выдачи. Звуковая запись делится на фрагменты, соответствующей единице текстовой выдачи – абзацу.

6. Морфологическая разметка диалектного текста осуществляется автоматически программой DIALING (разработана группой А. В. Сокирко).

7. Редактирование морфологической разметки. Проводится вручную, поскольку, во-первых, не существует автоматических анализаторов для разметки диалектных текстов и автоматическая разметка дает значительное число некорректных разборов. Во-вторых, задачи корпуса требуют введения в морфологическую разметку ряда дополнительных зон и помет. Например, вводится зона литературного соответствия ввиду возможного расхождения диалектной и литературной начальной формы (ср.: *церква* и *церковь*, *Паска* и *Пасха*, *пенёк* и *улей*); для поиска диалектных форм вводится помета *nstand*,

выделенная в особую зону разметки; используется отсутствующая в автоматическом анализаторе специальная система тегов для разметки имен собственных; ручного редактирования требует разметка разных видов неоднословных единиц.

8. Автоматическая проверка размеченного текста. Процедура проводится с помощью специально разработанной для СарДК программы: проверяется парность скобок, которая могла быть нарушена в процессе ручного редактирования морфологической разметки; осуществляется конвертирование размеченного текста в неразмеченный, что позволяет получить полностью соответствующие друг другу версии текста – размеченного и неразмеченного.

Завершается практическая работа по подготовке материалов для корпуса формированием папки текста-модуля. Состав папки:

- 1) звуковой файл с нарезкой;
- 2) файл с текстовой расшифровкой во вспомогательной транскрипции;
- 3) файл с морфологически размеченным текстом;
- 4) файл с метаинформацией;
- 5) файл с биографией диктора;
- 6) файл с фотографиями диктора;
- 7) файл с фонетическим комментарием к тексту;
- 8) файл, представляющий словарное описание диалектных слов и устойчивых сочетаний.

Таким образом, участие студентов проекте создания диалектного корпуса служит основой получения хороших теоретических знаний о разноуровневых языковых явлениях и практических умений и навыков в области современных корпусных технологий.

### **Литература**

Качинская И. Б. Диалектный подкорпус НКРЯ. Новый стандарт подачи. Новое рабочее место // Русская устная речь: Материалы Международной научной конференции (Саратов, 15–17 ноября 2010 г.). – Саратов: ИЦ «Наука». – С. 240–249.

Крючкова О. Ю. Задачи информационно-поисковой системы диалектного корпуса и лингвистическое обеспечение поискового механизма // Компьютерные науки и информационные технологии: Материалы Междунар. науч. конф. – Саратов: Издат. центр «Наука», 2018. – С. 214–215.

Крючкова О. Ю., Гольдин В. Е. Корпус русской диалектной речи: концепция и параметры оценки // Компьютерная лингвистика и интеллектуальные технологии. Вып. 10 (17). – М.: РГГУ, 2011. – С. 359–367.

Крючкова О. Ю., Гольдин В. Е. Русская диалектная речь в электронных информационных системах // Языки в современном мире: Материалы X Международной конференции. – М.: КДУ, 2012. – С. 24–31.

Крючкова О. Ю., Гольдин В. Е. Параметры обработки текстов для русского диалектного корпуса // Труды Международной конференции «Корпусная лингвистика–2015». – СПб.: С.-Петербургский гос. университет, Филологический факультет, 2015. – С. 307–314.

Крючкова О. Ю., Гольдин В. Е. Диалектный текстовый корпус: проблемы репрезентативности, сбалансированности, единиц хранения и выдачи // Труды международной конференции «Корпусная лингвистика-2017». – СПб.: Издательство С.-Петерб. ун-та, 2017. – С. 231–235.

Летучий А. Б. Корпус диалектных текстов: задачи и проблемы // Национальный корпус русского языка: 2003–2005. – М.: Индрик, 2005. – С. 215–232.

Летучий А. Б. Диалектный корпус: состав и особенности разметки // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. – СПб.: Нестор-История, 2009. – С. 114–128.

## **ОПЫТ СОЗДАНИЯ НАЦИОНАЛЬНОГО КОРПУСА АВАРСКОГО ЯЗЫКА**

Целью создания национального корпуса аварского языка является создание информационно-справочной системы, в которой будет собрано письменное и устное наследие аварского народа. С появлением такого электронного ресурса исследователи впервые получают возможность работать с большими объемами текстового материала без необходимости тратить время на поиски в книгах.

Национальный корпус аварского языка предназначен для всех, кому интересен аварский язык – для профессиональных лингвистов, специалистов по аварскому языку, преподавателей, составителей словарей и грамматик.

**Ключевые слова:** аварский язык, национальный корпус, информационно-справочная система, фольклор, словарь.

В настоящее время многие языки малочисленных народов Российской Федерации находятся под угрозой исчезновения. В их числе также многие языки народов Дагестана. Как известно, в Дагестане 32 языка, из которых 14 письменных и 18 бесписьменных. Современные информационные технологии предоставляют новые возможности для сохранения исчезающих языков.

Национальный корпус аварского языка представляется одним из важных проектов, который сможет внести свою лепту для сохранения языка в условиях постоянного уменьшения количества его активных носителей, под которыми понимаются те, кто не только понимает речь, но и осуществляет коммуникацию на том или ином языке.

Этот проект позволит осуществить лингвистические исследования с использованием корпусного подхода, что не только уточнит и дополнит имеющиеся описания языка, но и осветит те проблемы, которые еще не изучены в аварском языкознании. Информационно-справочная система должна выступить как материал для исследования лексики и грамматики аварского языка.

Информационно-справочная система позволит аккумулировать письменное и устное наследие аварцев.

Возникает вопрос: что такое корпус? Под корпусом понимается информационно-справочная система, которая содержит сведения о тексте и в то же время информацию о языковом знаке этих самих текстов. Все тексты делятся на части и слова. Практически каждая единица получает свое языковое описание.

Корпус языка – это собрание текстов на этом языке, в котором текстам приписана дополнительная лингвистически релевантная информация (аннотация) и которое снабжено поисковым механизмом, позволяющим производить поиск по этой информации. Аннотация может включать в себя любую информацию, в зависимости от задач, стоящих перед создателями корпуса и исследователями. При этом корпус, в отличие от электронной библиотеки, предназначен в первую очередь не для просмотра полных текстов имеющихся в нём произведений.

Следует отметить, что работа по разработке корпуса проводилась по двум направлениям. На первом этапе необходимо собрать тексты, включаемые в корпус. Источниками текстов послужили издательства, отсканированные книги и расшифровки устной речи.

На сегодняшний день оцифровано более 2 тысяч источников, включая художественные произведения, включая стихотворения и прозаические тексты, статьи в средствах массовой информации, фольклорные тексты, научно-популярные статьи, переводные тексты, учебники и словари.

Фольклор во всем своем многообразии жанров являет собой яркий образец метафоричности языка и содержит элементы архаики, по этой причине, представляется важным включение фольклорных произведений в Национальный корпус аварского языка.

Поэтические тексты, наравне с прозаическими и фольклорными, являются весьма важными источниками изучения аварского языка.

Были использованы также ранее оцифрованные тексты. Кроме того, тексты на аварском языке были получены от издательств и редакций аварских газет и журналов.

Не все стили представлены равномерно: основной массив текстов – это художественные произведения, затем идут фольклорные тексты, которые относятся к художественному стилю. Что касается текстов других стилей (научный и официально-деловой), то их не так много. Тексты делового и научного стиля составляют несущественное количество.

К сожалению, в Дагестане не сложилась традиция создания деловых документов или научной продукции на дагестанских языках. Существует лишь небольшое количество научных статей, написанных на дагестанских языках: ученые предпочитают писать на русском языке, чтобы быть доступнее более широкому кругу читателей [Магомедов 2016, 158].

На втором этапе необходимо внести в собранные тексты грамматическую информацию. Поэтому для разметки текстов необходимо разработать систему автоматического морфологического анализа. Следующим этапом является создание или адаптация существующей поисковой платформы и вывеса корпуса в Интернете. Разработка системы морфологического анализа является одним из центральных этапов создания корпуса. Задача создания такой системы для аварского языка состоит из нескольких частей. Первым шагом является описание морфологической системы языка, т. е. описание имеющихся в нём частей речи, грамматических значений и способов их выражения, описание классов лексем с одинаковым словоизменением (напр., склонений существительных) и т. п. Вторым шагом является создание грамматического словаря, включающего в себя словоизменительные таблицы и перечень лексем с указанием их словоизменительного типа, основы и всей остальной информации, необходимой для построения полной парадигмы. Третьим шагом является создание программы, способной размечать текст, используя данные грамматического словаря [Магомедов 2012, 205].

Далее разрабатывается программа по сегментированию текста на части и составляется грамматическая характеристика к слову. Например, берем слово и четко определяем: к какой части речи оно относится, в каком падеже, числе стоит и т. д. Среди информации, которой могут снабжаться тексты корпуса, следует в первую очередь выделить пословную разметку, т. е. дополнительную информацию, которая указывается при каждой словоформе текста, содержит прежде всего грамматическую информацию, а также информацию о тех или иных семантических признаках словоформы и др. Грамматическая информация может включать в себя начальную форму, часть речи, словоизменительные грамматические признаки и т. д. Важная часть корпуса – создание системы морфологического анализа.

Электронный ресурс создается в целях сохранения и развития аварского языка. И ориентирован, в первую очередь, на молодежь. Он будет полезен и станет большим подспорьем в работе лингвистов, филологов, преподавателей и учителей аварского языка, школьников и студентов, словом, для всех, кто занимается исследованием аварского языка.

Мы планируем оцифровать все номера журнала «Гьудулльи», начиная с 1-го номера и до сегодняшнего дня. Это материал даст нам возможность широкого привлечения аварского текстового материала.

Проблема сочетаемости лексических единиц заключается в выявлении синтагматических свойств слова, что значительно важно, наряду с парадигматическим описанием лексики.

В проекте предусматривает изучить наиболее частотные употребления слов и выражений.

В результате реализации проекта будет возможным выявить зафиксированное в языке восприятие национальной картины мира, которая для каждого народа уникальна, поскольку отражает индивидуальный жизненный опыт, традиционные устои и способы познания определенного языкового коллектива.

В проекте предлагается исследование сочетаемостных свойств слова в аварском языкознании. Результаты работы могут быть использованы в спецкурсах и спецсеминарах по лексикологии, представляет определенную практическую ценность для преподавания аварского языка в школах и вузе, в лексикографической практике (семантическом конструировании различных типов словарей).

Корпус аварского языка является компьютерным инструментом исследования языка, поэтому будет иметь очень важное значение при составлении различных видов словарей. Особое внимание уделяется вопросам составления словарей и автоматической грамматической разметки текстов. Работа корпуса очень эффективна для всех видов лингвистического анализа текста.

Кроме того, любому лингвисту при написании научной работы необходимы иллюстративные материалы по исследуемому языку. В этом плане национальный корпус аварского языка будет востребован специалистами для поиска примеров к разрабатываемой проблеме. Традиционный способ получения языковых примеров – чтение книг – имеет ряд очевидных недостатков, например, большая трата времени на поиск какого-то одного примера, тогда как при помощи корпуса моментально можно получить необходимый иллюстративный материал. Таким образом, основная задача корпуса – предоставить исследователю языка возможность быстро получать реальные языковые примеры по заданному запросу.

Возможен также поиск по переводу словоформы на русский язык, что удобно для исследователей, не владеющих аварским языком. В настоящее время для аварского языка параллельные тексты на русском и аварском языках имеют большое значение в аспекте сохранения аварского языка. В про-

цессе изучения аварского языка школьники, студенты и просто желающие изучать язык опираются прежде всего на знания первичного языка – русского, которым они владеют хорошо.

Параллельные тексты позволят исследователям сопоставить два разнотипных языка и вывести соответствия на лексико-грамматическом уровне.

Одной из главных целей дальнейшей работы над корпусами является увеличение текстовой базы и расширение грамматического словаря. При сборе новых текстов важен не только их объём, но и тип. Чтобы корпус был репрезентативным, т. е. действительно мог представлять язык во всех его вариантах, в нём должны быть представлены тексты разных стилей и жанров: художественные и нехудожественные; переводные и непереводные; поэзия и проза; пресса, официальные документы, тексты из интернет-форумов и т. д., причём желательно, чтобы ни один из типов текстов не имел бы существенного количественного преобладания над остальными. Следует несомненно стремиться разнообразить состав корпуса. Расширение грамматических словарей позволит повысить качество грамматической разметки.

Ещё одним возможным расширением функциональности является размещение на страницах корпусов статистической информации, такой как частотные списки словоформ и лексем.

Таким образом, национальный корпус аварского языка – это фундаментальный проект, который направлен на фиксацию и сохранение аварского языка и создание условий для его возрождения и развития как полноценного средства коммуникации.

Проект адресован очень большому кругу пользователей. Например, если студента или школьника интересует, как то или иное слово переводится на аварский язык, он может воспользоваться поиском по переводу. Для исследователя-авароведа интерес представляет вся совокупность разработанной разметки как объект лексико-грамматического исследования, а сами тексты как материал исследования с разных точек зрения – в диахронном или синхронном аспекте. Для преподавателей аварского языка материал корпуса можно использовать для разработки упражнений для учеников и студентов в средних и высших учебных заведениях. Преподаватели аварского языка, разрабатывающие основы преподавания языка, могут найти в созданном ресурсе информацию и сочетаемости слова в лексическом и грамматическом аспекте.

Корпус предназначен не только для лингвистов, но и для историков и этнографов: тексты содержат уникальную информацию по разным вопросам этнографии и истории, поскольку художественные произведения отражают окружающую реальность, а значит и описания событий, обрядов, традиций и т. п.

Сайт будет находиться в открытом доступе, и каждый желающий сможет свободно использовать его материалы.

### Литература

Магомедов Д. М. Сохранение языковых и этнокультурных ценностей народов Дагестана как средство сохранения национального самосознания // Ингушетия в контексте научных проблем и перспектив изучения Кавказа (к 90-летию научно-исследовательского института). – Магас, 2016. – С. 156–159.

Магомедов М. И. Создание и развитие корпусных ресурсов по литературным языкам народов Дагестана // Известия Кабардино-Балкарского научного центра РАН. – Нальчик, 2012. № 3. – С. 200–207.

## ПОДКОРПУС ДИАЛЕКТНОЙ ЛЕКСИКИ ЯКУТСКОГО ЯЗЫКА В УСЛОВИЯХ ЦИФРОВИЗАЦИИ (ПОСТАНОВКА ПРОБЛЕМЫ)

В данной статье автор ставит проблему создания диалектного подкорпуса якутского языка. Первые опыты показали, что для сбора качественных материалов и их лингвистической обработки, записи устной речи носителей языка требуют детальной подготовки. Атмосфера записи, стиль речи собеседника, техника записи, морально-этические качества и интеллектуальный уровень диалектолога могут сыграть значительную роль в создании мультимедийной базы данных.

**Ключевые слова:** якутский язык, национальный корпус языка, диалектолог, диалектный подкорпус.

Диалектная лексика якутского языка является одним из существенных показателей изменений, происходящих в современном якутском языке. В условиях развития полилингвизма, введения новых стандартов в сфере образования и науки, фонетических изменений в устной речи носителей языка, создание Национального Корпуса якутского языка (далее – НКЯЯ) может стать универсальным фондом (базисом) для научных исследований лексики и грамматики языка, а также различных языковых изменений.

Прообразом для создания национальных корпусов в России, по сути, является Национальный корпус русского языка [URL: <http://www.ruscorpora.ru/corpora-stat.html>]. Национальный корпус представляет данный язык на определенном этапе (или этапах) его существования и во всем многообразии жанров, стилей, территориальных и социальных вариантов и т.п. Главной целью создания НКЯЯ является «обеспечение академических исследований языка» [Плунгян, 2005].

НКЯЯ может включать следующие подкорпусы:

1. Газетно-журнальный корпус (корпус современных СМИ), в котором представлены статьи, фразы, цитаты, выдержки из СМИ, рекламные тексты;
2. Корпус специальных текстов на якутском языке (научные статьи, монографии и т.д.);
3. Корпус диалектной лексики, включающий тексты и записи (аудио, видео), в которых можно выявить диалектную лексику говоров якутского языка;
4. Мультимедийный корпус, включающий записи частной устной речи (бытовой разговор), записи публичной устной речи (записи радиопередач, «круглых столов», бесед на определенные темы и т.д.).

Также НКЯЯ может иметь подкорпус материалов частной переписки, дневников, так называемый *эпистолярный подкорпус*.

Среди данных корпусов существенным в сохранении живой речи является сбор материала для *мультимедийного корпуса*, который можно совместить с диалектным подкорпусом. Живая речь носителей языка, сохраненная в настоящее время, является ценным материалом не только для науки, но и для будущих поколений. Нормативно-литературная речь преимущественно развивается в письменном виде, а характерной чертой живой речи являются диалектизмы. Диалектные особенности можно выявить не только в качестве отдельных лексем, но и в фонетических отличиях устной речи. Несмотря на укрепление позиций литературного якутского языка, диалектная речь ярко выражена в отдаленных улусах среди старшего и среднего поколения (например, в Сунтарском, Среднеколымском, Верхоянском улусах).

Первые опыты аудиозаписи показали, что в будущем необходимо учитывать следующие моменты:

1. **Стиль речи.** Как и предполагалось, диалектная лексика сохраняется в основном в речи людей пожилого возраста (старше 60 лет). К сожалению, устная речь представителей более младших поколений, пусть даже пенсионного возраста выхолощена единообразием стиля не только общения, но и художественной литературы. Люди в основном излагают мысли на том языке, в том стиле, в среде которой они выросли. В результате, даже многочасовые записи могут иметь нулевой результат.

2. **Атмосфера записи.** Запись с приготовлением, подготовкой и готовыми вопросами, введением в тему, «разогревом» имеет свои существенные минусы. Во-первых, собеседник «закрывается», старается говорить правильно, как многие сами объясняют – «по-литературному». Во-вторых, само присутствие техники записи (видеокамера, диктофон, телефон) затормаживает даже самого хорошего собеседника. Таким образом, диалектологу приходится чаще скрывать, что он ведет запись разговора.

При этом собеседник знает, что с ним разговаривает ученый, который занимается якутским языком. Тут возникает этический вопрос. Если запись без предупреждения подразумевает дальнейшую обработку и публикацию, то необходимо заранее согласовать, что будет осуществляться запись. Да, можно собирать, записывать все подряд, всех пожилых людей. Но ограниченное время и языковые процессы требуют собранности и более качественных записей, от которых больше КПД.

Значительным фактором для перспективных исследований является хранение и использование данных из записей. Некоторые факты из жизни села, односельчан могут быть сугубо личными, которые нельзя распространять, тиражировать. В таких случаях единственным ограничителем является совесть и морально-этические принципы исследователя.

**3. Уровень подготовки диалектолога.** Тематика беседы с людьми старшего поколения требует знаний о сельской жизни. Есть темы, без которых невозможно разговаривать собеседника: сенокос, охота, местная флора и фауна, погода и ее приметы, история села, улуса, заготовка дров, национальные традиции, национальная кухня, одежда, утварь, украшения, орудия труда, местная политика и т.д. В лучшем случае, изучить тему, о которой возможно будет идти речь. Знание родного языка может играть ключевую роль в диалоге со старшим поколением. Также есть темы, которые относятся к табу в том или ином населенном пункте, например, тема прозвищ односельчан; тема, касающаяся трагедии местного масштаба (о сгоревшей школе) и т.п.

Для осуществления поставленной цели необходимо выполнение следующих задач: обеспечение специальной техникой (диктофоны, видеокамеры, фотокамеры и т.п.); организация краткосрочных диалектологических командировок (экспедиций); составление соглашений (договоров) между авторами, издательствами, редакциями и научным учреждением в целях создания базы данных художественных, публицистических текстов и их изучения и т.д.

Существенной базой данных для диалектного подкорпуса якутского языка может стать газетный подкорпус. Улусные (районные) газеты никогда не входили в поле изучения в контексте диалектологии. Начальный этап изучения газет Томпонского (относится к акающей группе говоров) и Сунтарского (относится к окающей группе говоров) улусов выявил около десяти новых диалектизмов, которые, в виду их узколокальности могли быть незафиксированными в словарях. Письменный текст сильно отличается от устного тем, что ошибок может быть больше. Например, правописание на основе современного орфографического словаря может быть не совсем правильным, но имеющим все признаки диалектизма. Также сложность газетных текстов заключается в том, что стали мало публиковать художественные произведения, воспоминания ветеранов. Отмечается отсутствие писем читателей. Новостной, калькированный характер подачи материалов нивелировало практически все диалектные отличия.

Создание НКЯЯ является новым подходом в решении насущных языковых проблем. Связующим звеном в комплексе работ по сохранению якутского языка является носитель языка, которому присущ творческий подход, энтузиазм, компетентность. В условиях цифровизации общества создание НКЯЯ является требованием времени и приоритетной задачей государственной политики ведущих структур (Совета по языковой политике РС (Я), Правительства РС (Я) и др.) в сфере развития национальных языков.

Тем не менее, несмотря на актуальность, корпус языка является лишь инструментом в изучении и развитии родного языка. Он не должен стать самоцелью даже для академических изысканий. Национальный корпус в нашем понимании это только аргумент или каркас для различных исследований. Главная цель корпусных работ – сбор и сохранение образцов живой речи носителей языка, которая станет ценным материалом для различных исследований.

#### **Литература:**

Национальный корпус русского языка. [Электронный ресурс]. URL: <http://www.ruscorpora.ru/corpora-stat.html> (дата обращения: 12.05.2019).

Плунгян В. А. Зачем мы делаем национальный корпус русского языка? // Отечественные записки. – 2005. – № 2. – С. 296–308.

## **НОВЫЕ ТЕКСТОВЫЕ КОРПУСЫ МИНОРИТАРНЫХ ЯЗЫКОВ НА СЕРВЕРЕ KORP.CSC.FI.**

Сервер Korp.csc.fi в Финляндии предоставляет текстовые корпуса нескольких разновидностей для многих языков – больших и малых. Инфраструктура Корп разработана шведской компанией Språkbanken в университете и Гетеборге, исходный код выпущен под лицензией MIT. Открытый характер системы позволяет легко переноситься в новую среду, и уже существуют многочисленные Корп установки. Тот, который мы обсуждаем, поддерживается языковым банком Финляндии.

**Ключевые слова:** корпус уральских языков меньшинств, минимальные данные для Корпа, корпус языков с ограниченными ресурсами.

*J. M. Rueter, N. Partanen  
Helsinki, Finland*

## **ON NEW TEXT CORPORA FOR MINORITY LANGUAGES ON THE HELSINKI KORP.CSC.FI SERVER**

The korp.csc.fi server in Finland provides text corpora of multiple varieties for numerous languages large and small. The Korp infrastructure is developed by the Swedish Språkbanken in the University and Gothenburg, and the source code is released under MIT license. Open nature of the systems makes it easily transferred into new environments, and there are already numerous Korp installations available. The one we discuss is maintained by the Language Bank of Finland.

**Keywords:** minority-language Uralic corpora, minimal data for Korp, low-resourced language corpora.

The korp.csc.fi server in Finland provides text corpora of multiple varieties for numerous languages large and small. The Korp infrastructure [Ahlberg et. al. 2013] is developed by the Swedish Språkbanken in the University and Gothenburg, and the source code is released under MIT license. Open nature of the systems makes it easily transferred into new environments, and there are already numerous Korp installations available. The one we discuss is maintained by the Language Bank of Finland.

In addition to the official languages of Finland, Finnish and Swedish, there is an “other languages” section featuring materials from Hurro-Urartian, Indo-European, Niger-Congo, Semitic, and Uralic language families. While the Indo-European and Niger-Congo materials are listed by language and focus on large modern languages (e.g. English, French, German, Russian, Spanish and Swahili), the Oracc collection features Hurro-Urartian, Semitic and other Mesopotamian materials, and the Uralic collections include materials from non-majority languages. It is the Uralic materials that will be described more in detail here.

The Uralic materials are divided according to licensed group: ERME, Fenno-Ugrica, SUS-kenttätyö, Wanca [Jauhiainen 2015], Kildin Saami, BeserCorp. The ERME, Fenno-Ugrica, Wanca and SUS-kenttätyö collections contain multilingual materials; Erme – mainly Erzya and Moksha literary materials, Fenno-Ugrica (National Library of Finland) represents pedagogical and literary materials from the 1920s through the early 1950s for ten Uralic languages, Wanca (Kone, FINCLARIN), in contrast, represents internet materials for 29 Uralic languages, and the SUS-kenttätyö (Finno-Ugrian Society Fieldwork) collection is just the beginning of Uralic fieldwork materials to be made available from the Pre-Soviet Era (normalized text search with fieldwork transcriptions and research-language translation, i.e. German, Russian, or other). The Kildin Saami sample corpus provides original unannotated orthographic text as are currently found in the Finno-Ugrica and Wanca collections. And finally, the BeserCorp provides a mixed encoding representation of Beserman Udmurt texts, where obvious Russian loanwords are written in Cyrillic script and other words are written in IPA-like Latin transcription with morphological segmentation and analyzed glossing.

Making materials available on a korp server, with concordance and more specific search options, presupposes some rudimentary language resources. First, there is the text, which may be any variety of orthographic, phonetic or other UTF-8 supported encoding. Second, the texts should be broken into sentence-like segments that eventually might be parsed syntactically, or annotated at other levels. Finally, each sentence is given a unique identification and tokenized with 1,2,3... unique identification of each individual token for each sentence (see figures 1–2). These are the minimal requirements, and any additional information including lemmas, part-of-speech, morpho-syntactic analysis, glosses and sentence translation are enhancement.



Figure 1. Illustrates the beginning “<sentence>” tag with attributes and values for unique identification as well as the original text and German translation. The attribute `paragID` for paragraph identifier along with the `sent` attribute are sufficient data for unique identification when accompanied by the text identifier. Superfluous data might be the `pgNo` (page number) and `pgLi` (page line) attributes, which align with reference provided in the printed texts, where the original transcriptions and translations are aligned at the line level.

```
<sentence paragID="1" sent="1" pgNo="0008" pgLi="1" orig_string="rutš
šue turili: «vetlan, voe, me ordę geštitiŋ!»" deu="Der Fuchs sagt zum
Kranich: «Laß uns, mein Brüderchēn, zu mir zu Besuch gehen!»">
```

Fig. 1.

Fig. 2. Provides an XML structure where the original-string tokens from fieldwork transcriptions have been copied to a line by line set of “<w/>” elements with “sID” string identification attributes.

```
<sentence paragID="1" sent="1" pgNo="0008" pgLi="1" orig_string="rutš
šue turili: «vetlan, voe, me ordę geštitiŋ!»" deu="Der Fuchs sagt zum
Kranich: «Laß uns, mein Brüderchēn, zu mir zu Besuch gehen!»">
<w word="руч" lemma="" pos="" msd="" sID="1" orig_string="rutš"/>
<w word="шүө" lemma="" pos="" msd="" sID="2" orig_string="šue"/>
<w word="турилы" lemma="" pos="" msd="" sID="3" orig_string="turili"/>
>
<w word=":" lemma=":" pos="CLB" msd="CLB" sID="4" orig_string=":"/>
<w word="«" lemma="«" pos="CLB" msd="CLB" sID="5" orig_string="«"/>
<w word="ветлан" lemma="" pos="" msd="" sID="6" orig_string="vetlan"/>
>
<w word="," lemma="," pos="CLB" msd="CLB" sID="7" orig_string=","/>
<w word="воө" lemma="" pos="" msd="" sID="8" orig_string="voe"/>
<w word="," lemma="," pos="CLB" msd="CLB" sID="9" orig_string=","/>
<w word="ме" lemma="" pos="" msd="" sID="10" orig_string="me"/>
<w word="ордө" lemma="" pos="" msd="" sID="11" orig_string="ordę"/>
<w word="гөсьитиŋы" lemma="" pos="" msd="" sID="12"
orig_string="geštitiŋ"/>
<w word="!" lemma="!" pos="CLB" msd="CLB" sID="13" orig_string="!"/>
<w word="»" lemma="»" pos="CLB" msd="CLB" sID="14" orig_string="»"/>
</sentence>
```

Fig. 2.

Each token has then been copied to both a “word” and “orig\_string” attribute, where Komi-Zyrian Cyrillic normalization has been applied to the “word” attribute value and the fieldwork transcription value has been retained in the `orig_string` attribute. Initial work has been carried out for the Mordvin languages Erzya and Moksha by these authors using simple lines of perl script for ordered normalization, but continued work will move towards practices used in early Modern English Normalization [Hämäläinen 2019].

Figure 3. illustrates the VRT representation of this minimal tokenization and string identifier data. The first line of the individual columns, in red, is merely an indicator of the data type. Only punctuation symbols have been lemmatized with part-of-speech tagging and morphological analyses. The absence of data value is indicated by a stroke below symbol “\_”.

Wordform	Lemma	PoS	Msd	StringId	Transcription
руч	—	—	—	1	rutš
шүө	—	—	—	2	šue
турилы	—	—	—	3	turili
:	:	CLB	CLB	4	:
«	«	CLB	CLB	5	«
ветлан	—	—	—	6	vetlan
,	,	CLB	CLB	7	,
..	..	—	—	~	..

Fig. 3.

If the language has a little more resources, which is the case for Komi-Zyrian, lemmas, parts-of-speech and morphological analyses may also be had with minimal manual work (see Fig. 4.). Even with current language technology that is available, however, this manual correction does not scale well into larger corpora, for which the further development of the available tools seems to be the best approach. That said, even a small manually annotated corpora can be very valuable in many tasks, and such work is always very recommendable. Pipelines described in connection to different treebanks [Partanen 2018, Rueter 2018], which combine finite state transducers into manual disambiguation, can be easily adapted in this context as well.

Wordform	Lemma	PoS	Msd	StringId	Transcription
руч	руч	NOUN	N.Sg.Nom	1	rutš
шуö	шуны	VERB	V.Ind.Prs.Sg3	2	šue
турилы	тури	NOUN	N.Sg.Dat	3	turiḷi
:	:	PUNCT	CLB	4	:
«	«	PUNCT	CLB	5	«
ветлан	ветлыны	VERB	V.Ind.Fut.Sg2	6	vetlan
,	,	PUNCT	CLB	7	,

Fig. 4.

As in Fig. 3. before it, the first row of the columns in Fig. 4. is merely intended to illustrate the data type. The order of the columns, it should be noted, is documented in each project separately, i.e. there is no specific ordering to the individual columns. Hence a similar project from Moksha might be illustrated with column ordering the same as in the Universal Dependencies forth-coming release [15 Nov. 2019; cf. Nivre, et al, 2019] (see Fig. 5.). Here the ordering applied is by column as follows: 1= string id, 2 = word form or token, 3 = lemma, 4 = part-of-speech, 5 = alternate encoding, 6 = features, 7 = dependency, 8 = dependency relation, 9 = NA, and 10 = miscellaneous.

```
<sentence sent_id="MishaninaValentina_LiendenyOchkonyasa_Moksha-1972-No2-pp38-39:36" text="Лётчикне, улема, кядьса токседазь коволнятнень.">
1 Лётчикне лётчик NOUN N Case=Nom|Definite=Def|Number=Plur 6 nsubj _
  SpaceAfter=No
2 , , PUNCT CLB _ 3 punct _ _
3 улема улема PART Pcle _ 6 advmod _ SpaceAfter=No
4 , , PUNCT CLB _ 3 punct _ _
5 кядьса кядь NOUN N Case=Ine|Definite=Ind|Number=Plur,Sing 6 obl _
  _
6 токседазь токсемс VERB V Mood=Ind|Number[obj]=Plur|Number[subj]=Plur|Person[obj]=3|
  Person[subj]=3|Tense=Pres|Valency=2 0 root _ _
7 коволнятнень коволня NOUN N Case=Genl|Definite=Def|Number=Plur 6 obi
```

Fig. 5.

This kind of additional information can be added at the collection, text, paragraph, sentence or token level, and all of these attributes can be queried in the search interface.

Korp uses VRT XML format, in which each tokenized sentence is represented in a tabular structure within XML's text element. This tabular structure is split so that each line represents a wordform, and the annotations about that wordform are gathered at different columns at the same row (cf. fig 3–4.). In principle the number of columns is arbitrary, and any information could be added this way.

The fields and tags in Korp can be translated within a configuration file, and this provides many possibilities for a multilingual interface. At the moment the whole interface is available in Finnish, Swedish and English. There are three search types: Simple, Extended and Advanced. The Advanced search mode uses CQP, which is a widely used and very expressive query language. Besides search form itself, Korp also has an API, which allows all the same search actions and the ordinary interface. Korp API can also be accessed through a Python package [Hämäläinen 2018].

It is relatively easy to add new languages and corpora to this infrastructure, especially in cases where the licenses are easily defined and the ownership of data is clear. It is also possible to add materials to the infrastructure that need various restrictions. One alternative is to make the data available for academic

research, which means that anyone with an account in a university that belongs to the HAKA-system can use the materials after login. This is relatively strong authentication, with a problem that the universities within this system are mainly located in Europe. However, for partially controlled use it is currently the best option offered in this infrastructure. It is also possible to make the materials available on request, which would demand an application and explanation of intended use. Although laborious to apply and process, this access method would in principle be open to anyone who fills the conditions corpus creators have set for their materials.

Table 1. provides a statistical summary for token and phrase coverage of language corpora on the Helsinki korp.csc.fi server.

English name	Project	Number of word forms	Number of sentences
Eastern & Meadow Mari	Fenno-Ugrica, Wanca	2,968,864	272630
Erzya	Erme, Fenno-Ugrica, SUS-Kenttätyö, Wanca	2,042,388	201,104
Moksha	Erme, Fenno-Ugrica, SUS-Kenttätyö, Wanca	1,673,480	142,248
Tundra Nenets	Fenno-Ugrica, Wanca	1,570,794	226,736
Hill or Western Mari	Fenno-Ugrica, Wanca	1,267,860	120,925
Mansi	Fenno-Ugrica, Wanca	936,869	120,598
Võro	Wanca	903,900	73,043
Veps	Fenno-Ugrica, Wanca	655,549	91,721
Khanty	Fenno-Ugrica, Wanca	620,815	72,934
Udmurt	Wanca	613,043	57,873
Ingrian	Fenno-Ugrica, Wanca	465,672	65,823
Selkup	Fenno-Ugrica	386,354	76,335
Komi-Zyrian	SUS-Kenttätyö, Wanca	233,596	22,286
Olonets-Karelian	Wanca	128,916	10,442
Komi-Permyak	Wanca	93,010	8,595
Skolt Sami	Wanca	91,154	8,402
Udmurt Beserman	BeserCorp	47,229	5,234
Karelian	Wanca	33,609	2,849
Ludic	Wanca	8,802	802
Kildin Sami	Kildin Saami Sample, Wanca	2,515	222
Nganasan	Wanca	2,161	67
Votic	Wanca	191	22

Table 1. Coverage for minority Uralic languages spoken in Russia

The infrastructure described in this article offers a flexible and robust platform for various types of linguistic corpora. Although numerous alternatives exist for tools described here, many with much larger resources, the relationship of different corpus providers should probably be best seen as complementing than competing. For example, for Komi-Zyrian komicorpora-website maintained by FU-Lab (FU-Lab 2019), is certainly the source of widest and first priority. Similarly, for Sami languages the materials provided by Giellatekno are in many ways crucial. However, wider representation of partially overlapping resources in differences interfaces, with possibly differing annotation schemes, is perhaps ultimately only beneficial for the field at large.

Looking onward, there are several questions for which the best practices still have to be adopted, and possibly invented. The versioning of the corpora, which would allow easy and flexible updating, is one of the tasks where further development is certainly needed. This, however, is not so much a technical problem, but relates to the practices and responsibility sharing around the corpus maintenance. As far as we see, consistent versioning, clear licensing, persistent identifiers and automatic validation must somehow be integrated to the workflows how these corpora are maintained and updated in the future.

## References

Ahlberg Malin, Lars Borin, Markus Forsberg, Martin Hammarstedt, Leif-Jöran Olsson, Olof Olsson, Johan Roxendal, and Jonatan Uppström. // Korp and Karpa bestiary of language resources: the research infrastructure of Språkbanken. Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA 2013), P. 429–433. 2013.

FU-Lab. Корпус коми языка: [Электронный ресурс]. URL: <http://komicorpora.ru/> (дата обращения: 28.10.2019).

Hämäläinen M., Säily T., Rueter J., Tiedemann J., & Mäkelä E. Revisiting NMT for normalization of early English letters. In B. Alex, S. Degaetano-Ortlieb, A. Kazantseva, N. Reiter, & S. Szpakowicz (Eds.),

Proceedings of the 3rd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature. P. 71–75. (ACL Anthology; No. W19-25). Stroudsburg: Association for Computational Linguistics, 2019.

Jauhiainen, Tommi, Heidi Jauhiainen, and Krister Lindén // The Finno-Ugric languages and the internet project. First International Workshop on Computational Linguistics for Uralic Languages Proceedings of the Workshop. Septentrio Academic Publishing. 2015.

Mika Hämäläinen. (2018, January 9). Python Korp Library (Version v1). Zenodo. <http://doi.org/10.5281/zenodo.1143374>

Nivre, Joakim; Abrams, Mitchell; Agić, Željko; et al., 2019, *Universal Dependencies 2.4*, LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, <http://hdl.handle.net/11234/1-2988>.

Partanen, Niko, Rogier Blokland, KyungTae Lim, Thierry Poibeau, and Michael Rießler. The First Komi-Zyrian Universal Dependencies Treebanks // Second Workshop on Universal Dependencies (UDW 2018), November 2018, Brussels, Belgium, P. 126–132.

Rueter, Jack M. & Tyers, Francis. Towards an open-source universal-dependency treebank for Erzya. Proceedings of the International Workshop for Computational Linguistics of Uralic Languages. Helsinki, Finland, 08/01/2018–09/01/2018.

## О КОНЦЕПЦИИ НАЦИОНАЛЬНОГО КОРПУСА БАШКИРСКОГО ЯЗЫКА

В статье дается анализ работы лаборатории лингвистики и информационных технологий по созданию корпусных проектов по башкирскому языку. Ставится задача создания на основе этих проектов национального корпуса языка. Дается определение национального корпуса, критерии разграничения от простых корпусов и корпусных проектов. Автором раскрывается концепция создания национального корпуса башкирского языка, определяются этапы его реализации.

**Ключевые слова:** корпус, корпусная лингвистика, количественная лингвистика, башкирский язык.

Зарубежное лингвистическое направление, возникшее в середине прошлого века, связанное с компьютерной обработкой информации больших объемов, сформировалось в новую динамично развивающуюся область филологии — корпусную лингвистику. Объектом данной области языкознания являются речевые материалы, реализованные в виде как письменных текстов, так и устных (фонетических) массивов данных. Корпусная лингвистика занимается созданием общих унифицированных принципов представления таких сверх-больших массивов языковых данных (корпусов), непосредственным составлением самих корпусов и выполнением теоретических и практических лингвистических исследований на базе этих данных [Захаров 2005, 3].

На сегодня в мире насчитываются более тысячи корпусов, их количество растет с каждым годом экспоненциально. Такой интерес к корпусам обусловлен тем, что это направление лингвистики активно влияет на все остальные направления языкознания, изменяет теоретические приоритеты и создает новые идеологии в понимании того, что же представляет собой язык [Плунгян 2008, 7-8]. Корпусы также открывают перспективу для новых исследований не только в области лингвистики, но и в смежных областях: в литературоведении (для стилистических исследований, определения нормативности употребления языковых реалий), в общественных науках (изучение социальных объектов через язык, используя такие параметры текстов, как период, автор или жанр, семантический контент текстов), в информационно-технических разработках (создание автоматизированных систем машинного перевода, распознавание речи, информационный поиск).

Сейчас все крупные языки обзавелись своими национальными корпусами. Ведутся корпусные разработки и по языкам народов России. Сегодня в сети Интернет функционируют корпусные проекты ряда языков: коми, калмыцкого, лезгинского, осетинского, тувинского, шорского.

Учитывая актуальность корпусных разработок, лабораторией лингвистики и информационных технологий Института истории, языка и литературы УНЦ РАН разработаны следующие корпусные проекты: прозаических, публицистических и фольклорных текстов башкирского языка.

1. Проект корпуса прозаических текстов. На сегодня этот корпус содержит тексты 1299 произведений 163 прозаиков. Паспортизация текстов включает: название текста, ФИО автора, объем текста (указывается объем в словоформах, например: “Дус булайык” — 53419 словоформ), время создания (если автором указано время создания, то указывается эта дата, если нет, то дата издания (например: “Мөхәббәт һәм нәфрәт”, 1964).

2. Проект корпуса публицистических текстов. Корпус содержит тексты республиканских газет “Башкортостан” (2016-2017), “Йәшлек” (2010, 2016-2017), “Киске Өфө” (2009-2011); районных газет “Ашказар” (2009), “Табын” (2000, 2009, 2012-2013); журналов “Ағизел” (2010-2013), “Башкортостан кызы” (2016-2017). Система экстралингвистических разметок публицистического корпуса включает название прессы, год, месяц и день выхода, название статьи, автора. Все тексты размечены по тематике и жанру. В данном корпусе выделены следующие тематики и жанры:

А. Тематика: политическая и социальная жизнь (политика, право, философия); экономика (производство, строительство, бизнес, финансы, коммерция); сельское хозяйство; искусство, культура и литература; наука и техника; образование; природа, путешествие; частная жизнь; спорт; религия; психология; медицина; красота и здоровье.

В. Жанры текстов: интервью, беседа; статья, очерк, репортаж, обозрение; советы; письма; обзор печати (новости из других источников); поздравления; художественно-публицистические жанры (эссе, фельетон, рассказ, стихи, эпиграммы); рецензия.

3) Проект корпуса фольклорных текстов. Данный корпус содержит фольклорные тексты из многотомных изданий фольклора башкирского народа и неизданных материалов из архивного фонда УФИЦ РАН.

Система экстралингвистических разметок фольклорного корпуса включает: название текста, жанр, жанровую разновидность/тематику, источник.

Определены следующие жанры и жанровые разновидности:

эпосы: исторические, мифологические, бытовые;

сказки: волшебные, богатырские, бытовые, о животных;

легенды: мифологические, этнонимические, этногенетические;

предания: бытовые, исторические;

песни: исторические, лирические;

баиты: исторические, бытовые;

такмаки: плясовые, игровые, обрядовые, такмаки айтыша, такмаки-памятки, сценические, небылицы;

обрядовый фольклор: родильный, свадебный, погребальный;

мунажаты: религиозные, философские, судьбы человеческие, дидактические;

поговорки: труд, общественная жизнь, быт, обычаи, родственные отношения, сущность человека, практическая философия;

пословицы: труд, общественная жизнь, быт, обычаи, родственные отношения, сущность человека, практическая философия;

загадки: земля и небо, явления природы, растительный мир, животный мир, человек и его жизнь;

поверья;

скороговорки;

приметы;

заклятья;

проклятья;

клятвы.

Почему я так подробно остановился на этих жанрах: пользователь может при поиске использовать их как фильтры.

В Государственной программе “Сохранение и развитие государственных языков Республики Башкортостан и языков народов Республики Башкортостан” на 2019-2024 гг. В разделе “Башкирский язык” есть пункт 2.3.3. “Осуществление мероприятий по созданию и поддержке системы “Национальный корпус башкирского языка”.

Что же такое национальный корпус языка и чем он отличается от простого корпуса? В корпусной лингвистике даются следующие определения национального корпуса языка.

«Национальный лингвистический корпус – огромная коллекция устных и письменных текстов различных жанров, стилей, региональных и социальных вариантов, представленных в языке и интересных для изучения языка» [Сысоев 2010, 100].

«Национальный корпус – это собрание текстов в электронной форме, представляющих данный язык на определенном этапе его существования, отображающий данный язык во всем многообразии жанров, стилей, социальных и территориальных диалектов и т.п. Корпус должен быть представительным, т.е. содержать по возможности все типы письменных и устных текстов» [Волоснова 2009, 43].

Учитывая все вышесказанное, планируется перейти с этапа корпусных проектов на этап разработки Национального корпуса башкирского языка, который включает в себя разработки подкорпусов, охватывающих новые формы, жанры и стили языка.

Предстоит работа по созданию следующих подкорпусов:

подкорпус учебной и научной литературы, включающий все школьные, сузовские и вузовские учебники, гуманитарные научные публикации, материалы из энциклопедий на национальном языке;

подкорпус параллельных текстов, включающий русско-башкирские параллельные тексты;

подкорпус поэтических текстов;

подкорпус официальных текстов и документов.

Составители Национального корпуса русского языка, являющегося эталоном для корпусных разработок по национальным языкам народов Российской Федерации, также указывают, что национальный корпус имеет две важные особенности, одним из которых является содержание в нем всех типов письменных и устных текстов, представленных в данном языке (художественные разных жанров, публицистические, учебные, научные, деловые, разговорные, диалектные и т.п.) [Национальный корпус русского языка]. Поэтому любой корпус, который не содержит устную или письменную составляющую или не представляет все стили и жанры языка, не может претендовать на звание национального корпуса, является просто корпусом.

Учитывая это, мы планируем создать и подкорпусы устной речи: фонетический подкорпус произвольного разговорного языка, фонетический подкорпус литературного языка, фонетический подкорпус говоров диалектов башкирского языка.

В лингвистике также определен минимальный объем корпуса в словоупотреблениях, когда он может считаться национальным. Большинство создателей и исследователей считают, что объем национального корпуса должен превышать 100 млн. словоупотреблений [Богоявленская 2016]. Такие требования большого объема для национального корпуса основываются на мнении Синклера [Sinclair 1991] о том, что сколько-нибудь значащие статистические лингвистические данные и результаты можно получать только на достаточно большом объеме материала.

В 70-е годы прошлого столетия К. Б. Бектаевым [Бектаев 1978, 33] была предложена формула для расчета достоверного объема текстов для построения частотных словарей:

$$N = \frac{Z^2(1-f_{\text{тр}})}{\delta^2 f_{\text{тр}}}$$

Если использовать эту формулу для расчета граничной достоверной частоты слова, то для  $Z_{p=1,96}$  (при надежности  $p=0,95$ ) и допустимой относительной ошибки в  $\delta \leq 0,33$ , граничная частота будет определена как  $f_{\text{тр}} \approx 5 \cdot 10^{-6}$ , что для тюркских языков означает область сверх низкочастотных слов. Т.е. корпус такого объема или его превышающий не только даст достоверные статистические результаты по высокочастотной и среднечастотным словам языка, но также будет включать в себя редко употребительные низкочастотные слова национального языка.

Следующим из важных требований к национальному корпусу является репрезентативность. В корпусной лингвистике репрезентативность понимается как пропорциональное представление в корпусе текстов различных периодов, жанров, стилей, авторов и т. п. [Волоснова 2006, 47]. Репрезентативность должна обеспечить национальному корпусу возможность получать представления о языке в целом.

Данное понятие как основное в корпусной лингвистике, тем не менее не является тривиальным. Так, по анализам Д. Байбера, корпус, составленный на базе пропорционального охвата типов речи и языковых стилей, должен был бы содержать около 90% обычной разговорной речи, 3% писем и замечаний и 7% опубликованных и неопубликованных текстов классических стилей и жанров [Biber 2000]. Учитывая трудоемкость составления корпусов устной речи, составить корпус, реально отражающий язык в представленных выше пропорциях, весьма проблематично.

Однако определение репрезентативности к национальному корпусу как содержание всех типов письменных и устных текстов, и их разных жанров, по возможности пропорционально их доли в языке, декларируется как основное требование и в Национальном корпусе русского языка [НКРЯ]. Сегодня составители национальных корпусов, не касаясь проблемы соотношения устных и письменных видов речи, которая не решится в ближайшее время, предлагают разные пропорции включения в устную и письменную часть корпусов стилевых и жанровых типов языковых текстов.

Считаем, что вопрос репрезентативности будущего Национального корпуса башкирского языка с учетом лингвистических и экстралингвистических факторов, например: временного среза, объема стилей и жанров в географической перспективе (районы, издательства) и др. должны стать объектом дальнейшего совместного обсуждения широкого круга филологов республики. На начальном этапе нами для текстовой части национального корпуса предложена следующая пропорция: проза – 40%, публицистика – 40%, учебная и научная литература – 5%, фольклор – 5%, поэзия – 5%, официально деловой стиль – 5%.

Сбалансированность корпуса ставится как задача финальная, а не соблюдаемая на всех этапах построения национального корпуса башкирского языка, как и удовлетворения национального корпуса требованию полноты. На этапе доведения его объема до 100 млн словоупотреблений эти вопросы не стоит ставить, поскольку они будут не просто затягивать работу, но и весьма затруднительными даже при значительных людских ресурсах, например: поиск текстов для слов, которые зафиксированы в толковых словарях, но не отражены в корпусе. Но на стадии заявления о статусе корпуса как Национального, этот вопрос должен подниматься.

Преодоление объема корпуса в 100 млн. словоупотреблений будет достигнуто за счет увеличения текстовых баз существующих корпусных проектов и добавления новых подкорпусов.

Структурированность представления данных также является одним из важных требований к Национальному корпусу. Какие информативные единицы будут закладываться в базу корпуса, что может пользователь искать и находить в корпусе?

В существующих корпусных проектах башкирского языка заложена морфологическая информация, которая включает а) частеречную характеристику; б) совокупность морфологических признаков по типу агглютинативных аффиксов словоизменения, которые подразделяются на именные и глагольные формы. Выделяются 12 частей речи: имена существительные, числительные, прилагательные, наречия, глаголы, местоимения, подражательные слова, междометия, модальные слова, союзы, частицы, послелоги.

В Национальном корпусе будет информация по семантике слова. Сейчас идет работа выделения семантических групп имен существительных, которые разделены на семантические группы: предметные имена, неперечисленные имена и имена собственные. Определены следующие семантические группы предметных имен:

- лица (кеше, укытыусы, табиб, ғалим),
- этнонимы (башкорт, татар, сыуаш),
- части тела (кул, аяк),
- термины родства (эсэй, килен, ағай),
- пища и напитки (һөт, икмәк, ит, кымыз),
- постройки (өй, торлак, каралды),
- одежда (баш, өс, аяк кейемдәре),
- украшения (беләзек, мунсак, һакал),
- животные (өй һәм кырағай хайуандар, коштар),
- растительный мир (кайын, үлән, япрак),
- транспорт (сана, машина, трактор),
- вещества и материалы (һыу, таш, сатин, камыр),
- пространство и место (йыһан, кала, тайга, йырын),
- инструменты и приспособления (сүкеш, балта, энә),
- механизмы и приборы (телефон, телевизор),
- оружие (кылыс, ук),
- посуда (шәшке, самауыр, калак),
- текст (хикәйә, китап, афиша),
- мебель (*шкаф, өстәл, ултырғыс*),
- **музыкальные инструменты (курай, кумыз, аккордеон).**

На базе определенных групп составлен словарь предметных групп имен существительных объемом. В дальнейшем предстоит составить словарь неперечисленных существительных, глаголов и остальных частей речи, разработать систему поиска по семантике.

В национальном корпусе должна быть представлена информация по лексико-грамматической омонимии и осуществлена возможность автоматического снятия этих неоднозначностей. Данная проблема требует проведения практических исследований на основе корпусных проектов и теоретического обобщения.

В существующих корпусных проектах реализована возможность представления информации по вкраплениям из иноязычных слов. Работа в этом направлении пока не идет, но в национальном корпусе все вкрапления должны быть маркированы по языкам вхождения.

Национальный корпус башкирского языка несомненно должен представлять и синтаксическую информацию на основе синтаксической разметки текстов. Поиски в этом направлении начаты.

Что касается звуковой составляющей национального корпуса, она должна содержать корпуса говоров диалектов, разговорный литературной речи, наддиалектной непринужденной разговорной речи, общим объемом звучания в 1000 часов.

Сегодня в лаборатории ведется работа над созданием аудиокорпуса говоров башкирского языка. Для сбора полевых материалов выбраны населенные пункты, которые были в свое время определены в качестве опорных пунктов для диалектологического атласа башкирского языка.

Аудиозаписи сопровождаются информацией, которая составляет экстралингвистическую разметку материала:

- пол: мужской, женский;
- образование: начальное, среднее (среднее школьное или суз), высшее;
- возраст;
- язык обучения: башкирский, русский, татарский, чувашский;
- язык общения в семье: башкирский, русский, татарский;
- национальность информанта;
- имя, отчество, фамилия;



- место последнего долгого проживания до переезда в данное место (в случае переезда);
- время проживания до последнего места проживания (в случае переезда);
- время записи.

Выделяется тип общения: монолог, диалог, полилог.

Для записи информантов определены следующие 15 тем:

- свадьба, свадебные обычаи;
- обычные блюда (что любят готовить на каждый день, что на праздниках);
- какой скот и как содержат;
- дети и близкие родственники (система родства);
- приусадебное хозяйство (огород, сад);
- о доме (когда построен, кто строил, крыша, рамы);
- топонимия в окрестностях поселения;
- история села, школы, рода;
- работа, школа, друзья;
- времена года, погода;
- частушки, пословицы, поговорки, сказки (какие знают);
- поездка в райцентр (по каким делам, каким транспортом пользуются);
- игры детей (какие игры и как играют);
- друзья детей (какие у детей друзья, где они живут);
- животный мир около поселения (какие птицы и звери обитают).

Запись осуществляется на цифровой диктофон в несжатом формате (.WAV, 16бит/22kHz – 16бит/48kHz), при отсутствии посторонних звуков. Первичная обработка аудиозаписей (очистка от посторонних шумов и длительных пауз), паспортизация производится в программе Sound Forge.

Разработанная для данного корпуса система аннотирования и транскрибирования будет использована и при составлении других фонетических подкорпусов.

Таковы основные моменты концепции Национального корпуса башкирского языка, который будет несомненно создан при условии финансирования со стороны Правительства Республики Башкортостан.

В заключении хотелось бы отметить, что система финансирования путем проведения тендера для научных исследований не является оптимальным. Обычно финансы поступают в конце года, а работа планируется на год. В соседних национальных республиках аналогичные мероприятия финансируются на постоянной основе. Так сотрудники Института Семиотики АН РБ (РТ), Центра инновационных лингвистических исследований (Сыктывкар, КР), Лаборатории в Марий-Эл и в Калмыкии финансируются на постоянной основе без проведения конкурсов.

#### **Литература:**

Бектаев К. Б. Статистико-информационная типология тюркского текста. Алма-Ата: Наука, 1978. – 167 с.

Богоявленская Ю. В. Репрезентативность лингвистического корпуса: метод верификации достоверности полученных данных // Политическая лингвистика, 2016. – С.163–166.

Волоснова Ю. А. Корпусная лингвистика: проблемы и перспективы // Лесной вестник. Филология, 7/2006, – С.43–49.

Захаров В. П. Корпусная лингвистика: Учебно-методическое пособие. – СПб., 2005. – 48 с.

Национальный корпус русского языка. [Электронный ресурс]. URL: <http://www.ruscorpora.ru> (дата обращения: 17.09.2019).

Плунгян В. А. Корпус как инструмент и как идеология: о некоторых уроках современной корпусной лингвистики // Русский язык в научном освещении, 2008, № 16 (2), – С. 7–20.

Сысоев П. В. Лингвистический корпус в методике обучения иностранным языкам // Язык и культура. 2010. № 1. – С. 99–111.

Biber, D.: Repräsentativnost v projektu korpusu. Studie z korpusové lingvistiky. Acta Universitatis Carolinae. Philologica 3-4. Praha: Univerzita Karlova – Nakladatelství Karolinum 2000, – С. 107–136.

Sinclair J. Corpus. Concordance and Collocation/Oxford: Oxford University Press, 1991. – 137 p.

## **О ПОЛУАВТОМАТИЧЕСКОЙ МЕТАРАЗМЕТКЕ КАЗАХСКОГО КОРПУСА**

В статье рассматривается проблема металингвистической разметки текстов национального корпуса казахского языка, составляемого сотрудниками Института языкознания им. А. Байтурсынова Министерства науки и образования Республики Казахстан. Описываются системы метаразметок национальных текстов, которые включают такие параметры как фамилия и имя автора, время создания, тип текста, тематическая и стилевая принадлежность, объем текста в словоупотреблениях и др., так и данные о лице проводившем саму разметку текста. Метаразметка обеспечивает в корпусе возможность поиска и отбора текстов пользователем при составлении подкорпусов с заданными свойствами.

**Ключевые слова:** корпус, аннотированный корпус, метатекстовая разметка, окно, поиск.

*Г. Тлегенова, А. Қожахметова  
Алматы қ., Қазақстан*

## **ЖАРТЫЛАЙ АВТОМАТТЫ МЕТАБЕЛГІЛЕНІМ СИПАТТАМАСЫ**

Мақалада ұлттық корпусстың құрамындағы мәтіндерді металингвистикалық белгілеу мәселесі қарастырылған. Ұсынылған қазақ мәтіндерінің мета-белгілеу жүйелері сипатталған. Мета-белгілеуде мәтінді тұтасымен сипаттайтын параметрлер мәтініне сілтеме жасау жатады. Мета-таңбалау, қолданушының көрсетілген қасиеттері бар ішкі құрылымды құрастыру үшін мәтіндерді іздеу және таңдау мүмкіндігін береді.

**Түйін сөздер:** регистр, аннотацияланған регистр, метамәтін белгілеу, терезе, іздеу.

Тіл білімінде соңғы жылдары компьютерлік технологиялардың көмегімен автоматтандыру мәселесі қолға алынып келеді. Ендеше ақпараттық технологияларды үйлесімді пайдалану қазіргі таңда өзекті мәселелердің біріне айналып отыр. Ғылымда, білім беру жүйесінде компьютерлерді пайдаланудың маңызы зор. Өйткені компьютер адам қызметінің барлық саласында еңбек тиімділігін арттыру құралына айналды.

Корпустық лингвистика қазақ тіліне еніп отырғанына көп уақыт бола қоймаса да, өз кезегінде практикалық-теориялық маңыздылығы арту үстінде. Көптеген жаңа терминдер еніп, оларға анықтама беріліп, заман сұранысына сай жаңа бағыттың қарқынды дамуына көңіл бөлініп отыр.

Корпус құруда метабелгіленім қоюдың маңыздылығы өте қажетті. Айталық Қазақ тіл біліміндегі корпуссты қолдану барысында ең әуелі таңдама мәтіндердің параметрлерін анықтап алуымыз қажет. Мысал ретінде, көркем әдебиет, оның ішіндегі әңгіме не роман, әңгіменің жазылған жылы, жазған адамның ер не әйел екендігі (жынысы), яғни алып отырған мәтін туралы жалпы ақпараттың бірізді формамен, математика тілімен айтқанда алгоритмі құрылуы тиіс. Осы алгоритмнің негізінде программа құрылып компьютердің көмегімен жүзеге асады. Метабелгіленім қою мәтін туралы ақпаратты алуда және біршама мәселелерді шешуде көмегін тигізеді.

Метабелгіленімдер – ғылыми-зерттеу жұмыстарында белгілі бір кезеңге, стильге, авторға, тақырыпқа т.б. қатысты материалдар жинаудың таптырмас дереккөзі десек те болады.

Метабелгіленімдер қойылған корпустардан зерттеушілер өзіне қажетті стиль, кезең, автор т.б. ақпараттар бойынша мәліметтерді тез тауып алуына мүмкіндік алады. Ал бұл жетістік қазіргі дамыған еліміздің ғылыми-зерттеушілік әлеуетін, ғылым мен білімді жан-жақты дамытатын бірден-бір күш екендігі сөзсіз. Корпустың қандай мәтінмен жұмыс жасасақ та ең алдымен жинақталған мәтін туралы ақпарат беретін метамәтіндік параметрлер нақтылануы қажет. Метамәтіндік белгіленімдер корпустың қай түріне болмасын енгізілуге тиісті маңызды ақпараттар болып табылады.

Жартылай метабелгіленім енгізу бағдарламасы **23** параметрден тұрады:

1) Автордың аты-жөні. Бұл параметрде автордың толық аты-жөні, болмаған жағдайда бос қалдырылады.

2) Құрастырушы. Ұжымдық, жеке құрастырушылардың аты-жөні беріледі.

3) Мәтін аты. Бұл жерде мәтіннің атауы беріледі. (Мәтін оқулық ішінен алынған тақырыптан алынса, тақырып атауы жазылады. Ал оқулық атауы дереккөзде көрсетіледі).

4) Мерзімі (жазылған уақыты). Мұндай белгіленімдер автордың шығарманы жазу барысында мәтіннің соңында қалдырған мәліметтерден алынады.

5) Белгілеуші. Бұл жерге метабелгіленімді жасаған адамның аты-жөні жазылады.

6) Аудармашының аты-жөні. Мәтін аударма болған жағдайда аудармашының аты-жөні көрсетіледі.

7) Автордың жынысы. Автордың жынысы мәтін авторы біреу болған жағдайда көрсетіледі; яғни ер не әйел деп біреуі қойылады. Ұжымдық мәтін авторлары әр түрлі жынысты болған жағдайда «ер/әйел» деген белгіленім қойылады.

8) Автордың туған уақыты. Автордың туған жылы, күні туралы нақты мәліметтер көрсетіледі. Ұжымдық авторлар болған жағдайда бос қалдырылады.

9) Серия аты. Мәтін телеарна, радиодан алынған (ауызша, жазбаша) болса, бағдарлама сериясы жазылады.

10) Аудитория жасы. Мектепке дейінгі, бастауыш мектеп жасы, жасөспірімдік кезең, жастар әдебиеті, бейтарап.

11) Субкорпус. Газет мәтіндері, ауызша, мультимедиялық, көркем мәтіндер, поэтикалық мәтіндер, оқыту, ғылыми-техникалық корпус, ісқағаз мәтіндері.

12) Мәтін хронотопы. Мәтінде суреттелген оқиғамен шамамен алғандағы мерзімі, өткен орны көрсетіледі.

13) Стиль. Көсемсөз (публицистикалық), ғылыми-көпшілік, ғылыми стиль, аралас, көркемсөз, ісқағаз, сөйлеу стильдері көрсетіледі.

14) Мәтін типі. Белгілі бір стильдің типтері көрсетіледі.

15) Сөзформа саны. Жеке файлда тұрған әрбір мәтінді ашқанда компьютер экранының төменгі жағында сол мәтіндегі сөз саны шығып тұрады, осы сан метабелгіленімге енгізіледі.

16) Таралым типі. Мәтіннің қай жерден алынғаны жазылады (газет, журнал, ғаламтор-ресурс, кітап, мультимедиа)

17) Мәтін формасы. Мәтіндер жазбаша мәтіндер болса, «жазбаша», егер видео/аудио жазбалар болған жағдайда «ауызша» деген белгіленім қойылады.

18) Графика. Мәтіннің қандай графикада жазылғандығын көрсетеді (кирил жазуы, латын жазуы, төте жазуы).

19) Мәтін тақырыбы. Тақырып белгілі бір салаға байланысты қойылады. Тақырып көркем әдебиетке қойылмайды.

20) Мәтін көзі. Бұл жерде мәтін алынған дереккөзі көрсетіледі (баспа, көшірме (сканер), интернет-ресурс, электронды көшірме, қолмен терілген, қағаз көшірме).

21) Корпусқа енгізілу уақыты. Мәтіннің корпусқа енгізілген мерзімі жазылады.

22) Жарық көрген уақыты. Кейбір мәтіндер (шығарма, монография, оқулық т.б.) жазылған уақытында жарық көрсе, кейбірі кейін өңделіп қайта басылады. Бұл жерде кітаптың жарыққа шыққан, яғни қайта басылып шыққан уақыты көрсетіледі.

23) Дереккөзі. Мәтіннің дереккөзі ретінде ол туралы толық ақпарат жазылады (жері, баспа, жылы).

Жоғарыда аталған метабелгіленім параметрлері бойынша корпусқа енгізілетін мәтіндерге белгіленім осы жартылай автоматты метабелгіленім қою бағдарламасы арқылы жүзеге асады (1-сурет).

Сурет 1. 23 параметр енгізілген метабелгіленім бағдарламасы

Демек, корпусстың қандай түрін алсақ та, қандай мәтінмен жұмыс жасасақ та ең алдымен жинақталған мәтін туралы ақпарат беретін метамәтіндік параметрлер компьютерлік базаға салынуы қажет. Сондықтан корпус құрастыруда метабелгіленімдер (шығарма атауы, авторы, жылы, шыққан орны, стилі, жанры, көлемі т.б.) жасаудың әдіс-тәсілдерін, түрлерін, компьютерлік бағдарламаларын т.б. зерттеудің маңызы зор.

Қазақ тілінің ұлттық корпусы тиімді құрастырылып, өз дәрежесінде қолданысқа енгізілуі үшін әлемдік тәжірибедегі метабелгіленімдер қоюдың ортақ ұстанымдары мен әдіс-тәсілдерін зерттеу қажет. Ал бұл жетістік қазіргі дамыған еліміздің ғылыми-зерттеушілік әлеуетін, ғылым мен білімді жан-жақты дамытатын бірден-бір күш екендігі сөзсіз.

#### **Әдебиет**

Национальный корпус русского языка. Параметры текстов. [Электронный ресурс]. URL: <http://ruscorpora.ru> (дата обращения: 10.09.2019).

## КОРПУС КОМИ ЯЗЫКА КАК БАЗА ДЛЯ НАУЧНЫХ ИССЛЕДОВАНИЙ

В статье рассматриваются возможности использования письменного корпуса коми языка для лингвистических исследований.

**Ключевые слова:** коми язык, корпус коми языка, лингвистика, частотность, репрезентативность.

Письменный корпус коми языка (<http://komicorpoга.ru/>) как информационно-справочная система, основанная на собрании текстов коми литературного языка в электронной форме, начал создаваться в 2015 г. Он постоянно модернизируется и дополняется. На данный момент он включает 103 871 текст и его объем составляет 56 930 079 словоупотреблений.

В корпусе коми языка представлено восемь подкорпусов: **художественная проза** — в данном подкорпусе преимущественно представлена художественная проза (романы, повести, рассказы и т.п.); **документальная проза** — прочие прозаические произведения небеллетрического характера, опубликованные в книжных и журнальных изданиях (очерки, публицистические статьи и т.п.); **драма** — произведения коми драматургов (драмы, комедии, трагедии и т.п.); **газетный** — тексты из комиязычных средств массовой информации (газеты «Коми му», «Йӧлӧга», «Парма гор», «Вьль туйӧд», «Звезда», Радио «Юрган»); **учебно-научная** — работы учебного и научного характера, написанные на коми языке; **поэтический** — поэтические произведения на коми языке (стихотворения, поэмы, песни и т.п.); **фольклор** — произведения коми народного творчества, опубликованные в орфографии литературного языка; и **другое** — в данный раздел собираются тексты, которые пока не распределены по отдельным подкорпусам.

Корпус коми языка дает очень широкие возможности для разноплановых лингвистических научных исследований как репрезентативная база функционирования исследуемых языковых единиц в естественной и реальной языковой среде на основе объективных статистических данных. «На корпусах текстов можно проверять лингвистические гипотезы, теории, они могут использоваться для любых практических работ с языковым материалом, в частности, как массивы текстовых примеров на все трудные явления языка. Достаточно большой (репрезентативный) объем корпуса гарантирует типичность данных и обеспечивает полноту представления всего спектра языковых явлений. Именно корпус позволяет получать данные, недоступные при традиционных методах лингвистического анализа» [Ганиева 2007, 104].

При этом важнейшую роль играет время сбора и обработки фактического материала. Скорость и эффективность поиска информации зависят от программного обеспечения. В корпусе коми языка поиск и вывод искомой информации составляет на данный момент несколько секунд, к примеру, поиск и вывод данных по слову «морт» (человек), который встречается в 169 887 предложениях, составил 6 секунд.

Поиск в корпусе разработан с учетом запросов разноподготовленных пользователей — достаточно просто и удобно. При нажатии кнопки «Поиск» внизу формы поиска справа от формы поиска появится область вывода результатов, при этом форма поиска сдвигается влево. В области вывода результатов поиска представлены предложения, в которых встречается искомая комбинация символов (слова, части слова, сочетания слов). Одновременно выводится до 30 предложений. К следующим тридцати можно перейти, нажав кнопку «Далее». Можно также перейти в конец списка найденных предложений либо к его началу. Перед каждым найденным предложением указываются метаданные: автор, название произведения, источник, год (дата) издания. Искомая комбинация символов выделяется красным цветом. При поиске словосочетаний выделение цветом не происходит. При нажатии левой кнопкой на любое слово в области вывода появится выпадающее меню, где будут указаны морфологическая характеристика искомого слова, его лемма и перевод.

При этом есть несколько вариантов поиска:

А) все варианты — будут найдены все возможные комбинации, в которых представлена искомая последовательность символов;

Б) точное совпадение — результатом поиска будет только та последовательность символов, которая введена в поиск;

В) только как часть слова – будут найдены все комбинации, в которых последовательность символов представлена как часть слова. При выборе поиска по части слова можно уточнить позицию в слове (в начале слова, в конце слова, в составе слова, все варианты);

Г) поиск по лемме (словарной форме) – будут учитываться все словоформы данного слова.

Кроме традиционных видов поиска, в корпусе коми языка предусмотрен и поиск по регулярным выражениям, когда с помощью специальных символов организуется поиск по различным формальным критериям. Основными из них являются следующие:

Символ	Результат/Использование
Любой символ	Представление данного символа, если не указано иное.
.	Представление любого отдельного символа, кроме разрыва строки или конца абзаца. Например, условие поиска «п.ск» возвращает и «пуск», и «писк».
^	Возможность найти термин, только если этот термин находится в начале абзаца. Особые объекты, например, пустые поля или привязанные к символу врезки, в начале абзаца игнорируются. Пример: «^Петр».
\$	Возможность найти термин, только если этот термин находится в конце абзаца. Особые объекты, например, пустые поля или привязанные к символу врезки, в конце абзаца игнорируются. Пример: «Петр\$». \$ сам по себе означает конец абзаца. С его помощью можно искать и заменять разрывы абзацев.
*	Поиск нулевого или большего числа символов перед «*». Например, при поиске «Аб*в» будут найдены «Ав», «Абв», «Аббв», «Абббв» и т.д.
+	Поиск одного или большего числа символов перед «+». Например, при поиске «АХ.+4» будет найдено «АХх4», но не «АХ4». Всегда будет найдена самая длинная возможная строка, соответствующая данному искомому элементу в абзаце. Если в абзаце содержится строка «АХ 4 АХ4», выделяется весь фрагмент.
?	Поиск нулевого числа или одного символа перед «?». Например, при поиске термина «Тексты?» будут найдены «Текст» и «Тексты», а при поиске «х(ab c)?у» будут найдены «ху», «хабу» или «хсу».
\	Поиск воспринимает специальные символы после «\» как обычные символы, а не как регулярное выражение (за исключением сочетаний \n, \t, \> и \<). Например, «текст\.» находит «текст.», а не «тексты» или «тексту».
^\$	Поиск пустого абзаца.
^.	Поиск первого символа абзаца.
[абв123]	Представление одного из символов, находящихся в скобках.
[а-д]	Соответствует любому из символов между «а» и «е», включая начальный и конечный символы. Символы упорядочены по кодовым значениям.
[а-де-х]	Представление любого символа, находящегося между «а» и «д» или между «е» и «х».
[^а-с]	Соответствует любому символу, кроме символов между «а» и «с».
	Поиск терминов, которые встречаются до « » и после « ». Например, результатом поиска «этот тот» будет «этот» и «тот».
{2}	Определяет допустимое количество вхождений данного символа перед открывающей скобкой. Например, результатом поиска «сине{2}» также будет «синева».
{1,2}	Определяет максимальное и минимальное допустимое количество вхождений данного символа перед открывающей скобкой. Например, результатом поиска «сине{1,2}» также будет «синее» и «синева».
{1,}	Этот параметр определяет, какое минимальное число раз может встречаться символ перед открывающей скобкой. Например, при поиске «сине{2,}» будет найдено «синее», «синеее» и «синееее».

В разделе «Дополнительное условие» можно ограничить поиск по одному из следующих критериев: автор, название, источник и год издания.

Если требуется произвести поиск с учетом регистра, необходимо поставить галочку в поле «Учитывать регистр».

Результаты поиска можно отсортировать по алфавиту (как по возрастанию, так и по убыванию). Сортировка может производиться по тексту, названию произведения, автору либо источнику.

Во-вторых, «одной из основных функций лингвистического корпуса являются возможности получения лингвостатистических данных о частоте использования тех или иных форм, слов или словосочетаний... Данная функция позволяет определить различия семантики синонимов, установить контексты, характерные для синонимичных слов, а также разграничить жанровые и стилистические особенности и оттенки значений лексических единиц» [Райскина 2015]. На основе данных по частотности употребления тех или иных языковых единиц можно не только определить более употребительные формы по синонимам, например, в случае синонимического ряда «ичёт», «дзоля», «посни» (маленький, малый, мелкий): «ичёт» используется наиболее часто – в 57666 предложениях; далее «посни» – в 15222; и «дзоля» наименее употребителен – в 7346. Можно анализировать частотность и фонетических вариантов, закрепленных в коми литературном языке, например, слово «чай» встречается в 6658 предложениях, а его вариант «тшай» – в 1891; слово «дзоля» – в 7346, а его вариант «дзöля» – в 380. Кроме того, исследования по частотности слов можно использовать и в методических целях при изучении коми языка, как это было сделано на портале по изучению коми языка «Коми кыв» (<http://komikyv.ru/>) в разделе «Словарный состав» (<http://komikyv.ru/node/344>) с выделением наиболее частотных коми существительных, глаголов, прилагательных и наречий. Стоит заметить, что эффективность использования лингвистических корпусов при обучении как родным, так и иностранным языкам постоянно отмечается методистами и учеными [Жокорева 2013, Сысоев 2010, Базарова 2015 и др.].

Кроме того, по корпусу языка можно проследить за изменением лексического состава данного языка, например, получить точные данные о времени вхождения нового слова. Так, слово «канму» (государство) начало встречаться в современных печатных коми текстах в 1996 г. Оно было использовано в двух источниках — в «Морт инöдъяс йылысь быдöнлы инана шыödчöм» (Всеобщая декларация прав человека) (перевод Красильниковой Н. А.) [Морт... 1996] и «Библия нимъяс коми ногон» (Библейские имена на коми языке) [Библия... 1996].

В лингвистике уже считается нормой составление словарей и грамматик на основе использования представительных корпусов текстов. Кроме реального иллюстративного материала с полной паспортизацией в виде метаданных, данные корпуса позволяют наиболее полно представить словарный состав языка. Так, анализ текстов в корпусе коми языка позволил выявить более 41 300 коми слов, не включенных в существующие нормативные словари коми языка, а анализ грамматических форм позволяет говорить о сильно развитой, в некоторых случаях даже излишней, вариативности в современном коми литературном языке.

В принципе, использовать данные корпуса можно в разных научных исследованиях. Как абсолютно справедливо отмечают В. П. Захаров и С. Ю. Богданова, «лингвисты-теоретики используют корпусы в качестве экспериментальной базы для проверки гипотез и доказательства своих теорий. Прикладные лингвисты (преподаватели, переводчики и т.д.) используют компьютерные корпусы при обучении языкам и для решения своих профессиональных задач. Особый класс пользователей представляют компьютерные лингвисты: они пытаются выявить и использовать статистические и лингвистические закономерности, присутствующие в текстах, для создания компьютерных моделей языка. Другие специалисты по языку (литературоведы, редакторы) также в ряде случаев могут получить ответы на интересующие их вопросы, обратившись к корпусу. Специалисты по общественным наукам (историки, социологи) могут изучать свои объекты через язык, используя такие параметры, как период, автор или жанр. Литературоведы используют корпусы для стилистических исследований. Наконец, корпусы используются для разработки и настройки различных автоматизированных систем (машинный перевод, распознавание речи, информационный поиск)» [Захаров, Богданова 2013, 81–82].

Хочется надеяться, что имеющийся качественный ресурс корпуса коми литературного языка будет более продуктивно использоваться учеными, особенно при описании разных аспектов языка, а также при обучении коми языку как родному и как государственному.

### Литература

Базарова Б. Б. Корпусная лингвистика и преподавание иностранных языков // Вестник Бурятского государственного университета, 2015. – Вып. 15. – С. 88–92.

- Библия нимъяс коми ногӧн // Войвыв кодзув. – Сыктывкар, 1996, № 6.
- Ганиева И. Ф. Об использовании корпусов в лингвистических исследованиях // Вестник башкирского университета, 2007. Т. 12. № 4. – С. 104–106.
- Захаров В. П., Богданова С. Ю. Корпусная лингвистика: Учебник для студентов направления «Лингвистика». 2-е изд., перераб. и дополн., – СПб.: СПбГУ. РИО. Филологический факультет, 2013. – 148 с.
- Кокорева А. А. Корпус параллельных текстов в обучении иностранному языку // Вестник Тамбовского госуниверситета. Гуманитарные науки. Педагогика и психология, 2013. – С. 57–62.
- Морт инӧдъяс йылысь быдӧнлы инана шыӧдчӧм (коми кыв вылын). – Мӧсква: «Права человека», 1996. – 16 лб.
- Райскина В. А. Применение современных методов корпусной лингвистики при анализе текста (на примере корпуса ВФМ) // Молодежный научный форум: гуманитарные науки: электр. сб. ст. по мат. XXV междунар. студ. науч.-практ. конф. № 6(24): [Электронный ресурс]. URL: [https://nauchforum.ru/archive/MNF\\_humanities/6\(24\).pdf](https://nauchforum.ru/archive/MNF_humanities/6(24).pdf) (дата обращения: 24.09.2019).
- Сысоев П. В. Лингвистический корпус в методике обучения иностранным языкам // Язык и культура, 2010. № 1(9). – С. 99–111.



## ЛИНГВИСТИЧЕСКИЕ БАЗЫ ДАННЫХ И ПРИЛОЖЕНИЯ. ФУНКЦИОНИРОВАНИЕ ЯЗЫКА В СЕТИ ИНТЕРНЕТ

*Х. Альнайар, М. Хаммалайнен,  
Н. Партанен, Дж. Рюйтер  
г. Хельсинки, Финляндия*

### ИНФРАСТРУКТУРА ОТКРЫТЫХ СЛОВАРЕЙ УРАЛЬСКИХ ЯЗЫКОВ

Мы представляем открытую онлайн-инфраструктуру для редактирования и визуализации словарей разных уральских языков (например, эрзя, мокша, скольт-саамский и коми-зырянский). Наша инфраструктура полностью интегрируется в существующую Giellatekno с точки зрения словарей XML и морфологии FST. Наш код в открытом источнике.

**Ключевые слова:** уральские языки, онлайн словари, словари в формате XML.

*K. Alnajjar, M. Hämmäläinen,  
N. Partanen, J. Rueter  
Helsinki, Finland*

### THE OPEN DICTIONARY INFRASTRUCTURE FOR URALIC LANGUAGES

We present an open online infrastructure for editing and visualization of dictionaries of different Uralic languages (e.g. Erzya, Moksha, Skolt Sami and Komi-Zyrian). Our infrastructure integrates fully into the existing Giellatekno one in terms of XML dictionaries and FST morphology. Our code is open source.

**Keywords:** Uralic languages, online dictionary, XML dictionaries

In order to revitalize severely endangered languages, such as many of the Uralic languages, enormous work is required to collect as many resources and knowledge about them as possible, while also involving their native communities. Digitizing the resources of endangered languages is crucial as it boosts the language resources in various ways, such as preserving them in a versioned manner and facilitating access to them globally. Scholars have produced valuable lexicographic resources (such as dictionaries and finite-state transducers) for endangered Uralic languages (e.g. Komi-Zyrian, Ingrian, Erzya, Moksha and Skolt Sami) in order to revitalize them.

We present a large-scale open-source MediaWiki-based dictionary for such languages, (named Akusanat) [Hämäläinen, Rueter 2018] and a customly-built and user-friendly web system (named Ve'rd) that improves and amending the knowledge presented in such dictionaries.

There is a myriad of active online dictionary projects targeting only one language that are under development by different people, who oftentimes are unaware of each other's contributions. In this section, we present some of the recent work on online dictionaries, which is heavily guided by the needs of one individual language. Our infrastructure differs from these projects in that its driving design principle is multilinguality and support for a multitude of different Uralic languages.

A recent dictionary for St. Lawrence Island Yupik [Hunt 2019] combines Foma-based morphological analyzers with an HTML based search interface. Unlike Akusanat, which does the morphological analysis and generation in the cloud, their solution runs the transducers on the client side with Foma's Javascript integration.

The Livonian dictionary consists of three databases, one – lexical, the second – morphological, and the third – a text corpus. While lemmas and their data are stored in the lexical database, and morphological forms are documented in the morphological database, all words indexed in the corpus refer to lemmas in the lexical database. Thus, all materials in the cluster can be accessed directly from the three databases [Ernštreits 2019].

There are also various attempts to build infrastructure for national majority languages. These projects also seem to be characterized by simultaneous use of different tools, with various connections to commercial software providers [Tavast 2018]. Also from this point of view there is clear demand for open and easily customizable dictionary editing and data retrieval platforms, such as the infrastructure presented here.

Akusanat is built using MediaWiki. MediaWiki is a well documented and open-source framework that comes with a set of fulfilled quality attributes such as support for multiple simultaneous users, user account management and a documented API. In addition, MediaWiki has been perceived as a useful framework for dictionaries in the past [Laxström, Kanner 2015].

Despite the features that MediaWiki has, it does not provide an intuitive editing interface. This hinders the involvement of users of non-technical backgrounds, which is often the case for many native speakers of endangered languages. As a result, involving the native community in improving and approving the recorded information in the dictionaries is not possible. Ve’rdd is built to tackle this issue while granting users and language experts the ability to contribute to different aspects of the knowledge of such endangered languages. Additionally, Ve’rdd makes different and scattered lexicographic resources in the system available for researchers and non-academic dictionary users alike. Figure 1 shows the infrastructure of our open dictionary on a high-level of abstraction showing how different users can interact with it, revealing the interplay of the two systems: Ve’rdd and Akusanat.

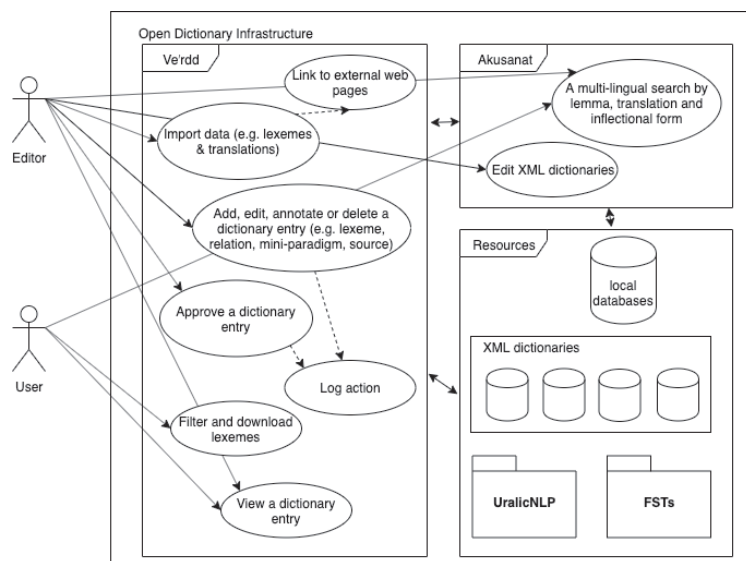


Figure 1. A UML diagram illustrating use-cases of the infrastructure

The Akusanat dictionaries offer a distinct presentation of synchronized data shared with the Giella (Giellatekno, Divvun) infrastructure. Like the Giella dictionaries [Moshagen 2013; Rueter 2013], Akusanat utilizes HFST-based [Lindén et al 2013] finite-state transducers but with an open-source python library (UralicNLP [Hämäläinen 2019]) in the search field, which allows users the option of entering virtually any word form to locate a possible lemma. Unlike the Giella dictionaries, however, Akusanat provides language internal links to associate words with derivational stems as well as external links to translations and cognates in other language dictionaries within Akusanat and entirely independent databases outside the domain.

The lexicographic data of Akusanat originates from the XML-based dictionaries in the Giellatekno infrastructure. Akusanat provides a user-friendly way of accessing the lexicographic data both as a regular dictionary user and as a dictionary editor solving the XML bottleneck. This means that, unlike XML, the lexicographic data can be edited simultaneously by multiple users. All the edits done in the Mediawiki-based Akusanat environment are synchronized with the XMLs residing in the Giellatekno infrastructure [Hämäläinen, Rueter 2019]. However, at the same time, also editing of original XML files is possible, as the synchronization works to both directions.

Ve’rdd is a customly developed system that fixes the shortcoming of Akusanat on the intuitu of editing, since Akusanat users must be familiar with the editing structure of the XML dictionaries. Ve’rdd stores information in an isolated database from Akusanat, which gives trusted editors the ability to perform amendments to information present in it without interfering with online dictionaries in Akusanat. Also, user experiences based on interactions with the system are continuously taken into account to facilitate the usability of the system and provide non-technical and technical users robust means of accessing and improving knowledge present in the database.

Figure 1 lists the core interactions of common users (speakers or learners of the endangered language) and editors with the system. The system supports import from XML dictionaries and CSV files. Whenever

data is imported, Ve'rdd consults multiple resources (e.g. Akusanat, UralicNLP and FSTs) to retrieve missing information such as part-of-speech, continuation lexica and mini-paradigms which ensures that imported information contains all the details present in other systems. Users and editors can then filter and order lexemes by multiple criteria (such as their language, consonance, etc).

By using Ve'rdd, editors have the ability to modify and comment on any present information in the database. To encourage the involvement of native speakers of endangered languages, especially speakers of another non-endangered language such as Russian or Finnish, the system allows approved editors with such criteria to add, edit, comment on and confirm the knowledge presented in the database. This guarantees that the information present in the system is validated and accurate. Whenever an editor performs any action (e.g. adding a lexeme or a translation), the system keeps a log which allows discovering cases of conflict and reverting back in the case of incorrect or non-verified actions are applied.

In a timely manner, Ve'rdd can then send the approved information (by authorized experts and native speakers) to Akusanat and other resources (e.g. UralicNLP and FSTs), which would then make retaining up-to-date information across multiple resources possible; hence, reducing the risk of providing inaccurate and misleading information.

Ve'rdd is already being used by the Skolt Sami dictionary editors as this is being written. Our development strategy involves direct interaction between the actual end users and designers, which has helped to address issues and features foreseen at the onset. A later goal would be to integrate Ve'rdd and Akusanat more completely into the infrastructure where morphological analyzers and other tools are being used, so that the end-user would have a natural and intuitive environment to work with the lexicon, but so that these changes would be automatically included into the newest compiler analyzer.

More work should also be done in connecting the lexicographic resources into various corpora that are openly available. There are various ways to proceed with this: the examples could be extracted automatically, the examples could be selected with references to the corpora, or the corpora could be tagged for representative examples that would be picked into dictionary.

The most important goal in the further development of Ve'rdd must, however, be further collaboration with the users. The system will be continuously improved with the received feedback, and the user base has to be widened to encompass a larger number of users in different languages included in the project.

## References

- Ernštreits V. (2019). Lexical tools for low-resource languages: A Livonian case study. In Proceedings of the eLex 2019 conference. P. 161–176.
- Hämäläinen M., & Rueter, J. M. (2019). An Open Online Dictionary for Endangered Uralic Languages. In Proceedings of the eLex 2019 conference. P. 819-830.
- Hämäläinen M., & Rueter J. M. (2018). Advances in synchronized XML-MediaWiki dictionary development in the context of endangered Uralic languages. In Proceedings of the XVIII EURALEX. P. 967–978.
- Hämäläinen M. (2019). UralicNLP: An NLP Library for Uralic Languages. Journal of open source software, 4(37), [1345].
- Hunt B., Chen E., Schreiner, S. L. R., & Schwartz L. (2019). Community lexical access for an endangered polysynthetic language: An electronic dictionary for St. Lawrence Island Yupik. In Proceedings of NAACL 2019.
- Laxström N., & Kanner A. (2015) Multilingual Semantic MediaWiki for Finno-Ugric dictionaries. In Septentrio Conference Series P. 75-86.
- Lindén K., Axelson E., Drobac S., Hardwick S., Kuokkala J., Niemi J., Pirinen T., et al. (2013). HFST a system for creating NLP tools. In International workshop on systems and frameworks for computational morphology P. 53–71.
- Moshagen S. N., Pirinen T., & Trosterud T. (2013). Building an open-source development infrastructure for language technology projects. In Proceedings of the 19th Nordic Conference of Computational Linguistics/ P. 343–352.
- Tavast A., Langemets M., Kallas J., and Koppel K. (2018) Unified Data Modelling for Presenting Lexical Data: The Case of EKILEX. In Proceedings of the XVIII EURALEX. P. 749-761.

## ЭЛЕКТРОННЫЕ РАЗРАБОТКИ ЦЕНТРА ИННОВАЦИОННЫХ ЯЗЫКОВЫХ ТЕХНОЛОГИЙ ГАУ РК «ДОМ ДРУЖБЫ НАРОДОВ РЕСПУБЛИКИ КОМИ»

В статье представлен обзор основных электронных разработок Центра инновационных языковых технологий (Республика Коми) за период с 2012 по 2019 гг., которые активно используются пользователями и дали новый толчок для развития коми языка.

**Ключевые слова:** миноритарные языки, финно-угорские языки, коми язык, электронное информационное пространство, электронные словари, электронные библиотеки, электронные образовательные ресурсы.

Существование и развитие национальных языков в наше время невозможно себе представить без их использования в информационном электронном пространстве.

Для языков малочисленных народов это имеет особенное значение, учитывая недостаточный ресурсный потенциал, который они могли бы использовать на развитие национального проекта.

Финно-угорские языки в Российской Федерации являются миноритарными, несмотря на то, что в ряде регионов они имеют статус государственных (фактически региональных) языков. На этих языках нет полноценного обучения в школах, отсутствуют независимые средства массовой информации, и, в целом, функции данных языков в общественной жизни имеют скорее орнаментальный характер.

Тем не менее, на региональных языках говорят тысячи граждан нашей многонациональной страны, и поддержка этих языков является обязанностью государства (что не всегда в полной мере осознается не только чиновниками, но и населением).

Для полноценного использования языка в различных сферах жизни общества необходимы значительные финансовые вливания, получить которые в современных условиях регионы вряд ли смогут. Однако есть реальная возможность на увеличение присутствия региональных языков в виртуальной реальности — мире компьютеров и интернета.

В 2012 году в Республике Коми начал функционировать Центр инновационных языковых технологий, частью которого стала Межрегиональная лаборатория информационной поддержки функционирования финно-угорских языков, известная под брендом FU-Lab.

Деятельность лаборатории направлена на развитие в электронном информационном пространстве трех родственных языков (коми, марийского и удмуртского), при этом приоритет, разумеется, имеет коми язык.

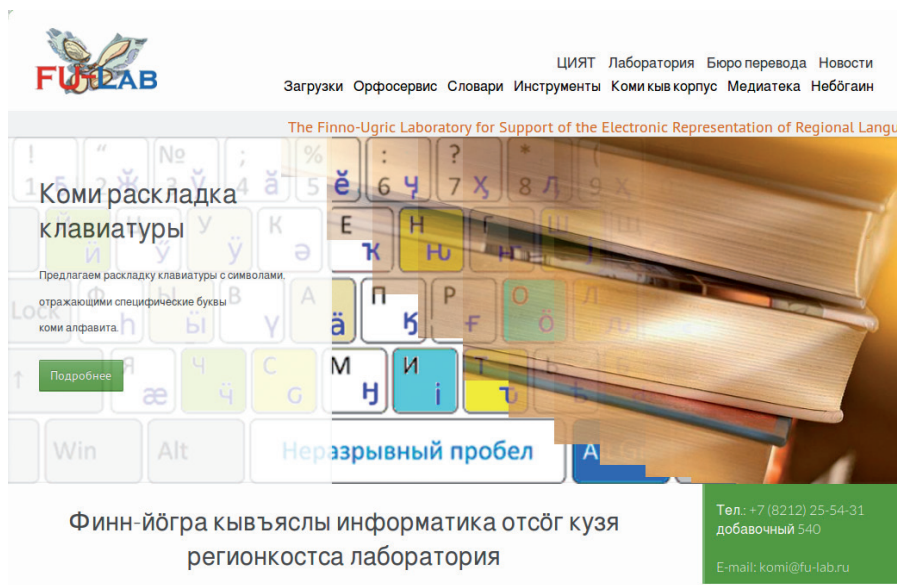


Рис.1. Начальная страница сайта лаборатории FU-Lab и Центра инновационных языковых технологий

Некоммерческий характер является отличительной особенностью работы FU-Lab. Все ресурсы, программы и инструменты, создаваемые сотрудниками лаборатории, являются бесплатными для пользователей. И это легко объяснимо, учитывая относительно небольшой круг пользователей, равно как и невысокую покупательную способность носителей региональных языков.

За минувшие с основания лаборатории шесть лет были созданы базовые инструменты и сервисы, позволяющие коми, удмуртам и марийцам применять родной язык при работе на персональном компьютере, использовании гаджетов и общении в социальных сетях, которые представлены на официальном сайте лаборатории FU-Lab и Центра инновационных языковых технологий расположенном по адресу: <http://fu-lab.ru/>.

С этого сайта можно по ссылкам перейти на все существующие электронные ресурсы по коми, марийскому и удмуртскому языкам, в частности по коми языку: коми онлайн словари, онлайн ресурсы по изучению коми языка, мультфильмы на коми языке, аудиозаписи на коми языке, коми теле- и радиопередачи, коми газеты, онлайн энциклопедии, в том числе коми википедия, коми онлайн библиотеки, комиязычные сайты государственных учреждений и других организаций, личные сайты коми учителей, коми блоги, коми группы “ВКонтакте”, финно-угорские сайты и др.

В данной статье мы освещаем основные разработки лаборатории со ссылками на соответствующие ресурсы интернета.

1. Онлайн словари коми и родственных языков.

Здесь представлены русско-коми, коми-русский, марийско-русский, удмуртско-русский, мансийско-русский и другие словари в онлайн версии. Также представлены коми-русский и русско-коми учебный словарь, специальный словарь для изучающих коми язык с морфоанализатором (возможностью нахождения слова в разных словоформах) и аудиосопровождением.

2. Сервис проверки орфографии коми и близких ему языков (<http://webspell.fu-lab.ru/>).

Сервис проверки орфографии языка – программное обеспечение, осуществляющее проверку правописания текста на выбранном языке.

3. Сервис сортировки в порядке коми, удмуртского и марийского алфавита (<http://fu-lab.ru/sortirovka>).

Данное программное обеспечение позволяет автоматически сортировать список или текст в алфавитном порядке для коми, удмуртского и марийского языков.

4. Онлайн конвертеры коми, марийских и удмуртских текстов с нестандартной кодировки на стандарт Юникод: (<http://fu-lab.ru/converter>).

5) Онлайн конвертер молодцовской графики: (<http://fu-lab.ru/convertermolodcov>). Данный сервис позволяет конвертировать коми тексты с молодцовской письменности (1920–30 гг.) на современную и обратно.

6) Коми онлайн библиотека «Коми гижӧд» (<http://komikyv.org/>).

Полнотекстовая онлайн библиотека на коми языке с возможностью поиска по различным параметрам (автор, год издания, жанр и т.д.). В разделе «переводная коми литература» представлены параллельные тексты на языке-оригинале.

7) Коми онлайн энциклопедия «Коми тӧданін» (<http://wiki.komikyv.org>).

Данная энциклопедия представляет гипертекстовую энциклопедию, включающую информацию о коми авторах (писатели, поэты, журналисты, ученые, переводчики), книжных и газетно-журнальных изданиях на коми языке с 1813 г. по 2017 г.

8) Национальный корпус коми языка (<http://komicorpora.ru/>).

Это электронная справочно-информационная система на основе аннотированного корпуса текстов литературного коми языка с возможностью поиска языковых единиц по различным параметрам. Она состоит из систематизированной базы текстов на коми языке и специальных помет к нему, необходимых для филологического анализа этих текстов.

9) Медиатека по коми диалектам (<http://videocorpora.ru>).

Это электронная справочно-информационная система на основе электронного архива аннотированных записей (видео, аудио, расшифрованные тексты на коми, в переводе на русский и английский языки) диалектных текстов коми языка с возможностью поиска языковых единиц по различным параметрам. В настоящий момент представлены тексты в основном на ижемском диалекте.

10. Материалы по коми лингвистике (<http://komikyv.com>).

Здесь представлены научные материалы по грамматике, лексике, фонетике, орфографии, диалектологии, истории языка, контактах коми языка с другими языками; научные словари и т.д. Сайт разрабатывается под эгидой Сектора языка Института языка, литературы и истории Коми научного центра УрО РАН.

11. Онлайн портал по изучению коми языка «Коми кыв». Одним из последних разработок стал портал для некоми пользователей «Коми кыв», на котором можно будет сконцентрировать обучающие ресурсы и методически обоснованно расположить их (диалоги, занимательная грамматика, лексические тренажеры, различного типа упражнения) при возможности выбора исходного и желаемого уровней владения языком.

На сайте лаборатории FU-Lab и Центра инновационных языковых технологий представлены следующие программы для скачивания:

1. Раскладки клавиатуры современных коми, удмуртского, марийского, алтайского и тувинского языков для наиболее популярных (Mac OS (Apple), Windows, Linux) операционных систем на основе кириллицы, в которых единообразно установлены коды данных букв в соответствии с международным стандартом Юникод. Эти раскладки клавиатуры удобны для набора текста на национальных и русском языках.

2. Коми оффлайн словари ([http://wiki.fu-lab.ru/index.php/Коми\\_электронные\\_словари](http://wiki.fu-lab.ru/index.php/Коми_электронные_словари)).

Коми оффлайн словари работают в оболочке программы GoldenDict. Словари коми языка поставляются вместе с рядом других словарей (удмуртский, марийский, алтайский, английский). В программу при желании можно подключать и другие словари.

3. Коми оффлайн спеллчекер (<http://fu-lab.ru/pravopisanie>).

Сервис проверки орфографии языка – программное обеспечение, осуществляющее проверку правописания. Здесь представлены программы для коми, марийского и удмуртского языков.

На сайте лаборатории FU-Lab и Центра инновационных языковых технологий представлены и учебные пособия по предметам школьного курса на коми языке, созданные в рамках международного проекта «Создание электронных учебных пособий по предметам школьного курса на родном языке». Реализованы следующие учебные пособия:

1. Электронное учебное пособие «Гулюсянь юёр» / «Весть от голубя» представляет собой образовательный продукт по коми литературе на коми языке для учащихся 5–6 классов. Автор пособия — доцент кафедры финно-угорской филологии и национального образования, руководитель НОЦ «Инновации в национальном гуманитарном образовании» Сыктывкарского государственного университета им. Питирима Сорокина, к. филол. н. Остапова Е. В.

2. Электронное учебное пособие «Роч муяслӧн важвылӧм (VIII–XIII-ӧд нэмъяс)» / «История России (VIII–XIII вв.)» представляет собой образовательный продукт по истории на коми языке для учащихся 6 классов. Автор пособия – к. ист. н., преподаватель ГПОУ «Сыктывкарский колледж сервиса и связи» Остапов Н. В.

3. Электронное учебное пособие по окружающему миру в 1 классе по программам Плешаковой А. А., Новицкой М. Ю. «Окружающий мир» – 1 класс по системе «Школа России» и Вахрушева А. А., Бурского О. В., Раутиана А. С. «Окружающий мир» – 1 класс по системе «Школа 2100» на коми языке. Авторы пособия – к. пед. н., доцент кафедры начального и дошкольного образования Института педагогики и психологии Сыктывкарского государственного университета им. Питирима Сорокина Поберезкая В. Ф., к. филол. н., доцент, зам. директора Института педагогики и психологии Сыктывкарского государственного университета им. Питирима Сорокина Терентьева С. Н., к. пед. н., доцент кафедры начального и дошкольного образования Института педагогики и психологии Сыктывкарского государственного университета им. Питирима Сорокина Габова М. А.

4. Электронное учебное пособие «Вежӧр куд» / «Золотая голова» представляет собой образовательный продукт по логике на коми языке для детей 6–7 лет. Авторы пособия — Тимерханова Н. Н., к. ф. н., доц., зав. кафедрой лингвистической типологии и лингводидактики Удмуртского государственного университета; Лекомцев А. Л., ст. препод. кафедры физической и общественной географии Удмуртского государственного университета; Федина М. С. к. ф. н., начальник Центра инновационных языковых технологий Коми республиканской академии государственной службы и управления.

5. Электронное учебное пособие «Йӧзкотыр йылысь тӧдӧм» / «Обществознание» представляет собой образовательный продукт по обществознанию на коми языке для учащихся 5–7 классов. Автор пособия – к. ист. н., преподаватель ГПОУ «Сыктывкарский колледж сервиса и связи» Остапов Н. В.

Несмотря на весьма впечатляющие результаты сотрудников Центра инновационных языковых технологий, впереди предстоит еще много работы, как по линии усовершенствования существующих продуктов, так и в плане создания новых. Так, корпус коми языка пока лишь в незначительной мере содержит газетные тексты, равно как и тексты на исторических системах коми письменности (молодцовица, коми латынь и пр.), дальнейшего усовершенствования требует и поисковая система, и вся архитектура корпуса.

Для пополнения коми онлайн библиотеки необходимо постоянно работать с правообладателями, добиваясь обеспечения возможности свободного доступа пользователей к максимальному количеству текстов на коми языке. В онлайн библиотеке желательно создание аудиоверсий к каждому тексту, а также возможность загрузки текстов в различных форматах.

Развитие коми онлайн энциклопедии может идти по нескольким направлениям. Например, развитие тематических разделов (сейчас они неактивны) могло бы привести к появлению регионального аналога Википедии. Литературный раздел в идеале следовало бы довести до статей о каждом художественном произведении (сейчас в основном существуют статьи по отдельным изданиям и по конкретным авторам).

Наряду с общими словарями и недавно открытым учебным коми словарём, желательно создание на основе имеющейся текстовой базы письменного корпуса коми языка интегрального коми-русского словаря для широкого круга коми пользователей, а также научного коми словаря для исследователей языка.

Предстоит и большая работа по внедрению региональных языков в те сферы, где они пока не представлены. Например, в образование (создание учебных приложений для мобильных устройств), что требует подключения к работе специалистов в данной области информатики.

Можно ожидать и предстоящей локализации браузеров, основных операционных систем и популярного программного обеспечения (коми, марийский и удмуртский интерфейсы), обеспечения возможности выбора в поисковых системах коми, марийского и удмуртского языка в качестве преимущественного языка поиска, создания системы автоматизированного перевода, включение и оптимизация сервиса онлайн-перевода на коми и родственные языки в Google и Яндекс переводчиках.

Реализация этих планов требует интеграции усилий всех лиц и организаций, заинтересованных в сохранении и развитии региональных языков, и, в первую очередь, самих носителей языка. Мы надеемся, что вместе мы сможем вывести коми, марийский и удмуртский языки на уровень, гарантирующий их дальнейшее устойчивое развитие в условиях быстро изменяющегося мира.

## ФУНКЦИОНИРОВАНИЕ ЛИТЕРАТУРНЫХ СООБЩЕСТВ В ГЛОБАЛЬНОЙ СЕТИ: КОММУНИКАТИВНЫЙ АСПЕКТ

В статье рассматривается функционирование литературных сообществ в глобальной сети с точки зрения коммуникаций. На основании статистических данных и опыте реализации тематических проектов исследуется процесс их взаимодействия с реальными и потенциальными целевыми аудиториями. В основе анализа лежат количественные критерии, демонстрирующие доступность и открытость рассматриваемых сообществ.

**Ключевые слова:** коммуникация, литературные сообщества, Интернет-проект, Интернет-издания, сайт, глобальная сеть, сетевое телевидение, современные технологии, целевая аудитория.

Современное общество все активнее использует телекоммуникационные технологии и возможности глобальной сети. Профессиональные сообщества, а также объединения по интересам нередко используют виртуальные площадки для общения в официальном и неофициальном форматах, решения текущих и стратегических вопросов. Обратимся к более детальному рассмотрению функционирования литературных сообществ в Интернет-пространстве с точки зрения стимулирования активности целевых аудиторий и принципов деятельности.

Эффективность передачи любого сообщения в глобальной сети, в том числе в стихотворной и (или) прозаической форме, в настоящее время все больше зависит от технологических и программных решений. Подача информации с графической точки зрения может значительно варьироваться, поскольку актуальные тенденции в дизайне требуют максимально учитывать потребности целевой аудитории, а это, в свою очередь, влечет за собой адаптацию дизайна под интуитивное восприятие пользователя, заставляет обращать внимание на эргономичность системных решений.

Временной фактор также играет важную роль. От того насколько быстро (скорость загрузки, обработка и передача данных, оперативность модерации и др.) работает система (сайт, форум, чат, социальная сеть и т.п.) во многом зависит выбор соответствующего тематического Интернет-ресурса. В данном случае, основной акцент будет сделан на литературные сообщества. Самыми многочисленными в стране общественными организациями, которые объединяют литераторов, являются Союз писателей России, Союз российских писателей, Российский союз писателей, Союз писателей XXI века. Все они имеют официальные сайты:

- <http://rospisatel.ru>, «Российский писатель» (Союз писателей России);
- <http://writers.ru> (Союз российских писателей);
- <http://rossp.ru> (Российский союз писателей);
- <http://writer21.ru> (Союз писателей XXI века).

Если традиция учреждения печатного издания (газеты, журнала, альманаха и т.д.) при литературной организации, своими корнями уходит в далекое прошлое, то наличие его полнотекстового аналога в глобальной сети, несмотря на развитие информационно-коммуникационных технологий, доступна не каждой из них. Как показывает практика, в некоторых случаях печатное издание трансформируется в сетевое издание, но это не является гарантией увеличения круга его читателей. Здесь целый ряд причин: от оптимизации поисковых запросов до времени загрузки и масштабируемости контента на мобильных устройствах.

С точки зрения коммуникации интересна позиция одного из экспертов литературного журнала для семейного чтения «День и ночь» Н. Ю. Патрихиной, владелицы арт-галереи «Рыжий кот» (г. Краснодар), которая отмечает, что Россия самая «литературоцентричная» страна. «Она в принципе такая – стремящаяся к самовыражению. Чуть ли не поголовно. Писатель у нас – он же и читатель. Других читателей теперь для писателей нет. Вот и с художниками, видимо, так: потребляют искусство те же, кто его создает». Кроме того она поднимает очень серьезный вопрос «Не таково ли вообще современное искусство?» и приводит собственное видение сложившейся ситуации: «Оно существует в виде отдельных «локусов», или, современным языком говоря, замкнутых «тусовок». Живет себе, само в себе варится, само себя опыляет, само себя потребляет». [Саввиных 2019, 3]

Если опираться на данные мониторинга, то ключевыми элементами оценки эффективности деятельности организации могут выступать качественные и количественные показатели. Перед современными читателями, реальными и потенциальными членами литературных сообществ все острее встает проблема выбора. И немаловажную роль в принятии того или иного решения играет доступ-



ность информации в глобальной сети в условиях тотального дефицита времени. Существует мнение, что «алгоритмы постепенно узнают нас лучше, чем знаем мы сами. Они уже не просто рекомендуют, они навязывают и диктуют наш выбор. Что купить, с кем жить, кого избрать – решаем не мы, бережно сохраняя иллюзию самостоятельного выбора. Тонны контента формируют наши привычки, мнения, решения. По сути, мы живем в эпоху контента» [Бачурин 2019].

Рассмотрим функционирование литературных союзов с позиции коммуникативной доступности, опираясь на количественные критерии. По данным LiveInternet.ru (статистика сайтов и ведения онлайн-дневников) по состоянию на 05.11.2019 года (за 31 день – период с 05.10 по 04.11.2019 года) в топ-5 электронных ресурсов, сегментированных по запросу «Литература» в России (из 2 859 сайтов), входят национальный портал современной поэзии «Стихи.ру» и национальный портал современной прозы «Проза.ру» Российского союза писателей (РСП). Число их посетителей за указанный период составило свыше 3 млн 140 тыс. и 2 млн 180 тыс. соответственно (табл. 1). Кроме того, следует отметить, что на «Стихи.ру» на данный момент зарегистрировано более 830 тысяч авторов и 47 млн стихов, а на «Проза.ру» почти 300 тыс. авторов и свыше 8,5 млн текстов.

Таблица 1.

### Топ-5 электронных литературных ресурсов России по данным LiveInternet.ru (на 05.11.2019 года) [Рейтинг]

Рейтинг	Ресурс	Процент	Посетителей
1	ИНТЕРНЕТ-БИБЛИОТЕКА iLibrary.ru	86%	3 869 082
2	LiveLib — социальная сеть читателей книг	87%	3 546 593
3	Стихи.ру - национальный сервер современной поэзии	71%	3 140 541
4	Проза.ру - национальный сервер современной прозы	71%	2 184 480
5	Лекции.Орг	75%	1 457 604
6	Мегабучалка	90%	1 150 570
7	Электронная библиотека БИБЛИОФОНД	85%	1 030 679
8	Электронная библиотека бесплатных книг: у нас есть все книги	69%	968 735
9	Альдебаран OCR библиотека	69%	916 683
10	Книги читать онлайн	84%	885 458
11	Сайт журнала "Мир фантастики"	74%	739 375

Использование возможностей Интернета и современных технологий – осознанный и целенаправленный процесс в деятельности РСП. На официальном сайте данной организации в разделе «Миссия» отмечено: «Письменный текст, литературное творчество и литература в целом как культурное, общественное и образовательное явление позволяют сохранять информацию, знания и накопленный человечеством бесценный опыт, а также осмыслять и развивать их путем личного участия в творческом освоении мира. Кроме того, сегодня Интернет предоставляет для всех новые возможности общения, в том числе и для литераторов – единомышленников, коллег по профессии и призванию. Значимость этих процессов трудно переоценить» [Раздел].

Особое внимание заслуживает Литклуб.TV (<http://www.litclub.tv/>) – сетевое телевидение РСП, представляющее программы с участием современных авторов. Трансляция эфира осуществляется через сеть Интернет и включает три канала: «Стихи», «Проза» и «Слово», а записи программ публикуются на Youtube (<http://youtube.com/LitclubTV>).

С точки зрения сохранения литературного наследия интересен Интернет-проект Союза писателей XXI века «Читальный зал» ([www.reading-hall.ru](http://www.reading-hall.ru)), в котором представлено более четырех десятков изданий, число которых увеличивается.

В разделе «О проекте» отмечено: «Мы объединяем в единое литературное Интернет-пространство достойные, на наш взгляд, издания. «Читальный зал» – широкий проект. Мы печатаем не только новинки толстых журналов, мы печатаем новинки тонких журналов, альманахов, газет, Интернет-изданий, небольших издательств. Нам представляется, что мир литературы должен быть многополярным. Только тогда читатель сможет составить адекватное представление о современной изящной словесности» [О проекте].

Если обратиться к статистическим данным, то «в среднем портал посещают более 33 тысяч человек в месяц. Например, 24 сентября 2019 года на портале «Читальный зал» было 1595 уникальных посетителей и 2344 просмотра» [Мальцев 2019, 1].

Для стимулирования читательской активности на официальных Интернет-ресурсах литературных сообществ применяются различные механизмы, призванные привлечь внимание целевых аудиторий. В их число входят различные дистанционные конкурсы с Интернет-голосованием, рейтинговая система участников, видео-интервью и лекции, онлайн трансляции поэтических баттлов и целый ряд других мотивационных инструментов.

Изменения, происходящие в сфере межличностных и социальных коммуникаций в целом, откладывают свой отпечаток и на функционирование литературных сообществ. Глобализация информационно-коммуникационных процессов позволяет расширить спектр вовлеченных участников, однако, приносит негативные моменты. Доступность и открытость вместе с положительными факторами может снизить качественные показатели, что может быть подтверждено или опровергнуто в ходе дальнейших исследований.

### **Литература**

Бачурин И. Выхожу из облака // Русский пионер. № 6 (93). 2019: [Электронный ресурс]. URL: <http://ruspioneer.ru/cool/m/single/6361> (дата обращения 02.11.2019).

Мальцев Ф. Новости портала «Читальный зал» // Литературные известия. № 9 (173). – С. 1. Сентябрь 2019: [Электронный ресурс]. URL: [http://litiz.ru/archive/litiz\\_2019\\_09\(173\).pdf](http://litiz.ru/archive/litiz_2019_09(173).pdf) (дата обращения 05.11.2019).

О проекте // Читальный зал: [Электронный ресурс]. URL: <http://reading-hall.ru/about.html> (дата обращения 03.11.2019).

Раздел «Миссия» // Официальный сайт Российского союза писателей: [Электронный ресурс]. URL: <https://www.rossp.ru/mission> (дата обращения 01.11.2019).

Рейтинг российских сайтов раздела «Литература» // Статистика сайтов и ведения онлайн-дневников LiveInternet.ru: [Электронный ресурс]. URL: <https://www.liveinternet.ru/rating/ru/literature/#period=month;geo=ru;group=literature> (дата обращения 05.11.2019).

Саввиных М. «Рыжий кот»: масс-культура и классика // День и ночь. Литературный журнал для семейного чтения. – Красноярск: Литера-принт, 2019. № 3. – С. 3–4.

## **МАШИННЫЙ ФОНД ЯКУТСКОГО ЯЗЫКА**

В статье описывается разработка многофункционального сервиса по хранению и обработке лингвистических данных. Основными задачами сервиса являются формирование ресурсной базы и быстрое предоставление информации для дальнейшего ускоренного развития якутского языкознания.

**Ключевые слова:** якутский язык, анализатор, корпус языка, машинный фонд, морфология.

Якутское языкознание за сравнительно короткий срок совершил огромный рывок в своем развитии, подытожив фундаментальные исследования по выявлению и систематизации основных законов таких лингвистических направлений как фонетика, морфология, лексикография, литературоведение, фольклористика. Это результат кропотливой работы огромной армии языковедов-тюркологов.

Современное развитие науки о языке требует более детального изучения материала и более тонкого анализа языковых данных. Это, в первую очередь, представляет собой работу с гигантскими объемами информации, справиться с которыми человек, даже при помощи огромного количества помощников не может в пределах разумного отрезка времени, что в условиях все убыстряющегося темпа жизни недопустимо.

Поэтому, в настоящее время, востребованными становятся разного рода методы автоматизации работы лингвиста и все более отчетливее становится потребность в систематизации всех научных работ с возможностью быстрого доступа. Использование возможностей вычислительно-аналитической мощи современных компьютеров позволяет получить инновационные результаты как в области теоретической лингвистики (получение новых знаний об устройстве языка), так и в области прикладной лингвистики. Для решения этой проблемы и упорядочивания хода работы было принято решение создать систему-оболочку включающую в себе все электронные ресурсы якутской лингвистики и в качестве “идейной основы” была взята браузерная онлайн система “Машинный фонд башкирского языка” которая представляет собой удобное консолидирование работ всех направлений башкирского языкознания [Сиразитдинов 2010; Сиразитдинов 2011].

Якутский вариант “Машинного фонда” в данное время готовится с четырьмя основными разделами: Лексика, Корпус, Анализатор, Труды. Раздел “Лексика” содержит словари с поисковиками; “Корпус” – фольклорный корпус; “Анализатор” – программа анализа отдельных слов и предложений; раздел “Труды” – сборник всех трудов ИГиИПМНС СО РАН.

Особенности развития якутского языкознания и непохожесть самого якутского языка на другие тюркские диктуют внедрения некоторых оригинальных решений. Например, “Лингвистический анализатор” разрабатываемый ИГиИПМНС СО РАН (Бочкарев В. В. – разработка алгоритма) в коллаборации с СВФУ ИЯКН (Торотоев Г. Г. – разработка матрицы) представляет собой программу, работающую по принципу логического перебора частей речи и последующего сопоставления с матрицей, составленной в соответствии с законами морфологии якутского языка. Якутский язык относится к агглютинативному типу. В этих языках словоформа образуется путем присоединения к основе в строгом порядке однозначных стандартных аффиксов; границы морфем отчетливы, фонетические изменения на стыках морфем подчиняются строгим правилам [Дыбо 2014, 20], что делает якутский язык очень удобным для компьютерного анализа, но в то же время попытки построить парадигму конкретного слова демонстрируют ее чрезвычайную сложность и многомерность, что обусловлено большим числом словоизменяющих аффиксов [Дыбо 2014, 20]. Поэтому при разработке алгоритма анализатора был выбран двухкомпонентный метод, который основывается на логической переборке частей слова и сопоставлении с матрицей. Если совпадения не обнаруживаются, то результат исправляется согласно дополнительному скрипту и начинается цикл исправлений и проверок на совпадение с матрицей. В случае, если после переборки всех возможных исправлений совпадения не обнаружатся, то выдается первый вариант с отметкой – «неизвестный аффикс». Это делается для того, чтобы программа продолжила работу. По результатам тестирования такая двойная система обеспечивает почти полное исключение ошибок вызываемых частыми случаями выпадения и слияния, а также вставными гласными. Кроме того, двойная система позволяет программе анализировать не только существующие слова, но и дает возможность генерации новых слов, т.е. иметь зачатки “искусственного интеллекта”, понимающего значение слов, что в будущем станет основой программ имитирующих мыслительные процессы.

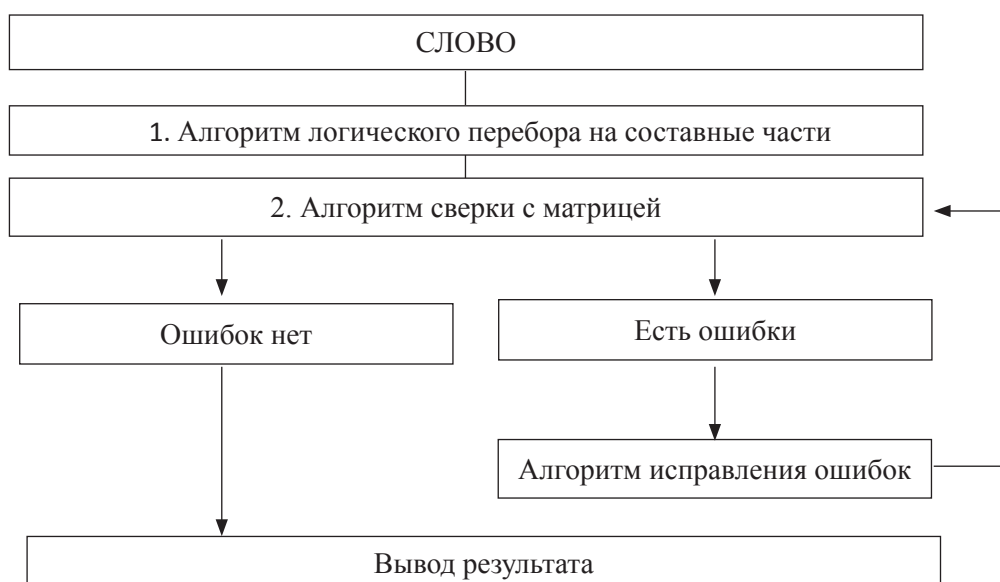


Рис.1. Блок-схема алгоритма анализатора

При аннотировании словоизменяемых морфологических показателей языка саха выделены 4 типа подхода к морфологической разметке:

1) использование тэгов Лейпцигской системы глоссирования;  
 2) заимствование тэгов, употребленных при описании структур татарских словоформ. Одним из главных условий при аннотировании языка саха является совместимость предлагаемых разметок с тэгами размеченных корпусов других тюркских языков.

3) использование разметки морфологического анализатора, разработанной для хакасского языка А. В. Дыбо.

4) использование собственной разметки. Проживание носителей якутского языка на географической периферии тюркоязычного мира способствовало развитию грамматических особенностей языка саха, например, появление третичных причастий или поздних отпричастных образований на почве самого якутского языка [Ноговицына 1998, 141].

Таким образом, результаты работы могут быть использованы другими учеными-тюркологами. На данный момент анализатор представляет собой универсальный “научный калькулятор” для лингвиста и является первой разработкой подобного рода в якутском языкознании. До сих пор все работы лингвистов производились вручную, без автоматизации, тогда как в других отраслях, например, у физиков и инженеров, используются программы для проведения расчетов и моделирования.

Основными функциями программы являются:

- Разбивка на слоги;
- Фонетический разбор;
- Морфологический разбор;
- Визуализация морфологического разбора в виде схемы;
- Нахождение этимона;
- Поиск из 15 томного “Большого толкового словаря якутского языка”.

Помимо основного предназначения анализатор будет работать как составная часть корпуса языка. Использование анализатора вместо многочисленных меток в базе данных корпуса позволяет на начальном этапе обойтись максимально упрощенной схемой без существенного урона для качества, что, в свою очередь, позволяет многократно ускорить процесс заполнения информацией. Выбор «Фольклорного корпуса» в качестве начального обусловлен наличием значительного объема уже готового текста и возможностью работать без оглядки на авторские права.

После достижения достаточного объема базы, который по подсчетам Н. А. Леонтьева составляет примерно 7 млн. словоформ [Леонтьев 2019], финишной настройкой универсального анализатора и окончательного определения нужных функций для фильтра при помощи специального скрипта, на основе цикла по количеству полученных в итоге словоформ, будет произведено автоматическое преобразование простейшей базы на стандартную с отметками, что повысит надежность, точность и скорость работы системы.

Есть возможность сделать систему таким образом, чтобы анализатор работал не при чтении информации, а при записи в базу и тем самым сразу же получить аннотированный текст. Но такой вариант плох тем, что любые вносимые изменения, такие как дополнение и удаление функций, разделов, оптимизации алгоритма работы, и другие всевозможные изменения, которых ввиду того, что корпус создается с нуля будет много, потребуют редактирования всей структуры базы с миллионами записей. Это с большой вероятностью может вызвать ошибки, поиск и исправления которых очень трудоемко. Следовательно, будет лучше, если анализатор будет работать на выводе.

Изначально такая схема организации работы выбрана для экономии человеческих, финансовых и временных ресурсов, как наименее требовательная к таковым.

Ключевые результаты по корпусу планируется связать со скриптом поиска в базе трудов сотрудников ИГИиПМНС для автоматического вывода вероятного научного обоснования. Для этого каждый вводимый в раздел “Труды” материал должен быть отмечен соответствующими тэгами. В идеальном варианте после нахождения по тэгам предполагается поиск по всему тексту материала для вывода конкретных отрывков, но с учетом низкой вычислительной мощности современных машин такой вариант отпадает из-за чрезмерно долгого времени ожидания результата.

### Литература

Дыбо А. В., Шейович А. В. Автоматический морфологический анализ для корпусов тюркских языков // Филология и культура. 2014 №2(36) – С.20–26.

Леонтьев Н.А. Вопрос о размере машинного корпуса на примере якутского языка // Электронный научно-практический журнал “Современные научные исследования и инновации”: [Электронный ресурс]. URL: <http://web.snauka.ru/issues/2015/11/58769> (дата обращения: 20.09.2019).

Ноговицына А.Н. Лингвистическое аннотирование причастий языка саха // Шестая Международная конференция по компьютерной обработке тюркских языков «TurkLang-2018». (Труды конференции) – Ташкент: Издательско-полиграфический дом «NAVOIY UNIVERSITETI», 2018.– 141 с.

Сичинава Д. В. К задаче создания корпусов русского языка // Научно-техническая информация. Серия 2: Информационные процессы и системы. 2002. №11 – С. 25–31.

Сиразитдинов З. А., Бускунбаева Л. А., Ишмухаметова А. Ш., Рафикова Л. Н. Лингвистическая информационная система «Машинный фонд башкирского языка» // Языки меньшинств в компьютерных технологиях: опыт, задачи и перспективы. – Йошкар-Ола, 2011. – С. 51-56.

Сиразитдинов З. А., Максutow А. Д., Полянин А. И., Бускунбаева Л. А. Информационная лингвистическая система «Машинный фонд башкирского языка» // Урал-Алтай: через века в будущее. Материалы IV Всероссийской научной конференции, посвященной III Всемирному курултаю башкир. – Уфа, 2010.– С. 286–290.

Торотов Г. Г., Торотова С. Г. Морфологическое аннотирование словоизменяемых категорий имени существительного в языке саха // Сравнительно-сопоставительное изучение тюркских и монгольских языков: Материалы Международной научно-практической конференции. г. Якутск, 18-19 октября 2018 г. – Якутск: Издательский дом СВФУ, 2018. – С.254.

## ТРУДНОСТИ МАШИННОГО ПЕРЕВОДА И ОБРАБОТКИ РЕЧИ

Компьютерные программы поддержки национальных языков связаны с эластичной структурой исторического времени, которое вызывает к диалектическому мышлению, сторонящемуся технического прогресса и соразмерной ему языковой и культурной традиции. Машинный перевод текста уничтожает отображение и порождение языковых структур. Только творческое, духовное мышление человека актуализирует само интеллектуальное зрение, которое освещается предметным видением мира. В работе дан анализ того, что само мышление выступает как некая магнитная стрелка развития речи. Переводчик выступает как соперник автора, а машина формально, механически, направляет автора перевода к утопически-материальному, но ещё не к открытому для творческих возможностей бытию.

**Ключевые слова:** компьютерные программы поддержки национальных языков, эластическая структура исторического времени, диалектическое мышление, языковые структуры, творческое мышление человека, проблема развития речи, переводчик как соперник автора, утопически-материальное бытие.

Актуальность настоящей работы связана с обсуждением проблемы сохранения языков в условиях техногенного развития цивилизации. Актуальность определяется развитием компьютерных программ поддержки национальных языков. В этом плане хотелось бы обсудить трудности сугубо машинного перевода текста и обработки письменной речи, что способствует развитию национальных языков, формированию научной и образовательной практики.

Машинный перевод, хотя и является до некоторой степени точным, тем не менее, в процессе данного перевода мы не в состоянии «вчувствоваться» в душевную жизнь другого человека. При этом возникает вопрос, а не подменяем ли мы самобытную жизнь своих переживаний переживаниями другого «Лица»? В этом смысле проблема машинного перевода и обработки речи связана с оправданием убеждения в реальности жизненного мира других людей, а также внешнего мира [Фришейзен-Келер 1912–1913, 332]. Главное, видимо, заключается в том, что машинный язык лишает собственное «Я», присущей ему идентичности, которое сознаёт факты личного, внутреннего опыта, переживаний со стороны иных субъектов. Машинный перевод, который должен удостоверить «раскол» между мной и чем-то отличным от меня, в принципе, направлен на интересубъективное переживание, связанное с экологическим сознанием, которое включает в себя и индивидуальные «миры», а также их «объекты». Экологическое сознание, которое мы не должны игнорировать при переводе одних смысловых содержаний на другой язык, очевидно, оказывается сопряжённым с абсолютно уникальным культурным опытом переживаний человека, в результате чего сам переводчик текста оказывается, практически, соперником его автора.

Компьютерные программы национальных языков связаны с экологической структурой исторического времени. Необходимо заметить, что время существует лишь потому, что нечто происходит. Однако данный факт недостаточно осмыслен. Машинный перевод проявляется в отношениях, которые связаны с «Что» (Was), т.е. с сущностью «в различных формах своего протекания» [Блох 1997, 157]. Но гуманный дух самой обработки машинной речи оппозиционен непосредственно переживаемому времени. Само восприятие времени, зачастую противоположно представлению о времени. Оживлённые часы «перевода» одного текста на другой язык пролетают почти мигом; пустые часы еле-еле ползут. Когда человек обращается к воспоминанию, то оживлённые часы кажутся продолжительными.

Видимо, сама структура исторического времени, вызывает к диалектическому мышлению. Сама истина, очевидно, не есть некая «отчеканенная монета» и эта истина оказывается сопряжённой с самими свидетелями диалектического мышления, которое сторонится технического и интеллектуального прогресса, который, видимо, отличается соразмерной ему языковой и культурной традиции. Однако сама диалектика всегда сопряжена с гармонией естественного и искусственного. Диалектическая мысль устремляется в этом плане сквозь возникающие противоречия в своём содержании. Однако сам человек, длительное время противоречащий себе, ещё не диалектик. Его разговоры как об искусственном, так и о трудностях перевода и обработки речи, зачастую, являются бессмысленными. С диалектическим бытием не имеет ничего общего и путаник, который в каждой вещи стремится отыскать некий крючок, «на который можно было бы повесить кривое, даже обманчивое» [Блох 1997, 207].

Машинный перевод текста, в этом плане, уничтожает отображение и порождение языковых структур. Только наблюдать языковые структуры – значит делать бессмысленными сами структуры.

Программирование языковой речи, часто связано с тем, что просто попадает на глаза. Однако прикосновение к самой природе языка «схватывает» саму творческую мысль человека, а эта мысль связана с интеллектуальным зрением, которое, в свою очередь, освещается предметным видением мира. Люди, в принципе, призваны сначала понять и «прояснить» самих себя, прежде чем устремиться к обработке речи, т.е. начать нечто новое. Люди должны осознать сами границы своей творческой способности, «чтобы не надорваться» [Блох 1997, 182].

Но возникает вопрос, а в какой степени люди могут принимать решения относительно самих себя? И здесь мы оказываемся обращёнными к самому творческому стремлению, которое находится в нас, но тоскует по продуктивной творческой жизни. Мы, ведь, живём не просто для того, чтобы жить, проявлять существующие артефакты культуры, а «мы живём» и именно в данное мгновение [Блох 1997, 229].

Сама острота мысли автора перевода не позволяет мышлению быть чем-то праздным. Но эта острота перевода, скорее, является, зачастую, ремеслом. Она прокладывает путь к чему-то одному. Значение смыслового подтекста перевода есть то, что «притягивает» и далее направляет себя к тому, что оно означает. Но сам перевод текста не является полным; ведь, духовная полнота существует там, где есть движение, а именно, движение самой основы текста. На это оказывается ориентированным лишь мышление, которое подобно «магнитной стрелке», поскольку мышление, зачастую, перехватывает сами пределы смыслов. Отсюда, сам перевод должен, в принципе, быть лаконичным. Рассудку не требуются никакие «приправы». Философская мысль, в этом плане, убедительна и поучительна.

Переводчик текста выступает, при этом, как соперник автора текста. Мысль переводчика лишь тогда его увлекает, когда мысль оказывается направленной на самую себя.

Переводчика текста озадачивает нехватка бытия в самом историческом явлении. Это означает, что творческий поиск автора перевода, стремление постичь могущество диалектики, означает изначальную загадку в каждом проживаемом мгновении перевода, в котором может быть и нечто фальшивое. Точность перевода, в этом плане, сопряжена с собственным бытием вещей и процессов. Но сам перевод, сама творческая обработка речи связаны с уплотнением социального и исторического времени. Все таинства перевода речи оказываются оваянными далью. Но изначальная загадка текста связана с понятием Родины, как дома, и в этом плане мы ощущаем саму почву «здесь» и «теперь», т.е. того, что стало нам близким и имманентным нашему самосознанию.

Переводчик, разумеется, выступает как соперник автора, тогда как сама машина формально, чисто механически, постепенно направляет автора перевода к «утопически-материальному», но, видимо, к ещё не открытому для творческих возможностей бытию. Переводчик всегда испытывает сложность гармонизации мыслей и жёстко сталкивающихся в пространстве вещей. Однако утопически-материальное существует потому, что сама материя мира ещё не закрыта для исторического исследования языка и речи. И здесь мы, очевидно, сталкиваемся с трудностью самого понятийного определения «потенциального», а это, в свою очередь, накладывает печать на саму диалектику машинного и человеческого перевода, искусственного и естественного.

### Литература

Блох Э. Тюбингенское введение в философию / Пер. с нем. Т. Ю. Быстровой, С. Е. Вершинина, Д. И. Криушова. – Екатеринбург: Изд-во Урал. ун-та, 1997. – 400 с.

Фришгейзен-Келер М. Вильгельм Дильтей как философ // Логос. Кн. 1-2. – М., 1912–1913.

## О ЛЕКСИКОГРАФИЧЕСКОЙ БАЗЕ ДАННЫХ МАШИННОГО ФОНДА БАШКИРСКОГО ЯЗЫКА

Лексикографический раздел Машинного фонда башкирского языка дает информацию о слове исходя из существующих словарей башкирского языка. В базе представлены академические и учебные словари: одноязычные и двуязычные, частотные, терминологические, синонимический, словари-справочники (названий населенных пунктов и улиц городов республики, названий водных и горных объектов Башкортостан) и другие. Лексикографический подфонд дает возможность одновременного поиска в 58 словарях.

**Ключевые слова:** башкирский язык, корпусная лингвистика, компьютерная лексикография, лингвистические базы данных, лексикография, словари.

Лексикографическая база данных Машинного фонда башкирского языка (далее МФБЛ) построена на основе изданных в разные годы печатных словарей с указанием авторов и года издания и на сегодня включает 58 словарей. Для удобства работы все они сгруппированы в шесть тематические группы:

– общефилологические двуязычные словари с башкирского языка (башкирско-английский, башкирско-русский, башкирско-турецкий);

– общефилологические двуязычные на башкирский язык русско-башкирский, англо-башкирско-русский);

– двуязычные терминологические словари (русско-башкирские словари по пчеловодству, по зоологии, по экологии и т.п.);

– одноязычные словари (толковый, фразеологический, синонимический, орфографический);

– словари-справочники (русско-башкирский словарь водных, горных объектов, улиц городов Республики Башкортостан);

– частотные словари (языка прозы Д. Юлтыя, эпоса “Урал-батыр”, по публицистике, прозе, науке).

**Двуязычные общие словари** (башкирско-русский, русско-башкирский, башкирско-английский, англо-башкирский, башкирско-турецкий). Словари данного типа имеют самые сложные строения, которые обобщены нами к структуре:

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 5,

10) иллюстративный материал в виде словосочетаний, предложений,

11) иллюстративный материал в виде фразеологизмов, пословиц и поговорок.

Дополнительные информации до перевода выделяют диалектизмы, уменьшительные и ласкательные формы, прямые и переносные значения, стили употребления (разговорное просторечие), грамматические принадлежности и формы падежного управления, разъяснения употребления слов, формы управления при глаголах, косвенные падежи личных и указательных местоимений, залоговые и другие формы глаголов и др. [БРС 1996]. Дополнительная информация 4 после русского перевода дает различные дополнительные пояснения и уточнения значений и оттенков значения при переводе, падежные управления. *Например: Ете I 1. числ. колич. семь; ете кеше семь человек; ете йыл семь лет; с аффиксом 3 л. ед. ч. поминки (у мусульман на седьмой день кончины) ...*

**Ете II диал.** 1) см. үткер; **ете бысак** острый нож; 2) см. **етеу II**.

**Двуязычные терминологические словари** имеют следующую структуру словарной статьи:

1) словарное слово,

2) показатель многозначности,

3) дополнительная информация или специальная помета,



4) перевод,

5) дополнительная информация, иллюстративный материал (иллюстративный материал может отсутствовать).

В некоторых статьях терминологических словарях наблюдается многозначие. *Например: перга* һитә кәрәз күзәнәктәрәнә тултырылып, бал менән һылап куйылған корт йәттәһе (Русско-башкирский словарь по пчеловодству, Г. Д. Зайнуллина, 2001). **конь** 1. ат шах. ат башы төшөрөлгән шахмат фигураһы 2. ат лег. атл. төрлө гимнастик күнекмәләр яһау өсөн тәғәйенләнгән күн тышлы озон бүкән; озонлого 160 см, киңлеге 35 см, бейеклеге 1,1–1,35 м; нәк уртала алынмалы тоткаһы була. (Русско-башкирский терминологический словарь по физическому воспитанию и спорту. Г. Д. Зайнуллина, 2009).

**Двуязычные словари-справочники** описываются обобщенной структурой:

1) словарное слово,

2) перевод,

3) вид объекта,

3) территория расположения (протекания),

4) этимология.

*Например:* словарное слово **Акташ**, перевод – Акташ, вид объекта – речка, территория расположения – Абзелиловском, Баймакском, Бурзянском, Зианчуринском, Кугарчинском районах. Этимология – от микропонима Акташ (ак ‘белый’ и таш ‘камень’).

**Толковый словарь башкирского литературного языка** имеет следующую структуру словарной статьи:

1) словарное слово,

2) показатель омонимии,

3) помета части речи,

4) толкование,

5) иллюстративный материал в виде словосочетаний,

6) иллюстративный материал в виде фразеологизмов, пословиц и поговорок.

*Например: бизәк* и. 1. Матурлыҡ өсөн төшөрөлгән төрлө-төрлө һүрәт. Алма бизәк. Сәскә бизәк. Өрә бизәк. Бизәк төшөрөү. 2. күсм. Йәш, матурлыҡ биргән нәмә. Кыз бала – йортка бизәк (Толковый словарь современного башкирского литературного языка. З.Г.Ураксин, 2005).

• **Синонимический словарь** также имеет структурную схему из 4 полей:

1) словарное слово,

2) информация о синонимическом ряде,

3) толкование,

4) иллюстративный материал.

*Например: ак* I. 1. күсм. саф, таза, пак, керһез. Бысраҡ, кер, тузан катышмаған. Башкалары һирәгерәк кулланыла; керһез, ак кейемгә карата ғына әйтелә. *Кулың ак булһа, йөзөң пак* (Әйтем); I. 2. күсм. саф, таза, пак, керһез. Әхлаки яктан бозолмаған, намысына тап төшмәгән, тотанаклы. Таза, пак, ак, керһез был мәғәнәлә үз-ара якыныраҡ. *Үзенең һөттән ак, һыузан пак икәнлегенә ышандырып куйы ул* (Һ. Дәүләтшина); II сал. Кар төсө; караның киреһе. Ак киң мәғәнәлә, төстө белдергән бөтә һүзәр менән кулланыла. Ак болот. Ак қағыз. Ак кайын. Ак сәс. Ак төшөү. Ак йөзлө, кара күзле, зифа буйлы, озон сәсле болгар кызы Айһылыу кайза? (М. Ғафури); III *һөт-катык*. Һөт ашамлыктары. *Ағы барзың тугы бар*. (Әйтем).

Здесь после реестрового слова дается лексика (выделена курсивом), образующая со словарным словом синонимический ряд. После идет толкование реестрового слова и его отличительных свойств от остальных синонимичных слов. Четвертое поле дает пример употребления (выделено курсивом).

**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:* меню, (ис.) -зың, -ға, -һы(на) 2. такси (ис.) -зың, -ға, таксийы.

**Частотные словари.** Изданные 3 частотные словаря по функциональным стилям башкирского языка имеют единую структуру:

1) словарное слово,

2) показатель части речи,

3) показатель омонимии,

4) частота употребления.

*Например: без А 539, без И 5* (Частотный словарь башкирского языка. Т.2 (проза) Сиразитдинов З.А., 2002).

Словари рассматриваемого раздела описаны в базах согласно принятым для групп структурам данных. При выдаче на экран пустые поля не выводятся.

Все вышерассмотренные лексикографические базы разделов МФБЯ активно используются лингвистами республики. В лексикографии сбору языкового материала всегда придается большое значение, так как чем больше иллюстративного материала, тем выше достоверность определения значения слова, тем шире сфера действия наблюдаемых закономерностей.

Главное преимущество лексикографической базы – это за считанные минуты найти интересующее нам слово по всем словарям с переводом, с толкованием, с иллюстративными материалами одновременно, что нереально в бумажном варианте. Лексикографическую базу МФБЯ можно найти на сайте Машинного фонда башкирского языка по адресу: [mfbl2.ru](http://mfbl2.ru).

Таким образом, Лексикографические данные МФБЯ сегодня представляют интегрированную автоматизированную картотеку для построения новых словарей. Сегодня сотрудники лаборатории лингвистики и информационных технологий ИИЯЛ, наряду с решением новых задач, выполняют и поддержку Машинного фонда башкирского языка.

### **Литература**

Машинный фонд башкирского языка. – URL: <http://mfbl2.ru> (дата обращения: 18.10.2019).

Сиразитдинов З.А., Бускунбаева Л.А., Ишмухаметова А.Ш., Ибрагимова А.Д. Информационные системы и базы данных башкирского языка. — Уфа, 2013. — 116 с.

## **НАЦИОНАЛЬНЫЕ ВИКИПЕДИИ РОССИИ И ПРОБЛЕМЫ ВИЗУАЛИЗАЦИИ НАЦИОНАЛЬНЫХ ЯЗЫКОВ В РУНЕТЕ**

Автор анализирует роль и значение Википедий на национальных языках России как фактор сохранения и развития национальных идентичностей, включая язык и культуру. Предполагается, что традиционная модель оффлайн культуры постепенно отмирает, а виртуальная онлайн культура становится более важным фактором развития языков и идентичностей. Автор полагает, что национальные Википедии станут стимулирующими факторами для развития национальных идентичностей нерусских групп, содействуя сохранению и визуализации национальных языков в интернете.

**Ключевые слова:** язык, национальный язык, Википедия.

Национальные языки народов современной Российской Федерации принадлежат к числу динамично меняющихся языков, но общие перспективы их развития и сохранения продолжают оставаться неопределенными, а будущее в условиях политики русификации, с одной стороны, и развитием массовой культуры общества потребления, с другой, может быть определено в лучшем случае как туманное, в худшем как катастрофическое. Новый этап в современной истории национальных языков начался в конце 1990-х годов, что стало следствием быстрого развития интернета и виртуализации культур, которые до этого момента были ограничены рамками и пределами традиционной оффлайновой книжной культуры мира национальных государств, которые существовали в условиях сохранения либроцентричной парадигмы существования и развития культур. Появление интернета, включая электронные онлайн энциклопедии, внесло значительные коррективы в развитие национальных языков, определив и изменив векторы и траектории их новейшей актуальной внутренней и внешней истории.

Одним из факторов развития и существования национальных языков в современной Российской Федерации стал проект свободной онлайн энциклопедии «Википедия». Поэтому в центре авторского внимания в этой статье будет роль Википедии, как формы актуализации идентичностей, которые отличаются от русской и альтернативны ей, в развитии национальных языков народов России.

Национальные субъекты Российской Федерации имеют право наделять языки титульных наций официальным государственным статусом. Этот статус формально закреплен в Конституциях субъектов Российской Федерации, но фактическая ситуация разительно и значительно отличается от той, которая декларируется на официальном уровне: анализ официальных сайтов органов законодательной и исполнительной власти субъектов Федерации, министерств и ведомств, высших учебных заведений свидетельствует, что страницы на национальных языках отсутствуют.

С одной стороны, государственной поддержки для сохранения и развития национальных языков явно недостаточно, в сфере образования они представлены в несравнимо меньшем объеме, чем русский. Поэтому, с другой стороны, проблемами развития и сбережения национальных языков занимаются только заинтересованные интеллектуалы, формальные и неформальные общественные организации, которые поддерживают проекты, направленные на более активное и широкое использование национальных языков субъектов Российской Федерации в интернете. Одной из форм поддержки и актуализации видимости национальных языков и стал проект Википедии на языках субъектов Российской Федерации.

Проект «Википедия» был запущен в 2001 году [Greenemeier 2007]. К 2018 году Википедия имела разделы на 301 языке. Формально «Википедия», с одной стороны, планировалась как чистая энциклопедия, а ее создатели полагали возможным избежать политики и втягивания в политические и идеологические дебаты и дискуссии. С другой стороны, политика стала важным фактором функционирования Википедии [Montellaro 2015]. На протяжении почти 20 лет своей истории Википедия успела стать фактором политической борьбы и идеологического противостояния, включая национальные проблемы, связанные с проявлением национальных идентичностей и развитием национализма [Кирчанов 2017].

Если исключить из анализа Википедии на языках бывших республик СССР, то к 2019 году Википедия была представлена на 69 языках субъектов Российской Федерации. Если добавить украинский язык, как один из государственных языков Республики Крым, то Википедия функционирует на 70 языках. Википедии на языках народов Российской Федерации крайне разнообразны по такому критерию как количество статей. Динамика развития и наполнения Википедии на национальных языках России различна [Динаев 2017; Нур 2009; Чобаев 2017; Сүзе 2010; Әмирхан 2012]. Татарская

Википедия к 2019 году содержала 87 126 статей. На 21 сентября 2019 года Чеченская Википедия содержала 244 977 статей, что позволяло отнести ее к числу тех национальных проектов, которые динамично развивались. Башкирская Википедия тоже среди тех, которые динамично развиваются: по состоянию на 21 сентября 2019 года башкирский раздел Википедии содержит 49 814 статей. чувашский раздел занимает 96 место среди всех языковых разделов Википедии, включая 42 110 статей. К 21 сентября 2019 года Осетинская Википедия содержала 11 696 статей. На 1 октября 2017 года калмыцкая Википедия содержала 2074 статьи (216 место). На 21 сентября 2019 года в Ингушской Википедии насчитывается 1041 статья, в то время как в эрзянской – 5698 статей.

Проанализируем основные проблемы и тенденции, которые существуют в сфере развития Википедии на национальных языках субъектов Российской Федерации.

Во-первых, проект «Википедия» является международным и, поэтому, национальные активисты, которые поддерживают разделы энциклопедии на языках субъектов РФ, периодически принимают участие в международных мероприятиях. Например, в 2012 году (20–21 апреля) башкирские редакторы были приглашены на международную конференцию тюркоязычных Википедий в городе Алматы (Казахстан), что актуализировало важность сотрудничества интеллектуалов из тюркских государств в сохранении национальных языков и идентичностей с использованием виртуальных ресурсов.

Во-вторых, Википедии на национальных языках стали формой политической активности и гражданской позиции. Например, 11 июля 2012 года башкирская Википедия поддержала протест Русской Википедии против Законопроекта № 89417-6, который предусматривал введение цензуры в Рунете, разместив баннер на главной странице [Schmidt, Cohen 2014]. Авторы Удмуртской Википедии предпочитают создавать статьи, сфокусированные на национальных и этнических особенностях удмуртской нации. Чеченская Википедия также могла стать фактором политической жизни: до 2012 года среди администраторов были сторонники размещения тех статей, которые актуализировали точку зрения сепаратистов.

В-третьих, Википедии на национальных языках народов Российской Федерации содействуют более широкой визуализации их идентичности, делая их более видимыми в европейском и мировом контекстах представительства других национальных языков в Интернете. Например, 1 июля 2018 года Татарская Википедия [Гатауллин 2011] по количеству статей обошла киргизскую, отодвинув её на шестьдесят восьмое место. 10 января 2018 года Башкирская Википедия по количеству статей обошла исландскую (отодвинув её на девяносто первое место), а 2 апреля 2019 года – пенджабскую, оттеснив её на девяностое место. Татарская Википедия также развивалась крайне динамично [Фэтхуллин 2018], обойдя по численности статей окситанскую, потеснив её на шестьдесят восьмое место.

В-четвертых, национальные Википедии стали формой поддержания и попыткой стимулировать как национальные идентичности, так и национальные языки. Ярким примером такой Википедии, с одной стороны, является Татарская Википедия [Шэйхетдинов 2018; Фэтхуллин 2018], которая имеет версию на кириллице и на латинице, что сближает ее с аналогичными проектами в других независимых тюркских государствах. С другой стороны, Калмыцкая Википедия стала ареной «войны переименований», связанной с трудностями перевода географических названий и выражения особенностей калмыцкого языка и специфики калмыцкой идентичности.

Подводя итоги статьи, во внимание необходимо принимать ряд факторов, которые отражают основные тенденции функционирования Википедии на национальных языках народов Российской Федерации. Во-первых, активность отдельных национальных Википедий является крайне различной: наряду с языками-лидерами (татарский, башкирский и чувашский), существуют разделы с крайне низкой активностью. Во-вторых, тематическая направленность статей в национальных Википедиях также является очень разнообразной: принимая во внимание родовые травмы проекта Википедии как виртуальной энциклопедии, вероятно, не имеет смысла ожидать от национальных Википедий особой оригинальности, потому что сам жанр энциклопедии требует унификации и фактически обуславливает ее. В-третьих, если русская Википедия пребывает в состоянии зависимости от английской версии, то национальные Википедии Российской Федерации зависят как от русской, так и английской Википедии, которые являются для них источниками статей, потому что национальные активисты переводят их на свои языки. В-четвертых, национальные Википедии на языках Российской Федерации несвободны от политических и идеологических стереотипов: в этом контексте Википедии фактически стали виртуальной сферой, где развиваются национальные идентичности, а интеллектуалы разных субъектов Федерации ведут свои «бои за историю», редактируя и корректируя статьи своих оппонентов и сторонников других точек зрения.

Таким образом, Википедии на национальных языках народов Российской Федерации успели стать культурным и интеллектуальным феноменом, а их создатели и активные пользователи играют

свою роль в развитии и сохранении национальных идентичностей. Поэтому, использование национальных языков Российской Федерации как языков Википедии делает их более видимыми в контекстах кризиса либроцентричной модели развития идентичностей. Эта культурная, интеллектуальная и политическая функция Википедии становится особенно важной в условиях, когда печатная книга на национальном языке утрачивает монополию представительства того или иного языка в сравнении с новыми техническими средствами и возможностями визуализации языковой и национальной идентичности.

Поэтому, Википедия на национальных языках, с одной стороны, актуализирует факт их существования, стимулируя дальнейшее развитие в условиях вытеснения национальных языков субъектов Российской Федерации из других форм и сфер коммуникации, которые осуществляются в традиционной оффлайновой форме. Кроме этого, каждый проект Википедии на национальном языке демонстрирует несколько тактик и стратегий, которые используют интеллектуалы и активисты для сохранения языка и идентичности – поэтому, как чувашская или татарская, как эрзянская или удмуртская национальные Википедии могут и должны стать объектами самостоятельных исследований. С другой стороны, национальные Википедии могут стать важным каналом сохранения национальных языков и идентичностей.

### Литература

Гатауллин Р. Татар интернетеның өч проекты «Белем жәүһәре». Бөтендөнья интернет-премиясенә лаек булды! (2011): [Электронный ресурс]. URL: <http://konkurs.belem.ru/node/78> (дата обращения: 15.10.2019).

Динаев А. Чеченский язык в интернете (2017): [Электронный ресурс]. URL: <http://heharho.ru/index.php/razv/133-chechenskij-yazyk-v-internete> (дата обращения: 15.10.2019).

Кирчанов М. В. От балканских крестьян к македонцам: интеллектуальная история македонского национализма. М.: Библио-Глобус, 2017. 304 с.

Нур Г. “Татнет йолдызлары”: житдилек һәм гайбәт арасында (2009): [Электронный ресурс]. URL: <https://www.azatliq.org/a/1908895.html> (дата обращения: 15.10.2019).

Сүзе Х. Башкорт Википедиясе бер меңле булды (2010): [Электронный ресурс]. URL.: <https://www.azatliq.org/a/2241952.html> (дата обращения: 15.10.2019).

Фәтхуллин Ф. Татар Википедиясен үстерергә кирәк дип саный (2018): [Электронный ресурс]. URL: <https://tatar-inform.tatar/news/2018/08/01/168876/> (дата обращения: 15.10.2019).

Фәтхуллин Ф. Телне саклауда татарча Википедия дәүләткә караганда күбрәк файда китерә (2018): [Электронный ресурс]. URL: <https://tatar-inform.tatar/news/2018/01/08/155696/?v=full> (дата обращения: 15.10.2019).

Чобаев Л. Нохчийн Википеди: Оьрсийчоьхь - шолг'аниг, дуьнент'ехь - 43-г'аниг (2017): [Электронный ресурс]. URL: <https://www.radiomarsho.com/a/chechen-wikipedia-/28749132.html?nocache=1> (дата обращения: 15.10.2019).

Шәйхетдинов Т. «Сәләт» татар Википедиясенә мәгълүмат язучу студентлар эзерли (2018): [Электронный ресурс]. URL: <https://tatar-inform.tatar/news/2018/09/27/172356/> (дата обращения: 15.10.2019).

Әмирхан З. Киләчәген кайгырткан халык үз телен һәм мәдәниятен үстерүгә акча табарга тиеш (2012): [Электронный ресурс]. URL: <https://www.azatliq.org/a/24607960.html> (дата обращения: 15.10.2019).

Greenemeier L. «Wikipedia» Good Samaritans» are on the money // Scientific American. 2007. October 19: [Электронный ресурс]. URL: <https://www.scientificamerican.com/article/good-samaritans-are-on-the-money/> (дата обращения: 15.10.2019).

Montellaro Z. How Does Political Wikipedia Stay Apolitical? // The Atlantic. – 2015. – November 18: [Электронный ресурс]. URL: <https://www.theatlantic.com/politics/archive/2015/11/how-does-political-wikipedia-stay-apolitical/450948/> (дата обращения: 15.10.2019).

Schmidt E., Cohen J. The Future of Internet Freedom // New York Times. 2014. 11 March: [Электронный ресурс]. URL: <https://www.nytimes.com/2014/03/12/opinion/the-future-of-internet-freedom.html> (дата обращения: 15.10.2019).

## ИЗ ОПЫТА СОЗДАНИЯ ЭЛЕКТРОННЫХ ВЕРСИЙ «БАШКИРСКОЙ ЭНЦИКЛОПЕДИИ» НА РУССКОМ И БАШКИРСКОМ ЯЗЫКАХ

Авторы статьи проанализировали опыт создания электронных версий «Башкирской энциклопедии» на русском и башкирском языках. Изучены особенности и возможности онлайн-энциклопедий на примере функционала Башкирской онлайн-энциклопедии.

**Ключевые слова:** опыт, онлайн-версия, Башкирская энциклопедия, функционал, гиперссылка.

*История создания электронной версии «Башкирской энциклопедии».*

В 1997 году Научное издательство «Башкирская энциклопедия» выпустило краткую энциклопедию «Башкортостан» на русском языке. Это издание признано первой национально-региональной энциклопедией, подготовленной в Российской Федерации. В 2000 году издана краткая энциклопедия «Башкортостан» на башкирском языке – первая в России энциклопедия на родном языке коренного народа российского национального региона [Латыпов 2013, 39].

В 2000 году принято решение о выпуске электронной версии краткой энциклопедии «Башкортостан» на CD-диске. Офлайн-версия была выпущена ограниченным тиражом – 1000 экз. на русском языке и 1000 экз. на башкирском. Для чтения CD-диска необходим оптический дисковод. Материалы офлайн-энциклопедии можно записать на жесткий диск ПК. Для работы с офлайн-энциклопедией доступ к сети Интернет не требуется. В электронных изданиях, выпущенных на CD, отсутствует возможность добавить новый материал, актуализировать или редактировать имеющийся материал. Эти моменты побудили изучить возможность создания онлайн-энциклопедий, лишенных перечисленных недостатков.

К 2011 году была завершена работа над «Башкирской энциклопедией» в 7 томах на русском языке [Башкирская энциклопедия 2005-2011]. В 2012 году принято решение о создании онлайн-версии многотомной «Башкирской энциклопедии» на русском языке. Был зарегистрирован домен с аналогичным именем: <http://башкирская-энциклопедия.рф>. В период подготовки был изучен функционал ряда известных офлайн- и онлайн-энциклопедий, разработан дизайн будущего сайта, определены принципы подготовки и загрузки материалов на сайт, решена проблема с корректным отображением башкирских букв в популярных интернет-браузерах.

В конце 2012 года в издательстве «Башкирская энциклопедия» был создан отдел электронных изданий в составе трех человек. Работа над русскоязычной версией сайта заняла два года. Были загружены свыше 18 тыс. статей и 4 тыс. иллюстраций. Одновременно отдел переводов готовил переводы статей на башкирский язык. В апреле 2015 года состоялась презентация онлайн-версии на русском языке, а в декабре 2018 года завершилась работа по созданию онлайн-версии на башкирском языке: в интернет были загружены материалы 7-го тома энциклопедии на башкирском языке. Сегодня «Башкирская энциклопедия» на русском и башкирском языках доступна в онлайн-режиме. За короткое время ресурс стал популярным среди пользователей интернета. Структура онлайн-энциклопедии выстроена таким образом, что наглядно демонстрирует систему знаний о Башкортостане. Организована работа по постоянному обновлению, актуализации информации, содержащейся в статьях энциклопедии [Сайтов 2017, 15]. В 2016 году начата работа по текущей актуализации материалов многотомника, планируется через каждые 5 лет осуществлять сплошную актуализацию многотомной энциклопедии. При каждой сплошной актуализации планируется увеличивать объем энциклопедии путем включения новых статей в её электронную версию. Вышеперечисленные меры по развитию электронных версий энциклопедий на русском, башкирском и английском языках позволили бы вести работу по наполнению мирового информационного пространства достоверной научно-справочной информацией о Башкортостане и башкирском народе более масштабно [Сайтов 2017, 18].

Ниже будут рассмотрены особенности и возможности онлайн-энциклопедий на примере Башкирской онлайн-энциклопедии.

*Сетевой доступ.* Несомненным достоинством всех онлайн-энциклопедий является их доступность широкому кругу пользователей из любой точки земного шара в любое время суток при подключении к сети Интернет.

*Мультимедийность.* Электронная энциклопедия может содержать текстовые, иллюстративные, анимационные, аудио- и видеоматериалы. Мультимедийный контент в значительной степени расширяет возможности представления информации, повышает наглядность и презентабельность энциклопедии.

*Интерактивность.* Электронная энциклопедия, позволяющая пользователю работать с ней в активном режиме, называется интерактивной. Интерактивный функционал обычно включает в себя интерактивное меню, интерактивные указатели, систему гиперссылок, поисковые возможности, пользовательские настройки. Интерактивный функционал онлайн-энциклопедии может включать в себя опции подписки, обратной связи, опроса пользователей, сбора статистики посещений и др.

*Неограниченный объем.* В отличие от печатного издания сайт онлайн-энциклопедии не имеет ограничений на объем и количество загруженных материалов. Поэтому в энциклопедии могут использоваться иллюстрации большего размера, лучшего качества, в большем количестве и разнообразии.

К примеру, при подготовке печатной версии «Башкирской энциклопедии» были разработаны критерии, определявшие, какие статьи могут иметь иллюстрацию. Именно по этой причине в печатной энциклопедии портреты встречаются не в каждой персональной статье. В электронной версии это ограничение может быть снято. В электронной версии можно отказаться от сокращения слов. В результате облегчается восприятие текста неподготовленным читателем, возникает возможность машинного перевода на иностранный язык.

*Возможность редактирования и актуализации.* Важным достоинством онлайн-энциклопедий является возможность редактирования и актуализации материалов, что подразумевает правку текста, замену текста или иллюстраций.

Актуализация материалов онлайн-энциклопедий является важной частью работы ГАУН РБ «Башкирская энциклопедия». При необходимости администратор интернет-сайта должен полностью заменить устаревший материал на новый, подготовленный в научных редакциях и хранящийся в электронном виде в базе издательства. Метод полной замены материалов гарантирует отсутствие привнесенных ошибок [Сайфуллин 2016, 135].

*Возможность масштабирования.* Имеется возможность добавить на сайт новую статью или иллюстрацию. Также можно исключить любой устаревший материал. Эти операции изменяют количество материалов онлайн-энциклопедии. В результате добавления новых материалов повышается содержательность и красочность энциклопедий.

*Создание списков и указателей.* Электронная версия, в отличие от печатной, может иметь множество различных указателей, повышающих ценность энциклопедии. Систематизация имеющихся материалов позволяет создать списки персоналий, предприятий, населенных пунктов, растений, животных и др.

Рассмотрим пример. Башкирская энциклопедия содержит статьи обо всех населенных пунктах Башкортостана с населением свыше 100 человек. В электронной версии имеется возможность опубликовать полный перечень населенных пунктов региона (включая исчезнувшие) с гиперссылками на уже имеющиеся материалы. Также можно опубликовать списки лиц, получивших региональные награды, почетные звания или премии, вне зависимости от того, имеется ли о них материал в энциклопедии, и актуализировать эти списки по мере появления новой информации.

*Быстрая замена материалов.* Пример 1. В электронной энциклопедии одна и та же иллюстрация может использоваться для иллюстрирования ряда статей. При замене иллюстративного материала в месте его хранения (папка с иллюстрациями) произойдет замена иллюстраций везде, где она используется. Пример 2. Одна и та же статья может быть включена в разные списки и указатели. Ее можно легко отыскать с помощью гиперссылок размещенных в других статьях. При замене заголовка статьи это изменение произойдет во всех списках, созданных с помощью модуля сортировки по критерию.

*Возможность модернизации.* При необходимости сайт онлайн-энциклопедии можно модернизировать: изменить дизайн, функционал, меню. Для этого нужно лишь использовать новые плагины или более современную версию конструктора сайтов. Онлайн-версия Башкирской энциклопедии была создана в 2012 году на базе CMS (Content Management System – система управления контентом сайта) Joomla! 2.5. В настоящее время 2-я версия не поддерживается. Для модернизации имеющегося сайта необходимо перейти на 3-ю версию.

*Рекомендация по хостингу.* Электронные версии Башкирской энциклопедии на русском и башкирском языках находятся на разных сайтах. Обе версии используют одни и те же иллюстрации, которые размещаются на отдельном сайте. Опыт использования выявил недостатки такой схемы. Многочисленные переходы по гиперссылкам с одного сайта на другой перегружают сервер хостинга и вынуждают платить за хостинг больше денег. Нет возможности одновременно просматривать башкирский и русский варианты статьи. Размещение одноименных статей (на русском и башкирском языках) на разных сайтах увеличивает время загрузки страниц на мобильные устройства и время, затрачиваемое на техническое обслуживание и администрирование сайта. Авторами рекомендуется размещать материалы многоязычных энциклопедий и иллюстрации к ним на одном сайте.

*Использование категорий.* Текстовые материалы лучше всего загружать в базу данных отдельно по каждой имеющейся категории. Все нужные категории следует создать заранее, еще до начала загрузки материалов. Использование категорий позволяет создавать различные списки и указатели автоматически за счет использования встроенных в CMS модулей.

*Использование гиперссылок.* Статьи онлайн-энциклопедии могут содержать гиперссылки как на собственные материалы, так и на материалы сторонних сайтов. Однако, следует помнить, что гиперссылка станет недействительной, если сторонний сайт закроется или исчезнет нужный материал на стороннем сайте или же он будет заменен на другой.

В текстовом материале гиперссылка используется обычно для перехода на смежную статью. Возможны два варианта: сделать гиперссылку непосредственно в тексте статьи либо разместить ее отдельно от текста (над или под текстом). Второй вариант имеет следующие достоинства: актуализация статьи может проводиться путем полной замены текста (быстро и удобно), название гиперссылки точно соответствует названию смежной статьи, число гиперссылок может быть любым. В электронной Башкирской энциклопедии этот метод используется для просмотра топографических карт, прослушивания аудиозаписей, просмотра фильмов.

*Гиперссылка на сторонний сайт.* Энциклопедическая статья, как правило, содержит краткую информацию о предмете статьи и не всегда имеет иллюстрацию. В этом случае ниже текста статьи можно прикрепить гиперссылку на сторонний сайт, содержащий более подробную и актуальную информацию. Данный подход можно использовать при описании учреждений и предприятий, имеющих собственный сайт.

*Обобщающая гиперссылка.* Список рек Башкортостана логично завершить обобщающей гиперссылкой на статью «Реки»; список редких растений — гиперссылкой на статью «Редкие растения». Подобные гиперссылки широко применяются в электронной версии «Башкирской энциклопедии».

*Использование краудсорсинга.* Современные онлайн-энциклопедии нередко создаются путем краудсорсинга. Краудсорсинг в энциклопедистике основан на принципе привлечения широкого круга лиц на добровольных началах для сбора и обработки информации с использованием инфокоммуникационных технологий. Для подтверждения полученной информации администратор проекта может потребовать от автора указать источник сведений (название книги или официального документа), предоставить фотографию или иное доказательство. Краудсорсинг удобен на этапе сбора первоначальной информации по отдельным темам.

Создание электронной версии «Башкирской энциклопедии» на русском и башкирском языках стало переходным этапом к научному portalу «Башкортостан». Как показывает опыт зарубежных энциклопедий, универсальные энциклопедии могут интегрироваться в систему начального, среднего и высшего образования.

### Литература

- Башкирская энциклопедия: в 7 т. – Уфа: Башк. энцикл., 2005–2011.
- Латыпов Б. Н. История становления и развития национально-региональных энциклопедий в России // Вопросы энциклопедистики: сб. ст. Вып. 3 / под общ. ред. У. Г. Сайтова. – Уфа: Башк. энцикл., 2013. – С. 36–42.
- Сайтов У. Г. Башкирская и российская региональная энциклопедистика: история и перспективы развития // Вопросы энциклопедистики. Вып. 6 / под общ. ред. У. Г. Сайтова. – Уфа: Башк. энцикл., 2017. – С. 15–18.
- Сайфуллин Ш. С. Из опыта создания электронной Башкирской энциклопедии // Региональные энциклопедии в современной научной инфокоммуникационной системе России: Материалы Всероссийской научно-практической конференции с международным участием (Уфа, 29-30 сентября 2016 г.). – Уфа: Башк. энцикл., 2016. – С. 129–135.



## СОЦИАЛЬНЫЕ СЕТИ КАК ИНСТРУМЕНТ ИЗУЧЕНИЯ ЯЗЫКОВ

В данной статье рассматривается образовательный потенциал социальных сетевых сервисов в процессе обучения иностранным языкам. Особое внимание уделяется основным характеристикам и особенностям социальных сетей с точки зрения их использования в образовательном процессе.

**Ключевые слова:** социальные сети, приложение, Интернет, грамматика, лексика, Ling, Livemocha, Lingualeo, ВКонтакте, Инстаграм.

Как известно, социальные сети с каждым годом набирают все большую популярность. В последнее время мы наблюдаем широкое внедрение компьютерных технологий во все сферы образовательного процесса. И одной из главных образовательных потребностей на сегодняшний день является изучение иностранных языков. К счастью, сегодня в Интернете можно ознакомиться с огромным количеством разнообразных ресурсов, выбрать для себя наиболее подходящий и изучать язык, получая удовольствие от занятий.

Разберем несколько актуальных социальных сетей и приложений, которые направлены на изучение иностранных языков.

К примеру, сеть Ling предназначена для людей, находящихся на любом этапе изучения языка. Особенностью данной сети является то, что она позволяет скачивать аудиоуроки, сохранять все выученные слова в личном кабинете, отправлять свои работы на проверку преподавателю и получать от него ответы.

Не менее популярная на сегодняшний день социальная сеть Livemocha предлагает обучение 27 языкам. В этой сети пользователи могут сами помогать друг другу в процессе обучения, применять полученные знания на практике, общаться с носителями языка.

Подобным средством изучения нового иностранного языка посредством общения является приложение Hello Talk. При регистрации пользователи определяют язык, которым они владеют, и язык, который они хотели бы изучить. Приложение удобно тем, что в нем можно отправлять голосовые записи, делиться своими видео и фото.

Одна из самых любимых пользователями сетей для изучения английского языка — Lingualeo. При регистрации пользователь должен пройти тест на определение уровня его подготовки. Вследствие этого приложение выводит индивидуальные задания, которые соответствуют характерному пользователю уровню. Главной особенностью данной сети является то, что ученик может сам составить список своих интересов и выполнять разного вида задания касательно им. Процесс обучения строится на активном использовании зрительной, слуховой и моторной памяти.

Приложение Easy ten для тех, у кого мало времени, но есть большое желание освоить английский язык. Каждый день сервис подбирает 10 новых иностранных слов, запоминает и учитывает ошибки в тестах, давая потом возможность повторить и запомнить трудные слова.

Также отличной сетью для изучения английского языка является приложение Simplex. Она направлена на изучение грамматики, улучшение восприятия на слух и обогащения лексики пользователей.

Однако вышеперечисленные социальные сети известны отнюдь не всем пользователям Интернета. Ко многим из них доступ, к сожалению, может быть закрыт, некоторые приложения занимают много памяти на телефоне или становятся платными после достижения определенного уровня, или же уровень можно продолжить только после приглашения друга и т.д. В этом случае можно использовать те же известные всем сети ВКонтакте или Инстаграм, где можно легко найти специальные паблики, группы, которые направлены на изучение иностранных языков. Более того, они позволяют делать собственные короткие видео и выкладывать их в сеть без какого-либо значительного редактирования. С их помощью можно реализовать популярные сегодня многодневные «челленджи». К примеру, записывать сторисы на иностранном языке, общаться с иностранцами и улучшать навыки говорения. Важно, чтобы эти ролики были не подготовленными, а спонтанными.

Таким образом, можно сделать вывод, что социальные сети играют огромную роль в изучении иностранного языка. Мы живем в веке информационных технологий, а это значит, что количество таких полезных сайтов и приложений будет только возрастать.

### Литература

Бурко Р. А., Терёшина Т. В. Социальные сети в современном обществе // Молодой ученый. – 2014. – №7. – С. 607–608.

## ЭЛЕКТРОННЫЕ СЛОВАРИ ФИННО-УГОРСКИХ ЯЗЫКОВ ИНСТИТУТА ЭСТОНСКОГО ЯЗЫКА

В статье дается краткий обзор развития направления электронных словарей в Институте эстонского языка с начала 1990-ых годов до сегодняшнего дня.

**Ключевые слова:** электронные словари, эстонский язык, финно-угорские языки

В настоящее время все больше словарей публикуются в виде онлайн-словаря и их никогда не печатают на бумаге. Такое развитие наблюдается и в Эстонии.

Первые онлайн-словари эстонского языка появились в интернете уже в середине 90-х годов XX в. К 2000 году на сайте Института эстонского языка (далее ИЭЯ) были опубликованы интернет-версии многих словарей, изданных в течение 1980-1990 годов, в том числе орфографический словарь, словари синонимов, антонимов, фразеологизмов, русско-эстонский и эстонско-русский, словарь сленга, терминологические словари и т.д. [Лангеметс 2000]. В то же время начали создавать словари, которые не были изданы раньше в традиционном печатанном виде. Эти словари изначально готовились для представления в сети Интернет.

В ИЭЯ для создания словарей была разработана система EeLex, которая позволила создавать словари в онлайн-режиме с подключением к Интернету. Эта система с 2019 года постепенно вытесняется системой нового поколения EKILex.

В Эстонии уже более ста лет назад (1911) был издан первый финско-эстонский словарь. Большие словари издавались и в советское время. В 2003 году издан финско-эстонский словарь объемом более 90 000 слов и в 2019 году – эстонско-финский словарь объемом более 40 000 слов. Венгерско-эстонский словарь издан в 2010 году (42 000 слов). Кроме словарей больших финно-угорских языков в Эстонии был издан (1990-2011) большой словарь вымирающего водского языка, созданный по инициативе профессора П. Аристе в 1959 году. Словарей других финно-угорских языков до 2015 года не издавалось.

С начала 90-ых годов XX в. в вузах Эстонии обучались многие представители финно-угорских народов. Их более 200 человек, больше всего марийцев и удмуртов. В 90-е годы несколько специальностей преподавалось на русском языке, сейчас таких специальностей практически не осталось и преподавание ведется либо на эстонском, либо на английском языках. Всем студентам необходимо знать эстонский язык. В середине 2000-ых годов была выдвинута идея создания словарей для студентов финно-угров и других интересующихся этими языками. Первым начали создавать эстонско-удмуртский словарь. Создателями были бывшие и нынешние студенты-удмурты эстонских вузов, всего 5 человек. Проект длился с 2007 по 2015 год, т.к. возникали трудности с финансированием проекта. Кроме Института эстонского языка проект был поддержан Программой Родственных Народов Эстонии, и Тартуским Университетом.

Над созданием эстонско-марийского словаря работали 7 человек – марийцы и один эстонец — главный редактор Свен-Эрик Соосаар. Основой для этого словаря был взят «Базовый словарь эстонского языка» (БСЭЯ) объемом 5000 словарных статей, который предназначен для изучающих эстонский язык начального и среднего уровня. В БСЭЯ выбраны слова исходя из частоты в корпусе эстонского языка и использования в повседневном общении [ЕКПС, 5]. Кроме слов базового словаря были добавлены еще более 4000 статей, в конечном итоге объем словаря составил 9800 словарных статей. Это первый словарь, куда были добавлены этимологические параллели марийских слов в эстонском языке, например слово *jõud* «сила» на марийском *вий*, которое имеет общий корень с эстонским словом *vägi*. В интернет-версии Этимологического словаря эстонского языка от этого слова идет гиперссылка к статье «*vägi*».

Следующий проект – Эстонско-эрзянский словарь создавался в течение двух лет, в 2017-2019 гг. Создателями являются 2 эрзянки, недавно защитившие докторскую (PhD) диссертацию в Таллиннском Университете, Наталия Абросимова и Наталия Ермакова. Редакторами словаря являются Ольга Ерина и Свен-Эрик Соосаар.

Эстонско-коми словарь начали создавать в марте 2019 года и прогнозируемый срок выполнения – март 2021 года. Создателем словаря стал лектор Тартуского университета Н. Кузнецов, коми-зырянин по национальности, выпускник Тартуского университета.

**Некоторые специфические термины для обыкновенных понятий  
в эстонском обществе с переводами в словарях ИЭЯ**

эстонский	русский	удмуртский	марийский	эрзянский
ID-kaart	ID-карта	ID-карта	ид-карточке	ИД-карточка, эсь са- лава тешкст марто карточка
haigekassa	больничная кас- са	-	больничный кассе	больницань касса
riigikogu	Рийгикогу, эстонский парла- мент	Риигикогу, эстон парла- мент	Эстоний кугыжаныш погын	Эстононь парламент

В начале 2010 годов в ИЭЯ при создании словарей начался переход с семасиологического принципа к ономасиологическому принципу [Langemets 2012]. Первым таким словарем является латвийско-эстонский словарь объемом более 40 000 слов. При создании этого словаря использован корпус эстонского и латвийского языков объемом более 120 миллионов слов в обоих языках.

По нашим предположениям, в будущем основным лексикографическим источником станет портал Sõnaveeb. Если раньше пользователь заходил на сайт ИЭЯ и выбирал один из 50 словарей, то на новом портале сведена воедино информация из многих словарей. Есть два варианта поиска слов. Если выбрать только эстонский язык — тогда рядом с каждым словом будет указано его значение и особенности грамматики, а если выбрать поиск по направлению эстонский-русский, то к эстонскому слову добавится и перевод. В будущем к русскому добавятся и словари на других языках.

#### Литература

- Лангеметс Маргит. Словари эстонского языка в интернете. – *Linguistica Uralica* 2000/2, – P. 143-148.  
 EKPS = Eesti keele põhisõnavara sõnastik. Eesti Keele Sihtasutus, Tallinn, 2014.  
 Langemets Margit. *Lingvistilisest kolonialismist sõnaraamatus*. – *Keel ja Kirjandus*, 2012. — P. 598-612

## О РАЗРАБОТКЕ БАЗЫ ДАННЫХ ПРАВИЛ КОНТЕКСТНО-СВОБОДНЫХ ГРАММАТИК СЕМАНТИКО-СИНТАКСИЧЕСКОГО АНАЛИЗАТОРА ТАТАРСКОГО ЯЗЫКА

В статье описывается создание семантико-синтаксического анализатора для анализа татарского предложения. Данный анализатор работает на основе правил и, соответственно, для работы требует наличия лингвистической базы данных большого объема. Результат работы анализатора представляется в виде модифицированных деревьев непосредственных составляющих, которые дополняются семантической информацией. Семантическая информация представляет собой реляционно-ситуационную модель предложения на основе которой построены правила контекстно-свободных грамматик.

**Ключевые слова:** семантико-синтаксический анализатор, татарский язык, деревья непосредственных составляющих.

В настоящее время в области обработки естественного языка наблюдается бум технологий, основанных на нейронных сетях и машинном обучении. Эти технологии пытаются использовать в любой области, где производится обработка больших объемов данных. Во многих случаях программы машинного обучения работают достаточно эффективно. Технологии нейронных сетей и машинного обучения используют также для разработки различных лингвистических анализаторов морфологического, синтаксического и семантического уровней. Особенностью этих технологий является потребность в больших объемах данных с готовой семантической и синтаксической разметкой, чтобы была возможность обучения на этом материале. Для татарского языка электронного корпуса с семантико-синтаксической разметкой на данный момент не существует, что является существенным препятствием для использования технологий машинного обучения. С другой стороны, в процессе разработки лингвопроцессоров, основанных на правилах, создается большой объем лингвистических баз данных, которые представляют ценность сами по себе и могут быть использованы в других различных задачах компьютерной обработки естественного языка. В качестве примера, поставщика лингвистических баз данных, можно привести систему машинного перевода ЭТАП. В процессе работы над этой программой был получен большой объем лингвистических ресурсов и технологий, в том числе синтаксический анализатор русского языка, который использован для синтаксической разметки корпуса РуТез. С использованием этого анализатора создан синтаксически размеченный подкорпус СинТагРус [Сизов, Валеев, Крейдлин 2008].

Семантико-синтаксический анализатор татарского предложения, описываемый в данной работе, состоит из целого ряда программных модулей, каждый из которых реализует определенный этап обработки текста (Рис.1). Соответственно, для каждого из этих этапов создается своя база данных.

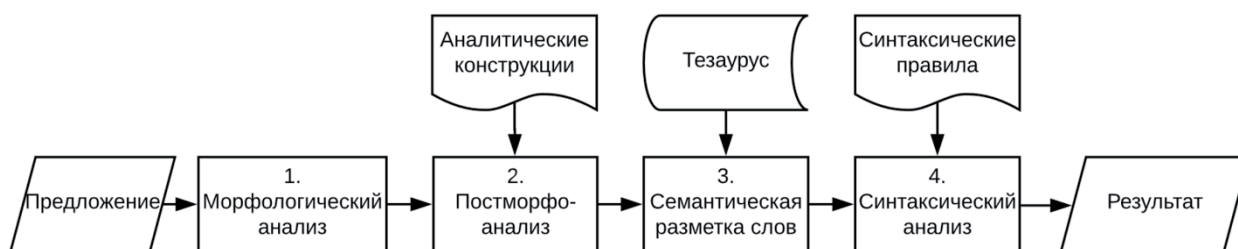


Рис.1. Последовательность работы анализатора

Модуль синтаксического анализа производит не чисто синтаксический анализ, а семантико-синтаксический анализ и в своей работе использует правила контекстно-свободных грамматик, дополненных семантической информацией [Шелманов 2015]. Семантическая информация получается на основе реляционно-ситуационной модели.

Рассмотрим фрагмент этой базы правил:

VP ->V\_Abounding\_with | V\_Absorb\_heat | V\_Abundance | V\_Activity\_finish | V\_Activity\_pause | V\_Activity\_start | V\_Activity\_stop | V\_Adding\_up | V\_Adjusting | V\_Aiming | V\_Amalgamation | V\_Amounting\_to | V\_Apply\_heat | V\_Arriving | V\_Assessing | V\_Assistance | V\_Attaching | V\_Attempt | V\_Bringing | V\_Building

**V\_Building** -> Pred\_Building |  
 Created\_entity1Pred\_Building |  
 Components Created\_entity1Pred\_Building |  
 Created\_entity2 Components Pred\_Building  
**Pred\_Building** ->V[lemma='төзө'] | V[lemma='кор'] | V[lemma='сал']  
**Components** -> N[last\_affix=ABL]  
**Created\_entity1** ->N[last\_affix=Sg]  
**Created\_entity2** ->N[last\_affix=ACC]

Первое правило показывает, что глагольная группа VP может описывать одну из перечисленных в списке ситуаций. Обозначения ситуаций приводятся в правой части правила, разделенные символом дизъюнкции.

Например: **V\_attaching** – ситуация, описывающая прикрепление одного объекта к другому, **V\_Building** – ситуация, описывающая строительство объекта.

Второе правило в примере описывает аргументно-предикатную структуру ситуации строительства V\_Building. Ситуация строительства V\_Building образуется с помощью предиката Pred\_Buildings в сочетании с аргументами предиката Created\_entityи Components. Здесь Created\_entity – актанта, обозначающий создаваемую сущность, а Components – актанта, обозначающий компоненты из которых строится новая сущность. Следует отметить такую особенность, что в зависимости от положения в предложении актанта имеет разную грамматическую структуру. Если Created\_entity расположен к предикату ближе, чем Components, то он имеет форму номинатива, а если дальше, то имеет аффикс аккузатива.

Например:

*кирпечлэрдэн йорт төзи 'из кирпичей строит дом'*  
*йортны кирпичлэрдэн төзи 'строит дом из кирпичей'*

Для выражения этих особенностей использованы два варианта актанта Created\_entity1 и Created\_entity2.

В данном примере представлены только те актанта, которые образуют ядро ситуации, кроме них в правило входят еще аргументы, не входящие в ядро, типа Место (Location) и Время (Time).

В 3-м правиле представлено, что предикат строительства может быть представлен любым из глаголов, перечисленных в правиле. В данном случае это леммы <төзө>, <кор>, <сал>.

В правилах 4 и 5 представлены ограничения для аргументов предиката, которые определяются их грамматическими формами. Так словоформа, выражающая актанта Components, стоит в форме аблатива (ABL), а словоформа, выражающая актанта **Created\_entity**, стоит в форме номинатива (Nom) или аккузатива (ACC).

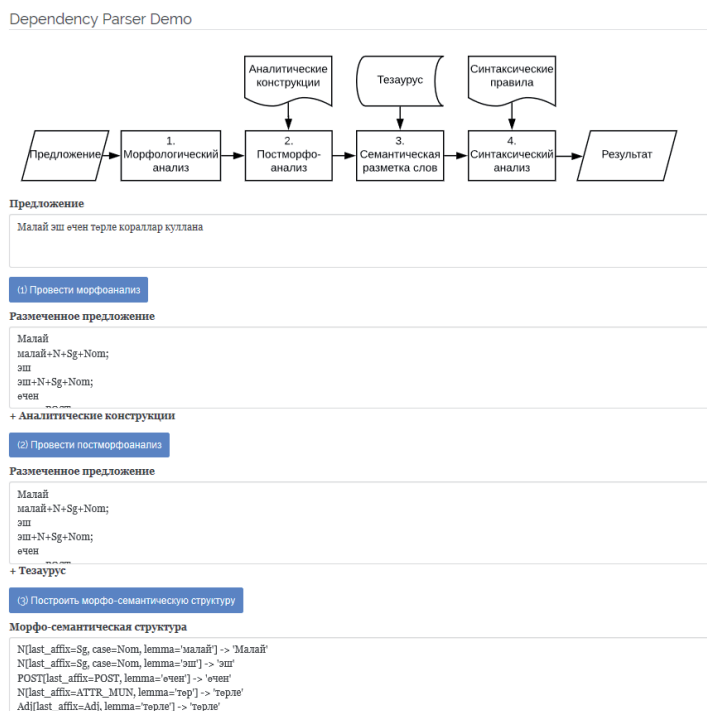


Рис.2. Интерфейс сайта для семантико-синтаксического анализа

Таким образом, учитывая, что с помощью правил описывается достаточно большое количество ситуаций, предикатов и ролей, то в результате формируется большая база правил, которая представляет собой представление ситуационных фреймов в виде правил контекстно-свободной грамматики.

Программа семантико-синтаксического анализатора реализована в форме веб-интерфейса (Рис.2). Она работает в двух режимах доступа: пользователя и эксперта. В режиме пользователя можно только запускать анализатор, подавая на вход текст для анализа. В этом режиме нет доступа к заполнению и редактированию базы данных анализатора. Режим Эксперта позволяет заполнять, редактировать и тестировать базу данных анализатора. Для получения доступа к режиму эксперта необходимо пройти аутентификацию.

Режим Эксперта реализован таким образом, что можно видеть результаты, получаемые на каждом промежуточном этапе (Рис.2): после морфологического анализа, после постморфологического анализа (сбор аналитических форм), преобразование результата в семантико-морфологическую форму и после непосредственно семантико-синтаксического анализа.

Результатом работы анализатора является дерево непосредственных составляющих, пример которого представлен на рис.3. В данном примере на вход анализатора поступает предложение на татарском языке:

*Малай эш өчен төрле кораллар куллана.* ‘Мальчик для работы использует разные инструменты’.  
На выходе получается следующая структура:

```
(S[]
-->(Agent[] (N[case=>Nom, last_affix=>Sg, lemma=>малай] Малай))
-->(VP[]
  []-->(V_Using[]
    []-->(Purpose[]
      []-->(N[case='Nom', last_affix='Sg', lemma='эш'] эш)
      []-->(POST[last_affix='POST', lemma='өчен'] өчен))
      []-->(Instrument2[]
        []-->(Adj[last_affix='Adj', lemma='төрле'] төрле)
        []-->(N[case='Nom', last_affix='PL', lemma='корал'] кораллар))
        []-->(Pred_Using[] (V[last_affix='PRES', lemma='куллан'] куллана))))
-->(EOS[] ))
```

Рис.3. Результат семантико-синтаксического анализа

Выданный анализатором результат представляет следующую актантную структуру.

Ситуация: V\_Using

Исполнитель (Agent): *Малай*

Цель (Purpose): эш өчен

Инструмент (Instrument2): төрле кораллар

Предикат (Pred\_Using): куллана

### Заключение

В статье описаны промежуточные результаты работы по созданию семантико-синтаксического анализатора для анализа простого татарского предложения. На данном этапе анализатор тестируется на предложениях с целым рядом структурных ограничений, что необходимо для отладки работы анализатора на более простых ситуациях и подтверждения выдвинутых гипотез. В дальнейшем предполагается постепенное снятие ряда ограничений.

Данная статья выполнена при поддержке Российского фонда фундаментальных исследований РФФИ 18-47-160014 «Разработка интегральной компьютерной модели и программного инструментария для семантико-синтаксического анализа татарских текстов».

### Литература

Сизов В.Г., Валеев Д.Р., Крейдлин Л.Г. Сравнение качества работы синтаксического анализатора системы ЭТАП и статистического синтаксического анализатора MaltParser на материале текстов из корпуса СинТагРус // Сборник трудов 31-ой Конференции молодых ученых и специалистов ИППИ РАН «Информационные технологии и системы (ИТиС'08). Геленджик, 2008 (27 сентября -04 октября). — С. 219-224.

Шелманов А.О. Исследование методов автоматического анализа текстов и разработка интегрированной системы семантико-синтаксического анализа: Дис. ... канд. техн. наук. – М., 2015. – 210 с.

## ПЕРСОНАЛЬНЫЕ САЙТЫ УЧИТЕЛЕЙ КОМИ ЯЗЫКА И ЛИТЕРАТУРЫ С ТОЧКИ ЗРЕНИЯ КОМИ ЛИТЕРАТУРНОГО ОБРАЗОВАНИЯ

Объектом исследования в данной работе являются сайты учителей коми языка и литературы на коми языке. Анализ содержания, структуры и дизайна ресурсов показал, что сайты находятся на стадии формирования, накопления научно-методической базы. В статье внесены предложения по расширению использования материалов на коми языке.

**Ключевые слова:** сайт, юзабилити, коми литература, учитель.

Использование информационных технологий в школьном литературном образовании может быть реализовано на уровне образовательной, психолого-педагогической и методической самоподготовки учителя; организации образовательного процесса на основе современных подходов; реализации образовательных потребностей учащихся.

В настоящее время на просторах интернета появляются ресурсы, позволяющие изучать коми литературу, как самостоятельно, так и в образовательных учреждениях. Многие учителя коми языка и литературы создают персональные сайты для связи с обучающимися и их родителями, коллегами. На данный момент, по нашим подсчетам, существуют сайты 24 учителей коми языка и литературы. Но отметим, что интернет – очень мобильное пространство, поэтому каждый день могут появляться новые ресурсы или исчезать имеющиеся.

Поскольку не во всех школах Республики Коми изучается коми литература, нами были изучены персональные сайты учителей именно коми литературы. В основу анализа персональных сайтов учителей положена разработка О. Н. Грибана «Дизайн сайта и юзабилити: практический аспект» [Грибан]. Автор статьи выделяет необходимые для сайта элементы: 1) заставка, в которой сообщается цель данного сайта; 2) навигация; 3) поисковая система. Кроме того, для удобства использования сайта информация на нем должна быть размещена крупным шрифтом не более четырех цветов.

На сайтах учителей коми языка и литературы размещены конспекты уроков и презентации к ним по творчеству коми писателей (В. Савин, Г. Юшков, Ф. Щербаков, И. Торопов, К. Жаков, А. Ванеев, В. Юхнин, Н. Куратова); авторские рабочие тетради; конспекты воспитательных мероприятий; работы учеников; переведенные на коми язык мультфильмы по проекту КРТК «Юрган»; коми песни в исполнении популярных коми артистов, личные фотографии учителей.

По разработкам на сайтах можно отметить личные литературные пристрастия учителей, авторский преподавательский стиль. Приведём два примера. Е. А. Игнатова вдохновенно пишет: «Просто обожаю произведения Нины Никитичны Куратовой, Елены Васильевны Габовой. Стихи Нины Обрезковой и Галины Бутыревой – это для меня каждый раз открытие. Неожиданное, иногда до боли точно сказанное мудрое открытие. Мне кажется что их стихи – это исповедь... Я преклоняюсь перед их талантом». Такой текст располагает к личности учителя, возникает желание познакомиться с выставленными на сайте материалами, посетить уроки. И действительно, на сайте Е. А. Игнатовой представлены урочные разработки и презентации учащихся, направленные на развитие эмоционально-чувственной сферы личности. Методические разработки А. А. Политовой позволяют говорить о сложившейся методической системе молодого учителя коми литературы. Рабочие тетради, контрольные итоговые тестовые задания интересны и своевременны, направлены на развитие познавательных способностей учащихся.

К сожалению, на некоторых сайтах мы видим недостаточно продуманный подход к представлению материалов. Так, на одном из сайтов («Методическая копилка») представлены презентации «К юбилею Нины Куратовой», «Писатели-фронтовики Великой Отечественной» на русском языке, и тут же помещены инструкции «Как сделать буклет в MicrosoftWord», «Один из способов создания анимационных фильмов», «Правила безопасного поведения в период летних каникул».

Между тем, учителю коми языка и литературы при подготовке к урокам коми литературы могут вполне доступные ресурсы. Позволим себе внести несколько продуктивных предложений. С целью популяризации коми литературы в мае 2016 года в социальной сети «ВКонтакте» нами была создана группа «Коми литература радейтысьяс» (Любители коми литературы). В эту группу выложены произведения И. А. Куратова, А. Е. Ванеева, А. Ельцовой, С. Говорухиной, Е. Афанасьевой, презентации творчества коми поэтов; фильмы о коми литературе, созданные студентами Сыктывкарского

государственного университета имени Питирима Сорокина. Кроме того, в группе есть ссылки на сайт «Ловья кыв». Самый популярный материал среди пользователей: стихотворение «Коми кыв» (Коми язык) И. А. Куратова, стихотворение «Ловья ва» (Живая вода) А. Е. Ванеева, стихотворение «Бара муна, чужан сиктöй, муна...» (Снова уезжаю, родное село) А. В. Ельцовой; видео с оригинальным исполнением французской группой «Chet Nuneta» песни «Варыш поз» на стихи В. Савина (гимн Республики Коми); фильмы студентов СГУ им. Питирима Сорокина о В. В. Пахоруковой «Кыпыд сьöлöм» (Большое сердце) и «Öнiя коми поэтъяс» (Современные коми поэты). Данный материал можно использовать как на уроках, так и во внеклассной работе.

На сегодняшний день наиболее профессионально разработанным коми литературным интернет-ресурсом, по нашему мнению, является онлайн-библиотека «Коми гижöд» / «Коми литература». В ней представлены фольклорные тексты, проза, поэзия, драма, переводы. Более полная и подробная информация дана в статье руководителя Центра инновационных языковых технологий ГАУ РК «Дом дружбы народов Республики Коми» М. С. Фединой [Федина 2017, 147–151]. Перечисленные разработки начинают широко использоваться в преподавании предметов этнокультурного направления в общеобразовательных учреждениях Республики Коми. Как отмечает руководитель научно-образовательного центра «Инновации в национальном гуманитарном образовании» СГУ им. Питирима Сорокина Е. В. Остапова, «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» [Остапова 2014, 273].

Нам видится, что данные ресурсы, широко используемые в литературно-образовательном процессе, могут найти своё место и на сайтах учителей коми языка и литературы. «Как и традиционные методы, информационные технологии, ЭОР являются средствами, с помощью которых педагог формирует картину мира у подрастающего поколения» [Остапова 2017, 145].

Итак, по «юзабилити» всех рассмотренных нами персональных сайтов учителей можно сделать вывод: не на всех имеются навигация и поисковая система; информация дана разными шрифтами и разными цветами, что затрудняет чтение текста. Можно предположить, что персональные сайты учителей коми языка и литературы находятся на стадии формирования, накопления научно-методического материала. В то же время, как нам видится, необходим качественно новый, профессиональный подход к репрезентативности деятельности учителей родного языка и литературы в интернет-пространстве.

### Литература

Грибан О.Н. Дизайн сайта и юзабилити: практический аспект: [Электронный ресурс]. URL: <http://www.griban.ru/blog/46-dizajn-sajta-i-juzabiliti-prakticheskij-aspekt.html> (дата обращения: 05.09.2019).

Остапова Е.В. Мультимедийные литературно-образовательные ресурсы на коми языке // Родной язык в современных условиях двуязычия: сборник статей / Под общей ред. Е.В. Остаповой и С.Н. Терентьевой. — Сыктывкар: ООО «Издательство «Кола», 2014. — С. 267–274.

Остапова Е.В. О внедрении в образовательный процесс электронных ресурсов по коми литературе // 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 — С.139–145.

Попова Р.П., Лудькова В.М., Остапова Е.В., Пунегова Г.В., Молчанова В.А. Кафедра коми и финно-угорской филологии Сыктывкарского государственного университета в условиях меняющейся образовательной парадигмы // Родной язык в современных условиях двуязычия: сборник статей / Под общей ред. Е.В. Остаповой и С.Н. Терентьевой. — Сыктывкар: ООО «Издательство «Кола», 2014. — С. 323–332.

Федина М.С. Коми онлайн библиотека «Коми гижöд»//Литературное наследие региона: сохранение, интеграция в цифровую среду, перспективы чтения: материалы международной научно-практической конференции (Сыктывкар, 30 окт.-1 нояб. 2017 г.) / Нац. б-ка Респ. Коми: сост. Н.В. Гурьева. — Сыктывкар, 2017. — С.147–151.



## ФУНКЦИОНИРОВАНИЕ НЕОЛОГИЗМОВ УДМУРТСКОГО ЯЗЫКА В СЕТИ ИНТЕРНЕТ

В статье рассматриваются способы образования новых слов в удмуртском языке на материале социальной сети ВКонтакте. В ходе анализа языкового материала установлено, что неологизмы созданы при помощи словосложения, суффиксации, калькирования, изменения лексического значения слов.

**Ключевые слова:** неологизмы, способы словообразования, терминология, удмуртский язык.

Развитие словарного состава современного удмуртского языка является одной из важнейших лингвистических задач. В связи с отсутствием собственно удмуртских терминов в различных сферах жизнедеятельности человека возникает необходимость создания новых слов за счет внутренних ресурсов языка.

В статье рассматриваются способы образования новых слов в удмуртском языке на материале социальной сети ВКонтакте. Данный сайт доступен более чем на 90 языках, в том числе на финно-угорских: эстонском, карельском, венгерском, финском, коми, марийском, эрзянском, удмуртском. ВКонтакте – это один из наиболее популярных и посещаемых ресурсов в интернете, поэтому изучение удмуртской версии сайта представляет большой интерес.

В ходе анализа языкового материала установлено, что самым продуктивным способом словообразования является словосложение – «образование новых слов путем объединения в одно целое двух или более основ» [Розенталь, Теленкова 1985, 306]. Словосложение в удмуртском языке, как и в других финно-угорских языках, в большинстве случаев осуществляется путем простого сложения компонентов, которые в отдельности представляют собой самостоятельные слова и при сложении не претерпевают никаких фонетических и морфологических изменений. Наиболее распространенным словообразовательным типом является присоединение существительного в основной форме к другому существительному, например: *вотэсбам* ‘сайт’ (*вотэс* ‘паутина; сеть’, *бам* ‘страница’); *интыпус* ‘местоположение’ (*инты* ‘место’, *пус* ‘знак, метка’); *ужтубат* ‘карьер’ (*уж* ‘работа; служба’, *тубат* ‘лестница’); *мутус* ‘карта’ (*му* ‘земля’, *тус* ‘облик, вид, форма, образ’); *ужкагаз* ‘документ’ (*уж* ‘работа’, *кагаз* ‘бумага’) и др.

В рассматриваемой лексике встречаются сложнопроизводные неологизмы, образованные путем сложения основ с одновременной суффиксацией, например: *кылкуэт* ‘текст’ (*кыл* ‘слово; речь’, *куэт* – сущ., образованное при помощи суффикса *-эт* от глаг. *куыны* ‘выткать, ткать’); *туспуктэм* ‘фотография’ (*тус* ‘облик, вид, форма, образ’; *пуктэм* – сущ., образованное при помощи суффикса *-эт* от глаг. *пуктыны* ‘поставить’); *тусбичет* ‘фотоальбом’ (*тус* ‘облик, вид, форма, образ’; *бичет* – сущ., образованное при помощи суффикса *-ет* от глаг. *бичаны* ‘собрать, собирать’) и др.

Имеются случаи, когда к русскоязычной основе присоединяется удмуртское слово, например: *видеосуред* ‘видео’ (*видео* < рус. *видео*, *суред* ‘картина’) и др.

Следующей продуктивной словообразовательной моделью является тип «имя прилагательное + имя существительное». В данном случае в качестве атрибута часто выступают качественные прилагательные, например: *лушкемкыл* ‘пароль’ (*лушкем* ‘тайный’, *кыл* ‘слово’, букв. ‘тайное слово’); *шонеррад* ‘правило’ (*шонер* ‘правильный, верный’, *рад* ‘ряд; порядок’); *нимазсйсьет* ‘приватность’ (*нимаз* ‘особый, отдельный; специальный’, *сйсьет* ‘фильтр’ – сущ. от глаг. *сйсьяны* ‘фильтровать’) и др.

Некоторые новые термины образованы по словообразовательной модели «наречие + имя существительное», например: *вачевераськон* ‘беседа’ (< *ваче* ‘вдвоем’, *вераськон* ‘разговор’) и др.

В отдельных случаях сложные слова образованы по типу «местоимение + имя существительное», например: *асбамы* ‘личная страница’ (*ас* ‘свой’, *бам* ‘страница’ + *-ы* – лично-притяжательный суффикс 1-го лица ед. ч.) и др.

В анализируемой лексике встречаются и сложные неологизмы, созданные по типу сочинения, представляющие собой соединение корневых слов, где оба компонента равноправны, грамматически независимы. В рамках данного структурного типа выделяются двухкомпонентные собирательные названия, например, *пи-ныл* ‘пол’ (*пи* ‘мальчик, парень’, *ныл* ‘девочка, девушка’). Данный способ образования слов непродуктивен.

При создании новых слов в удмуртском языке широко используется также суффиксальный способ словообразования – «образование новых слов путем присоединения суффикса к производящей основе» [Розенталь, Теленкова 1985, 306]. В зависимости от основы производного слова суффиксальные названия являются отглагольными и отыменными образованиями.

1) отглагольные названия: *ватсэтьёс* ‘приложения’ (< *ватсаны* ‘добавить’); *тодмостонъёс* ‘закладки’ (< *тодмостыны* ‘пометить, отметить, сделать метку’); *пуйсьськон* ‘регистрация’ (< *пуйсьськыны* ‘отметиться, отметить себя’); *пуктэтьёс* ‘настройки’ (< *пуктыны* ‘поставить’; данный неологизм чаще используется в значении ‘постановление, указ’); *чёлскон* ‘адрес; ссылка’ (< *чёлскыны* ‘сослаться, ссылаться’) и др.

2) отыменные названия: *удмуртлык* ‘удмуртскость; удмуртская ментальность’ (*удмурт* ‘удмурт; удмуртский’ + *-лык* – суффикс); *кышкыттэмлык* ‘безопасность’ (*кышкыттэм* ‘безопасный’ + *-лык* – суффикс); *кабет* ‘форма’ (на материале лексикографических трудов удмуртского языка можно проследить развитие значений слова *каб*: ‘колодка, болванка, на котором плетут лапти или шьют обувь’ [Борисов 1991, 122]; ‘1. колодка 2. форма’ [УРС 1983, 179]; ‘1. колодка 2. форма 3. фигура 4. модель’ [УРС 2008, 270] + *-ет* – суффикс); *лыдмет* ‘список’ (*лыд* ‘число, количество; счёт’ + *-мет* – суффикс, который, в отличие от других вышеприведенных аффиксов, в современном удмуртском языке непродуктивен) и др.

Следует отметить, что в настоящее время наблюдается тенденция использования новых методов суффиксального словообразования. К примеру, к удмуртским словам могут добавляться русские суффиксы, например: *чеберина* ‘красавица’ (*чебер* ‘красивый’ + *-ина* – суффикс) и др. Кроме того, удмуртские суффиксы могут присоединяться к иноязычным основам, например: *юзчи* ‘пользователь’ (< англ. *user* ‘пользователь’ + *-чи* – суффикс) и др. Данные способы пользуются большой популярностью среди молодого поколения удмуртов.

При создании отдельных неологизмов используется безаффиксный способ словообразования – «образование новых слов (имен существительных) от производящей основы (глагола или имени прилагательного) без добавления аффиксов (образующая основа становится основой существительного)» [Розенталь, Теленкова 1985, 306], например: *огин* ‘сообщество’ (*огин* – основа глагола *огинаны* ‘объединить, объединять’) и др.

Для обозначения новых понятий в современном удмуртском языке используется и калькирование – «образование слов и выражений посредством копирования, буквального (чаще поморфемного) перевода структуры иноязычных слов» [Стариченок 2008: 234], например: *сьёд лыдмет* ‘чёрный список’ (*сьёд* ‘чёрный’, *лыдмет* ‘список’ < *лыд* ‘число, количество; счёт’ + *-мет* – суффикс).

Следующим продуктивным способом является лексико-семантический способ словообразования – «образование новых слов в результате распада слова на омонимы, т. е. приобретение одной и той же лексической единицей разных значений» [Розенталь, Теленкова 1985, 305–306], например: *вотэс* ‘1. паутина 2. сеть, сетка’ и *вотэс* ‘интернет’; *пыриськись* ‘участник’ и *пыриськись* ‘пользователь’; *гожтэт* ‘1. письмо 2. записка 3. письмо, грамота’ и *гожтэт* ‘сообщение’; *борзытэм* ‘закрытый’ и *борзытэм* ‘заблокированный’: *борзытэм ватсэтьёс* ‘заблокированные приложения’ и др.

При переводе отдельных терминов с русского на удмуртский язык используется и устаревшая лексика, которая зафиксирована в ранних письменных источниках удмуртского языка, например, слово *варан* ‘служба’: *ожгар варан* ‘военная служба’ (*ожгар* ‘война; военный’; *варан* ‘служба, служение’ < *вараны* ‘dienen, erdienen, bedienen’ [Wiedemann 1880, 551] (= ‘служить, обслуживать’ – *О. Т.*); глагол образован от существительного *вар* ‘слуга, прислуга; работник’). Отдельные слова, которые в настоящее время уже не употребляются или употребляются только в отдельных говорах удмуртского языка, вновь вводятся в активный словарный фонд с новым дополнительным значением: например, слово *казь* ‘сотовый’: *казь пелепон* ‘сотовый телефон’ (*казь* ‘сеть для ловли рыбы’ [Борисов 1991, 123], ‘сеть (ловушка для зайцев)’ [УРС 1983, 181], ‘сеть (для ловли рыбы; для охоты на зайцев)’ [УРС 2008, 273]; *пелепон* < рус. *телефон*).

Следует отметить, что не все русские заимствования переведены на удмуртский язык, например: *меню* ‘меню’, *профиль* ‘профиль’, *контент* ‘контент’ и др. В отдельных русизмах при переводе произошли фонетические преобразования, например, *пелепон* ‘телефон’. Для данного наименования имеется и народная этимология: *пеле* ‘в ухо’ < *пель* ‘ухо’ + *-е* – суффикс входного падежа; *пон* – основа глагола *понины* ‘положить; вложить; приложить’. Указанная лексическая единица в сочетании с другими словами приобретает и дополнительное значение ‘мобильный’: *пелепон утялтэтьёс* ‘мобильные сервисы’ (*пелепон* ‘телефон’, *утялтэт* – сущ. от глаг. *утялтыны* ‘сберечь’ + *-ёс* – суффикс множественного числа).

Таким образом, в ходе исследования выявлено, что в образовании новой лексики в удмуртском языке самым продуктивным способом является словосложение. Некоторая часть слов образована при помощи суффиксального или лексико-семантического способов. Отдельные термины представляют собой кальки с русского языка, они образованы посредством копирования, буквального перевода структуры иноязычных слов.

### Литература

Борисов Т. К. Удмурт кыллюкам = Толковый удмуртско-русский словарь: Около 15 тыс. слов / УИИЯЛ УрО АН СССР. 2-е изд. – Ижевск, 1991. – 384 с.

Егоров А. В. Словарь терминов по информатике и информационным/компьютерным технологиям на удмуртском языке для общеобразовательных школ. – Сыктывкар–Ижевск–Йошкар-Ола–Саранск–Бадачоньтомай, 2011. – 47 с.

Розенталь Д. Э., Теленкова М. А. Словарь-справочник лингвистических терминов: Пособие для учителя. 3-е изд., испр. и доп. – М.: Просвещение, 1985. – 399 с.

Стариченок В. Д. Большой лингвистический словарь. – Ростов-на-Дону: Феникс, 2008. – 811 с.

Удмуртско-русский словарь: Ок. 35 000 слов / НИИ при Сов. Мин. Удм. АССР; Под ред. В. М. Вахрушева. – М.: Русский язык, 1983. – 592 с.

Удмуртско-русский словарь: Ок. 50 000 слов / РАН, УрО, Удм. ин-т ИЯЛ; отв. ред. Л. Е. Кириллова. – Ижевск, 2008. – 925 с.

Wiedemann F. J. Syrjänisch-deutsches Wörterbuch nebst einem wotjakisch-deutschen im Anhang und einem deutschen Register. – St. Petersburg, 1880. – XIV + 692 P.

## УПРАВЛЕНИЕ КОМПЬЮТЕРНЫМ УЗЛОМ ИНФОРМАЦИОННОГО КОНТУРА ЛИНГВИСТИЧЕСКОЙ БАЗЫ ДАННЫХ ДЛЯ ПОВЫШЕНИЯ ЕГО ЭФФЕКТИВНОСТИ

Функционирование современных высокосложных лингвистических баз данных обеспечивается специализированными информационными контурами. Эти контуры базируются на основе отдельных компьютерных узлов. Производительность обработки заданий на таких узлах должна соответствовать интенсивности входного потока. Рассматривается аналитическое моделирование выбранного компьютерного узла. На этой модели строится оптимизационная задача для определения оптимальной узловой производительности.

**Ключевые слова:** базы данных, компьютерные узлы, системы массового обслуживания.

Порождение текстов возможно на основе большого количества сведений как о семантике и синтаксисе языка, так и его лексических единицах [Метлицкая 2017]. Для представления этих сведений используются формальные языки. Такие языки обрабатываются специальными алгоритмами на основе баз данных. Автоматическое порождение текстов с применением баз данных выполняется при условии их наполненности и доступности.

Обработка естественных языков приводит к работе с текстовыми данными больших объемов [Макарич 2016]. Автоматическая обработка текстов требует временного их преобразования в абстрактное представление. На основе этого абстрактного представления и будет производиться анализ и последующий синтез текстов. Лингвистические процессоры автоматических систем используют различные структуры данных, занимающие большой объем памяти оперативной и накопителей. На этапах обработки текстов с использованием программных систем вносятся дополнительные изменения в используемые данные.

Создание лингвистических баз данных производится качественно функционирующими информационными системами, способными обрабатывать данные большого объема [Мишанкина 2013]. Формально представленная структурированная информация, которая в них может содержаться, требует для своего хранения специфических баз данных. Эти базы данных могут содержать разнородную и разнотипную информацию, что требует выделения больших объемов ресурсов.

Рост баз данных приводит к необходимости эффективного получения данных из них [Соловьев 2009]. Для эффективного получения данных созданы специальные алгоритмы и методы. В основе этих алгоритмов и методов лежит использование графовых моделей представления данных. Такие модели могут соответствовать естественным и искусственным языкам. Для обработки графовых моделей используются разнотипные количественные метрики. В частности, метрики применяются не только к рабочим данным, но и к их эталонным вариантам. Результаты сравнения отражаются на мере близости рассматриваемых данных к их теоретическим предшественникам.

Увеличение разнообразия количественных характеристик текстов приводит к увеличению времени их программной обработки [Шевелев 2004]. В результате затраченное на обработку время может, в некоторых случаях, стать нецелесообразно большим. Поэтому необходимы новые подходы для сдерживания роста времени при неумножении числа количественных характеристик. Такие подходы основаны на использовании распределенных систем управления базами данных. Использование распределенных систем позволяет повышать производительность результирующей системы. Получаемая производительность недостижима для отдельной рабочей станции.

При этом возникают задачи, связанные с управлением вычислительной мощностью отдельных компьютерных узлов, играющих роль рабочих станций. Производительность обработки заданий на них должна соответствовать текущей интенсивности входного узлового потока.

Для этого можно решать оптимизационные задачи. Чтобы получить оптимизационную задачу, следует, в первую очередь, построить модель. Для компьютерных узлов часто применяют модели систем массового обслуживания (СМО) [Клейнрок 1979; Гнеденко 1966; Гмурман 1972; Вентцель 1969; Вишневский 2004; Новиков 1969]. Для оптимизационных задач можно использовать аналитические модели СМО. В рассматриваемой информационно-поточной ситуации для компьютерного узла используется модель СМО типа М/М/1.

В СМО типа М/М/1 имеется входной поток заявок интенсивности  $\lambda$ , буфер неограниченной емкости, канал обработки заявок производительности  $\mu$ . Аналитическое моделирование СМО типа М/М/1 выполняется по соотношениям:

$$\begin{aligned}
 \rho &= \frac{\lambda}{\mu}, \\
 p_0 &= 1 - \rho, \\
 p_i &= \rho^i p_0, i = 1, 2, \dots, \\
 L_q &= \frac{\rho^2}{1 - \rho}, \\
 L_s &= \frac{\rho}{1 - \rho}, \\
 T_q &= \frac{\rho}{\mu(1 - \rho)}, \\
 T_s &= \frac{1}{\mu(1 - \rho)}.
 \end{aligned} \tag{1}$$

В (1)  $\rho$  – загрузка,  $p_0$  – вероятность простоя,  $p_i$  – вероятность наличия в системе  $i$  заявок,  $L_q$  – среднее число заявок в очереди,  $L_s$  – среднее число заявок в системе,  $T_q$  – среднее время пребывания заявки в очереди,  $T_s$  – среднее время пребывания заявки в системе.

Для построения оптимизационной задачи следует отметить, что производительность компьютерного узла может соответствовать требуемой интенсивности входного потока заявок, если вероятность простоя  $p_0$  и среднее время пребывания заявки в системе  $T_s$  минимальны. Тогда целевая функция определяется в виде:

$$F_{\square}(\lambda, \mu) = C_0 p_0(\lambda, \mu) + C_s T_s(\lambda, \mu) = \frac{\mu C_s + (\mu^2 - 2\lambda\mu + \lambda^2)C_0}{\mu^2 - \lambda\mu}. \tag{2}$$

В (2)  $C_0$  – оценка затрат на простой компьютерного узла,  $C_s$  – оценка затрат на пребывание заявки в системе. Тогда оптимизационная задача может быть записана следующим образом:

$$\arg \min_{\mu} F_{\square}(\lambda, \mu) = \arg \min_{\mu} \left[ \frac{\mu C_s + (\mu^2 - 2\lambda\mu + \lambda^2)C_0}{\mu^2 - \lambda\mu} \right]. \tag{3}$$

Оптимальное значение производительности  $\mu$  в задаче (3) находится из решения уравнения

$$\frac{\partial F_{\square}(\lambda, \mu)}{\partial \mu} = 0:$$

$$\frac{\partial F_{\square}(\lambda, \mu)}{\partial \mu} = - \frac{\mu^2 C_s + (-\lambda\mu^2 + 2\lambda^2\mu - \lambda^3)C_0}{\mu^4 - 2\lambda\mu^3 + \lambda^2\mu^2} = 0. \tag{4}$$

Единственным решением (4), удовлетворяющим (1), является:

$$\mu = \frac{\lambda\sqrt{\lambda C_0 C_s} - \lambda^2 C_0}{C_s - \lambda C_0}. \tag{5}$$

Численная оценка оптимальной производительности  $\mu$ , выполненная по (5), может быть после расчетов для заданной текущей интенсивности входного потока  $\lambda$  установлена на компьютерном узле.

Полученное решение оптимизационной задачи позволяет находить оптимальную производительность компьютерного узла. Аналитическое моделирование выбранного компьютерного узла

подходит для оценки производительности обработки узловых заданий в зависимости от изменения интенсивности входного потока. Формируемые на основе таких отдельных компьютерных узлов, подвергшихся аналитическому моделированию, информационные контуры обеспечивают функционирование современных высокосложных лингвистических баз данных.

#### Литература

- Вентцель Е. С. Теория вероятностей. – М.: Наука, 1969. – 576 с.
- Вишневский В. М. Теоретические основы проектирования компьютерных сетей. – М.: Техносфера, 2004. – 512 с.
- Гмурман В. Е. Теория вероятностей и математическая статистика. – М.: Высш. школа, 1972. – 368 с.
- Гнеденко Б. В. Введение в теорию массового обслуживания / Б. В. Гнеденко, И. Н. Коваленко. – М.: Наука, 1966. – 432 с.
- Клейнрок Л. Вычислительные системы с очередями. – М.: Мир, 1979. – 600 с.
- Макарич М. В. Реализация обучающей компьютерной программы на основе лингвистической базы данных автоматической системы обработки текста // Системный анализ и прикладная информатика, № 1, 2016. – С. 78–83.
- Метлицкая Н. А. Лингвистическая база данных системы автоматического порождения англоязычного рекламного текста // Системный анализ и прикладная информатика, №2, 2017. – С. 62–67.
- Мишанкина Н. А. Базы данных в лингвистических исследованиях // Вопросы лексикографии, №1 (3), 2013. С. 25–33.
- Новиков О. А. Прикладные вопросы теории массового обслуживания. – М.: Советское радио, 1969. – 398 с.
- Соловьев В. Д. Преобразование метрик, используемых в методах кластеризации для построения филогенетических деревьев языков // Ученые записки Казанского университета. Серия “Физико-математические науки”, т. 151, №3, 2009. – С. 229–239.
- Шевелев О. Г. Представление набора текстов в реляционной базе данных для целей лингвистического анализа // Вестник Томского государственного университета, № 284, 2004. – С. 222–226.

## **ОБЗОР ПЕРВОЙ НЕЙРОСЕТЕВОЙ СИСТЕМЫ РУССКО-ТАТАРСКОГО МАШИННОГО ПЕРЕВОДА**

В статье представлена первая система русско-татарского машинного перевода, построенная на основе нейросетевых алгоритмов. Описываются подготовительные этапы обучения на корпусах параллельных текстов. Приводится сравнение качества работы построенной системы машинного перевода с переводчиком Яндекс. Ставится задача по дальнейшему повышению качества перевода.

**Ключевые слова:** машинный перевод, нейросетевой перевод, русско-татарский машинный перевод.

**SecondAuthorb, ThirdAuthora,b,,P.A. Гильмуллин, Д.Ш. Сулейманов**

### **1. Введение**

2016 год стал годом, когда системы машинного перевода, основанные на нейронных сетях, превазошли по качеству работы систем, основанные на статистике (phrase- andsyntax-based) [Bojar 2016]. С тех пор многие компании разработали нейронные версии своих переводчиков для самых популярных языковых пар [Bojar 2017; Wu 2016]. Более того, большое количество исследований было посвящено улучшению качества перевода за счет использования лингвистически мотивированных или лингвистически информированных моделей, что привело, например, к использованию многофакторных моделей и морфем в качестве базовых единиц перевода (subwordunits).

Однако, как и в других областях искусственного интеллекта, например, при построении систем распознавания речи или диалоговых систем, использование современных методов машинного обучения для класса малоресурсных языков ограничено отсутствием необходимого объема обучающих данных. Даже компании с относительно неограниченным доступом к данным (например, Google, Yandex) используют различные методы, чтобы обойти это ограничение: объединение различных подходов к переводу и выбор лучшего результата [One model is better than two 2017], использование промежуточных языков при переводе (например, английского или родственного языка с большим количеством доступных данных).

С целью создания системы машинного перевода, способной демонстрировать хорошее качество работы для татарско-русской малоресурсной языковой пары, мы использовали последние достижения в области машинного обучения. Результирующая система включает в себя инструменты, которые позволяют дополнить обучающие данные (dataaugmentation), выполняют алгоритмы предварительной обработки текстов совместно с алгоритмом перевода на основе внимания (attention-basedencoder-decoder).

В разделе 2 данной статьи представлен обзор процесса сбора параллельного корпуса, в разделе 3 описаны основные характеристики нейросетевой модели, в разделе 4 приводятся результаты проведенного эксперимента, раздел 5 содержит заключение.

### **2. Подготовка данных**

Нейросетевой подход к построению систем машинного перевода подтвердил качество своей работы в экспериментах для многих языковых пар. Большинство аспектов перевода успешно моделируются нейросетями, при этом ключевое значение имеет репрезентативный параллельный текстовый корпус, поскольку системы нейросетевого машинного перевода, как известно, недостаточно эффективны при обучении на ограниченном наборе данных [Fadaee 2017; Wu 2016]. Таким образом, для решения задачи построения системы татарско-русского перевода необходимо создание достаточно большого параллельного корпуса. Ограничением здесь является небольшой объем источников, на основе которых он может быть построен.

Идея данной статьи заключается в разработке инструментов, которые позволят использовать максимум параллельных данных, доступных для татарского языка. Одним из основных источников двуязычной информации являются веб-сайты министерств и других государственных ведомств. В Республике Татарстан действует закон, обязывающий организации вести документооборот одновременно на русском и татарском языках. Другим источником данных являются литературные произведения: печатные книги с доступным переводом. Чтобы загрузить данные из веб-источников, мы разработали программу, которая может быть настроена для загрузки информации на основе списка сайтов

и конкретных правил, которые помогают определить соответствие между русскими и татарскими страницами (например, шаблоны url, ссылки на исходной странице). Доступные книги, для которых был доступен перевод, были отсканированы с использованием профессионального сканирующего оборудования.

Собранные данные были отфильтрованы в соответствии со следующими критериями: как исходное, так и переведенное предложения должны содержать от 1 до 80 слов; повторяющиеся предложения были удалены; все собранные тексты были выровнены с помощью инструмента ABBYY Aligner 2.0 [ABBYY Aligner 2.0].

Мы также провели ручную корректировку результатов процедуры автоматического выравнивания (литературный перевод книг, например, приводил к наличию пар предложений, которые сильно отличаются друг от друга). Данная работа была завершена примерно в течение 1 месяца.

По результатам проведенной работы было собрано 328 тысяч параллельных татарско-русских предложений. Данный объем параллельного корпуса использовался для обучения первой версии нейросетевого переводчика с русского на татарский язык.

Команда переводчиков осуществляла перевод текстов новостной тематики с русского на татарский язык. Процесс был организован с использованием инструмента Smart CAT для профессиональных переводчиков (Smart CAT tool for professional translators). Ручной перевод 35 тысяч предложений занял около 700 человеко-часов или 1 месяц работы команды. В течение данного месяца было завершено обучение промежуточных нейронных моделей для русско-татарского направления перевода. С помощью промежуточной системы готовились автоматические переводы всех новых текстов, которые поступали переводчикам для ручной корректировки. Это позволило существенно ускорить процесс перевода: суммарно за 2 месяца работы общее количество ручных переводов составило 189 689 предложений. Итоговый параллельный корпус объемом 517 тысяч предложений был использован для построения финальной версии русско-татарских моделей.

При обучении моделей для обратного, татарско-русского направления перевода, был применен back-translation подход для дополнительного увеличения объема обучающих данных. В его основе лежит предположение, что большое количество (до размера исходного корпуса) автоматически переведенных предложений способно повысить качество работы системы [Sennrich 2016]. Таким образом, с помощью созданного нами русско-татарского переводчика были подготовлены дополнительно 409 тысяч пар параллельных предложений, которые вошли в состав обучающего корпуса.

Основные этапы подготовки параллельного корпуса:

- Формирование базового корпуса на основе Интернет ресурсов и литературных произведений;
- Фильтрация, автоматическое выравнивание и ручная корректировка результатов (328 тысяч пар предложений);
- Построение промежуточных систем русско-татарского переводчика;
- Ручной и полуавтоматический перевод новостных текстов на русском языке (189 тысяч пар предложений);
- Построение финальной версии русско-татарского переводчика;
- Расширение обучающего корпуса за счет back-translated алгоритма: с помощью русско-татарского переводчика было автоматически подготовлено дополнительно 409 тысяч пар параллельных предложений;
- Построение финальной версии татарско-русского переводчика.

### **3. Описание архитектуры системы**

Для обучения системы нейросетевого машинного перевода мы использовали инструментарий Nematus [Open-source neural machine translation in Theano, 2017] с усовершенствованиями, предложенными в [Sennrich 2017]. В основе подхода лежит архитектура сети encoder-decoder-attention, каждая часть которой представляет собой одно- (для случая декодера с механизмом внимания) или двунаправленную (для энкодера) рекуррентную нейросеть. Размерность слоя векторного представления слов была выбрана равной 512, размерность скрытых слоёв – 1000, значение параметра dropout–0.2. Размер словаря установлен равным 15000, размера группировки batch–60 для этапа обучения и 5 для этапа валидации. Обучение нейросети проходило с помощью оптимизатора Adam (значение параметра learningrate – 0.0001).

Татарский язык является агглютинативным языком с богатой морфологией, что требует решения проблемы большого количества внесловарных слов (OOVproblem) из-за ограниченного размера словаря и данных для обучения. Чтобы преодолеть эту проблему, мы использовали базовые единицы, построенные на основе алгоритма BPE (byte-pair encoding) [Sennrich 2015]. Модель разбиения слов на составляющие части была применена к объединенному русско-татарскому корпусу.



#### 4. Эксперимент

Мы использовали метрику BLEU [Papineni 2002] для сравнения качества систем машинного перевода. Несмотря на то, что было показано, что значения BLEU не всегда коррелируют с ручной оценкой качества перевода [Baia 2009], данный критерий по-прежнему широко используется, поскольку ручное тестирование является дорогостоящим и требует большого количества времени.

Мы провели сравнение качества работы построенной системы машинного перевода и Яндекс.Переводчика [Yandex translate 2019], поскольку они являются единственными доступными инструментами для русско-татарской языковой пары. Результаты оценки качества перевода представлены в таблице 1.

Для проведения тестирования мы случайным образом отобрали 1000 русских предложений, которые были вручную переведены; данные предложения не вошли в состав обучающей и валидационной выборок.

Таблица 1.

#### Результаты сравнения качества работы систем машинного перевода

Система	Направление перевода	Обучающий корпус	BLEU	BLEU-1	BLEU-2	BLEU-3	BLEU-4
Яндекс Переводчик	РУ-ТТ	Н/Д	12.6	43.8	16.7	8.0	4.3
Представленная система	РУ-ТТ	0.5Мпар	39.6	62.8	44.1	34.6	28.7
Яндекс Переводчик	ТТ-РУ	Н/Д	17.2	47.7	21.6	12.0	7.1
Представленная система	ТТ-РУ	0.9М пар	45.7	65.4	49.2	40.7	33.5

#### 5. Заключение

В этой статье мы представили первую систему русско-татарского машинного перевода, построенную на основе нейросетевых алгоритмов. Был подготовлен обучающий корпус параллельных текстов, применены современные методы расширения объема данных и машинного обучения.

В дальнейшем планируем использовать многофакторные модели, так как грамматическая информация может помочь улучшить качество перевода, а также продолжить пополнение параллельного русско-татарского корпуса.

#### Литература

Bojar O., Chatterjee R., Federmann C., Graham Y., Haddow B., Huck M., Jimeno Yepes A., Koehn P., Logacheva V., Monz C., Negri M., Neveol A., Neves M., Popel M., Post M., Rubino R., Scarton C., Specia L., Turchi M., Verspoor K., Zampieri M.: Findings of the 2016 conference on machine translation. In: Proceedings of the First Conference on Machine Translation. pp.131–198. Association for Computational Linguistics, Berlin, Germany (August 2016), <http://www.aclweb.org/anthology/W/W16/W16-2301>.

Bojar O., Chatterjee R., Federmann C., Graham Y., Haddow B., Huang S., Huck M., Koehn P., Liu Q., Logacheva V., Monz C., Negri M., Post M., Rubino R., Specia L., Turchi M.: Findings of the 2017 conference on machine translation (wmt17). In: Proceedings of the Second Conference on Machine Translation, Volume 2: Shared Task Papers. P.169–214. Association for Computational Linguistics, Copenhagen, Denmark (September 2017), <http://www.aclweb.org/anthology/W17-4717>.

Wu Y., Schuster M., Chen Z., Le Q.V., Norouzi M., Macherey W., Krikun M., Cao Y., Gao Q., Macherey K., Klingne J., Shah A., Johnson M., Liu X., Kaiser L., Gouws S., Kato Y., Kudo T., Kazawa H., Stevens K., Kurian G., Patil N., Wang W., Young C., Smith J., Riesa J., Rudnick A., Vinyals O., Corrado G., Hughes M., Dean J.: Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. ArXiv e-prints (Sep 2016).

Schiebinger L., Klinge I.: Gendered Innovations: How Gender Analysis Contributes to Research. Luxembourg: Publications Office of the European Union (2013).

Fadaee M., Bisazza A., Monz C.: Data augmentation for low-resource neural machine translation. In: n Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL17). P. 567–573 (01 2017).

Zoph B., Yuret D., May J., Knight K.: Transfer Learning for Low-Resource Neural Machine Translation. ArXiv e-prints (Apr 2016).

One model is better than two. Yandex.Translate launches a hybrid machine translation system. <https://goo.gl/PddtYn> (2017), [Online].

ABBYY Aligner 2.0. <https://www.abbyy.com/ru-ru/aligner/>, [Online].

Sennrich R., Haddow B., Burch A.: Improving neural machine translation models with monolingual data. In: In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. P. 86–96. Berlin (2016).

Open-source neural machine translation in Theano. <https://github.com/rsennrich/nematus> (2017), [Online].

Sennrich R., Birch A., Currey A., Germann U., Haddow B., Heafield K., Valerio Miceli Barone A., Williams P.: The University of Edinburgh’s neural mt systems forw mt17. In: In Proceedings of the Second Conference on Machine Translation. vol. 2: Shared Task Papers. Stroudsburg, PA, USA (2017).

Sennrich R., Haddow B., Birch A.: Neural Machine Translation of Rare Words with Sub-word Units. ArXiv e-prints (Aug 2015).

Papineni K., Roukos S., Ward T., Zhu W.: BLEU: a method for automatic evaluation of machine translation. In: In Proceedings of the 40<sup>th</sup> Annual Meeting on Association for Computational Linguistics. P. 311–318 (2002).

Yandex translate. <https://translate.yandex.com/>(2019), [Online].

Baisa, V.: Problems of machine translation evaluation. In: In Proceedings of Recent Advances in Slavonic Natural Language Processing. Brno (2009).

## ПРИНЦИПЫ ПОСТРОЕНИЯ И ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ WORDNET ДЛЯ ТЮРКСКИХ ЯЗЫКОВ

В статье представлен обзор имеющихся на данный момент WordNet-подобных ресурсов для тюркских языков, проанализированы структура, принципы и порядок разработки. Также представлены первые шаги по построению WordNet для татарского языка (TatarNet) на основе перевода данных тезауруса русского языка РуТез и последующих обработок и дополнений. На примере достигнутых результатов обсуждается общая методология разработки ворднетов и рассматриваются ключевые проблемы, возникающие при построении синсетов для татарского языка.

**Ключевые слова:** WordNet, ворднет, татарский язык, тюркские языки, синсет.

Во многих языках мира имеется доступ к комплексным лексическим ресурсам. Традиционные ресурсы, такие как двуязычные и одноязычные словари, тезаурусы и лексиконы составляют лексикографами. Но в эпоху развития компьютерных технологий и масштабного распространения большого объема информации приоритетными становятся цифровые и автоматические системы сбора, систематизации, обработки и хранения лексических данных. Один из таких ресурсов – WordNet, впервые созданный для английского языка в Принстонском университете (США), поэтому получивший название Принстонского WordNet (далее – PWN) [Miller 1995]. Со времени публикации в 1995 году данный ресурс много раз дополнялся и совершенствовался и стал основой для создания многочисленных подобных ему ворднетов для других языков мира, но его сущность и цель остались неизменными.

По своей сущности WordNet представляет собой лексико-семантический ресурс, объединяющий информацию классического словаря и дополнительные данные по смысловым связям. В нем значения слов объединяются под общим концептами, которые называют синсетами (от английского *synonym sets* – «набор синонимов»). В результате получается комплексный словарь, доступный для использования компьютерными технологиями, удобный для анализа текста и других исследований в области языка [Лукашевич 2011, 68].

### **WordNet-подобные ресурсы для тюркских языков**

На данный момент составление WordNet-подобных ресурсов является популярным способом систематизации и автоматизации лексического инструментария во многих языках мира, в том числе и в тюркских, на которых мы остановимся подробнее.

**Турецкий WordNet** (<https://bitbucket.org/ozlemc/twn/downloads/>) [Özlem Çetinoğlu 2018; Bilgin 2004] был разработан в Sabancı University (Стамбул) в рамках проекта BalkaNet, целью которого было построение ворднетов среднего размера для шести языков, территориально расположенных на Балканском полуострове: болгарского, чешского, греческого, румынского и турецкого [Tufis 2004]. BalkaNet был построен на основе комбинации обоих подходов создания ворднетов. Сначала из Принстонского WordNet было отобрано множество синсетов и осуществлен их перевод. Затем, на основе подхода *merge*, было разработано множество синсетов по балканской тематике, общие для всех ворднетов BalkaNet. И, наконец, были разработаны синсеты, специфичные для каждого из языков BalkaNet. В частности, размер турецкого WordNet внутри данного многоязычного ресурса составляет 14,795 синсетов.

Помимо перечисленных выше методов использования, разработчики Турецкого WordNet видят первоочередную пользу данного ресурса в возможности экспортирования смысловых отношений в другие ворднеты и получения искомого слова по его описанию.

**KeNet** (<http://haydut.isikun.edu.tr/kenet.html>) – еще один турецкий WordNet, построенный на базе современных словарей турецкого языка, содержащий 113 217 синсетов [Razieh Ehsani, Ercan Solak, Olcay Taner Yildiz 2018; Razieh Ehsani 2018]. В отличие от BalkaNet, данный ресурс был построен в результате ручной обработки снизу-вверх на основе Турецкого современного словаря, составленного Турецким институтом языка [<http://sozluk.gov.tr>]. Из данного общедоступного онлайн-словаря были отобраны синонимы, которые две группы аннотаторов соотносили между собой по общим значениям с помощью специально созданного приложения. Затем для извлечения смысловых значений были использованы автоматические методы извлечения и обработка текста из Турецкой Википедии и Викисловаря.

**Extended Open Multilingual Wordnet** (<http://compling.hss.ntu.edu.sg/omw/summx.html>). Цель работы над данным ресурсом – создать условия исследователю лексической семантики одного или

нескольких языков для использования ворднетов этих языков без правовых и технических преград через единый интерфейс [Francis Bond 2013]. Результатом данного исследования стал Открытый многоязычный Wordnet, который содержит всего более 2 миллионов значений для 117,659 концептов из более 1000 языков мира. Ресурс был построен путем комбинирования открытых ворднетов с данными, автоматически извлечёнными из Wiktionary и Unicode Common Locale Data Repository (CLDR).

Открытый многоязычный Wordnet содержит данные и по русскому языку (20,138 синсетов, которые покрывают 64% ядерных концептов PWN), и по семи тюркским языкам: азербайджанскому (1923 синсетов, покрытие 35%), казахскому (1124 синсетов, покрытие 8%), киргизскому (793 синсетов, покрытие 7%), татарскому (550 синсетов, покрытие 5%), туркменскому (680 синсетов, покрытие 7%), турецкому (7953 синсетов, покрытие 35%) узбекскому (889 синсетов, покрытие 8%).

**BabelNet** (<http://babelnet.org/>) – еще один многоязычный, сверх-большой широкоохватный лексико-семантический ресурс [Roberto Navigli 2012]. Изначально, в 2013 году, он был построен при помощи комбинирования самой большой многоязычной открытой интернет-энциклопедии – Википедии – с самым популярным компьютеризированным лексиконом – WordNet. Синсеты WordNet и вики-страницы Википедии были интегрированы в единые концепты (в т.н. Babel синсеты) через автоматическое преобразование, а затем лексические пустоты в ограниченных ресурсами языках были заполнены при помощи машинного перевода.

Самое главное преимущество BabelNet перед другими ворднетами заключается в том, что он постоянно пополняется за счет расширения своих источников. На данный момент (последняя дата обновления – февраль 2018 года) BabelNet объединяет более 15 миллионов синсетов для 284 языков мира на основе обработки данных, полученных из 47 источников.

В данном ресурсе также представлены почти все тюркские языки. В частности, в нем содержится 69,756 определений и 609,628 текстовых входов татарских слов, которые относятся к 1,862,223 концептам.

#### **WordNet для татарского языка**

Изучив богатый опыт создания ворднетов по всему миру, и понимая, что ограниченность ресурсов и недостаточная разработанность лексической базы на данный момент не позволит создать полноценный WordNet для татарского языка (TatarNet) на основе монологических словарей, мы решили взять за основу Тезаурус русского языка РуТез [Лукашевич 2014]. В результате автоматической обработки данного ресурса, в том числе автоматического перевода, была сформирована таблица с текстовыми входами на русском языке с переводом на татарский язык. Основной задачей была проверка соответствия этого перевода на основе Русско-татарского словаря под редакцией Ф.А. Ганиева, редактирование и добавление возможных вариантов. Основанием для этого служили данные колонок с гипонимами, гиперонимами и глоссарием, так как приоритетной задачей была не оценка правильности перевода отдельных слов, а передача понятий оригинала на язык перевода.

При анализе и редактировании текстовых входов на язык перевода можно наблюдать интересные моменты передачи синсетов русского языка на татарский язык.

1. Синсеты из русского языка представляют собой существительные и отыменные глаголы. Почти все они на татарский язык переводятся в форме существительного (исем) или отыменного глагола (исем фигыль). Например: *величие – бөеклек, олылык; вескость – авыр булу, саллы буллу.*

В некоторых случаях при переводе отыменных глаголов употребляются обе части речи. Например: *бездействие – бер нәрсә дә эшләмәү; чара күрмәү, гамьсезлек; гегемония – гегемония, житәкчелек итү, өстенлек.*

2. В анализируемой таблице, как и во всех русско-татарских словарях, много слов, передаваемых на татарский язык при помощи описательной конструкции. Всего на конец файла насчитывается 375 примеров такой формы передачи текстовых входов, не употребляемых на татарском языке. Здесь решающим фактором являлось не количество слов, а наличие в словарях только отдельных компонентов словосочетания или описательного оборота. Например: *коренник – төпкә жигелгән ат; выскочка – сикергәк, ялагай, сәнәктән көрәк булган кеше.*

Эти описательные обороты можно поделить, в зависимости от лексического значения и состава слова источника, на 4 группы:

А) Коренные слова, не имеющие соответствующего варианта на татарском языке по причине того, что эти понятия не свойственны культуре этого народа. Например: *именинник – исем бәйрәмен үткәрүче; клюка – кәкре башлы таяк; конура – эт оясы.*

Б) Термины и понятия, не имеющие перевода на татарском языке, передаваемые заимствованиями и/или описательным оборотом. Например: *дротик – дротик, кыска саплы сөңге; горизонталь – ятма сызык, горизонталь, горизонталь сызык; котельная – пар казаннары бинасы; микрокосм – кечкенә зурлыклар дөньясы, микрокосм.*

В) Сложные слова, не имеющие идентичных по строению слов на языке перевода. Например: *водосток – су агып төшиз торган торба; двустволка – ике көпишале мылтык; естествоиспытатель – табигать фәннәре белгече.*

Г) Слова с повтором вида передаваемого понятия. Это или названия месяцев, или наименования растений и деревьев. Например: *январь – гыйнвар, гыйнвар ае; вяз – карама, карама агачы; липа – юкә, юкә агачы.*

Повторяться могут и слова, описывающие национальность или принадлежность к чему-либо другому. Например: *японец – япон, япон кешесе; девочка – кыз бала; иноходец – юрга, юрга ат.*

3. В обработанных данных часто встречаются текстовые входы, обозначающие названия профессий, национальностей и рода занятий женского рода. Из-за того, что в татарском языке отсутствует морфологическая категория рода, для их передачи применяются лексические средства, что приводит к увеличению описательных конструкций. Кроме того, в текстовых входах добавляется еще конкретизация возраста: *кыз (девушка) или хатын (женщина)*. Например: *активистка – активистка, актив хатын, актив кыз; караимка – караим хатыны, караим кызы; купальщица – су коенучы хатын, су коенучы кыз; манекенищица – манекенчы хатын, манекенчы кыз.*

4. Проблемную область составляют синсеты на русском языке, для которых в татарском языке нет своих соответствующих понятий. Это слова религиозной тематики – понятия православия, не свойственные религии татар – исламу. Статистически их немного – всего 32. Например: *ересь – ересъ; молебен – молебен; миропомазание – миро белән майлап чукиндыру.*

Как видно из примеров, они передаются 3 путями:

- 1) без изменений (ересь, канон, коляда, священник);
- 2) при помощи описательной конструкции (чиркәү манарасы, чиркәүдә чаң кагучы);
- 3) заменой близкого понятия из мусульманской терминологии: *исповедник – тәүбә иттерүче, тәүбәче, тәүбә-истиғфарчы.*

#### **Общие выводы**

По итогам анализа и обработки синсетов на русском языке с их передачей на татарский язык можно сделать несколько выводов:

1. Разработка WordNet представляет собой интересное и продуктивное направление для изучения, систематизации лексического богатства отдельных языков, в том числе и татарского языка. В процессе работы не только подтверждаются общепринятые особенности в переводе отдельных лексических единиц из одного языка на другой, но и выявляются не заметные до этого тонкости передачи культуры и своеобразия народного колорита языковыми средствами.

2. Работа над синсетами также четко показывает западающие моменты в лексикографии отдельных языков. К примеру, процесс работы был бы более продуктивным, а итоги более достоверными при наличии проработанных электронных версий словарей на обоих языках.

3. В процессе работы также подтверждается необходимость обновления и обогащения словарей татарского языка терминологией, современными и широко употребляемыми просторечными словами для более точной передачи заимствований средствами самого языка.

На данный момент завершена только первичная обработка автоматически полученных данных. На наш взгляд, в дальнейшем работу над TatarNet можно продолжить в следующей последовательности:

- проверка качества и репрезентативности полученных данных через сопоставление с частотным словарем, созданным на основе Татарского национального корпуса «Туган тел» [ТНК], добавление пропущенных значений;
- сопоставление полученных данных с ядром Принстонского WordNet, добавление пропущенных значений;
- формирование полноценных синсетов из полученных текстовых входов;
- выстраивание смысловых связей между полученными синсетами и их заполнение пустых синсетов;
- дальнейшее дополнение полученной системы другими частями речи и т.д.

#### **Литература**

Лукашевич Н.В. РуТез-Lite, опубликованная версия тезауруса русского языка РуТез / Н.В. Лукашевич, Б.В. Добров, И.И. Четверкин // Международная конференция по компьютерной лингвистике Диалог-2014. – 2014. – С. 340-349.

Лукашевич Н.В. Тезаурусы в задачах информационного поиска. – М.: Изд-во Моск. ун-та, 2011. – 511 с.

- Bilgin O., Çetinoglu Ö., Oflazer K. Building a Wordnet for Turkish // Romanian Journal of Information Science and Technology. – 2004. – V. 7. – No 1–2. – P. 163–172.
- Francis Bond, Ryan Foster. Linking and Extending an Open Multilingual Wordnet // Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013). – P. 1352-1362.
- Miller G.A. “Wordnet: a lexical database for english,” Communications of the ACM, vol. 38. – № 11. – pp. 39–41, 1995.
- Özlem Çetinoğlu, Orhan Bilgin, Kemal Oflazer. Turkish Wordnet // Kemal Oflazer, Murat Saraçlar (eds). Turkish Natural Language Processing. Springer, 2018. doi:10.1007/978-3-319-90165-7\_15.
- Razieh Ehsani, Ercan Solak, Olcay Taner Yildiz. Constructing a WordNet for Turkish Using Manual and Automatic Annotation // ACM Transactions on Asian and Low-Resource Language Information Processing, Volume 17 Issue 3, May 2018. Article № 24. doi:10.1145/3185664.
- Razieh Ehsani. KeNet: A Comprehensive Turkish Wordnet and Using It in Text Clustering. PhD Thesis. Işık University, 2018.
- Roberto Navigli, Simone Paolo Ponzetto. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network // Artificial Intelligence, Volume 193, December 2012. – P. 217–250.
- Tufis D., Cristea D., Stamou S. BalkaNet: Aims, Methods, Results and Perspectives. A General Overview // Romanian Journal of Information Science and Technology, Volume 7, Numbers 1–2, 2004. – P. 9–43.

## УДМУРТСКИЙ ЯЗЫК В СЕТИ ИНТЕРНЕТ

Сегодня интернет является незаменимой частью современного общества. Информация в глобальной сети представлена на различных языках народов мира. Не исключением являются и сайты на удмуртском языке, которые, в той или иной степени, функционируют в сети Интернет.

**Ключевые слова:** Интернет, медиа, многоязычие, русский язык, удмуртский язык, национальные языки

Возраст Интернета насчитывает всего несколько десятков лет, поскольку он появился на рубеже 60–70 годов прошлого века. Однако его вторжение в жизнь человека нельзя назвать иначе, чем ошеломляющим. Треть населения Земли использовали Всемирную Паутину хотя бы раз, причем большинство делает это регулярно. И, если суммарный относительно низкий показатель числа пользователей достигается за счет людей старшего возраста, то молодежь повально покупает смартфоны, чтобы оставаться на связи даже по дороге в школу.

Бесспорное преимущество, которое Интернет принес людям – это молниеносный обмен информацией. За считанные секунды найти в Сети нужную книгу, связаться с другом, находящимся на другом конце света, отыскать ответ на интересующий вопрос – сложно даже представить насколько такие возможности изменили представления о знаниях, информации, бизнесе, отношениях. Причем можно не только посмотреть фильм, но и тут же обсудить его с другими пользователями, не только прочесть статью, но и поинтересоваться у специалистов, насколько она компетентна.

Еще одно новшество, принесенное в реальность Всемирной Паутиной – это стирание границ между государствами. Благодаря Интернету человечество становится действительно единым, а появление электронных денег и электронных же способов их заработка и траты, вообще задает много вопросов о грядущем мироустройстве. Люди общаются без границ, совершают покупки, рассчитываются, договариваются, и весь этот процесс происходит практически бесконтрольно.

Одним из приоритетных направлений социолингвистических исследований является разработка эффективных стратегий сохранения лингвистического наследия человечества в условиях негативного влияния глобализации на сохранность миноритарных языков и культур. Важным ресурсом поддержки лингвокультурного плюрализма по праву может считаться Интернет. Являясь одним из символов глобализации, он может способствовать активному продвижению миноритарных языков и внести вклад в укрепление их позиций через создание «национального» или «миноритарного» Интернета, который определяется как совокупность сетевых ресурсов, ориентированных на аудиторию, значительную долю которой составляют люди одной этнической группы. Данная дефиниция включает в себя дихотомию: а) сайты на языке данной национальности; б) сайты на других языках, ориентированные на представителей данной национальности.

Целью настоящей статьи является исследование сетевых ресурсов национального удмуртского Интернета на предмет особенностей репрезентации в нём удмуртского языка, имеющего юридический статус государственного на территории Удмуртской Республики.

На основе мониторинга сайтов удалось обнаружить, что, несмотря на официально установленное двуязычие, русский язык доминирует во всех приоритетных коммуникативных сферах, в том числе в сфере массовой коммуникации, частью которой является Интернет.

К сожалению, на современном этапе с удмуртскоязычными сайтами дела обстоят не так радужно. «Целый ряд ярких независимых инициатив канул в лету. Так, например, произошло с удмуртской «Афишей» – сайтом «Ваньмон дыр», а это очень важный сегмент и важный жанр. Существование такого издания положительно сказывалось на престиже удмуртского языка и позволяло привлечь к удмуртскому чтению большее число пользователей. Большинство ссылок, представленных на сайтах «Удмурт инфо» [ЭР 1] и «Языки народов России в интернете» [ЭР 2], также оказались нерабочими. Научно-образовательные или общественно-политические порталы в целом отсутствуют».

Сайты Удмнета условно можно разделить на следующие ключевые сегменты: официальный, включающий сетевые ресурсы, созданные с целью презентации в Интернете государственных и коммерческих институтов и учреждений; медийный, к которому относятся сетевые ресурсы, ориентированные на передачу новостной информации; познавательный, включающий сетевые ресурсы, предназначенные для получения познавательной информации; социальный, к которому относятся сетевые ресурсы, необходимые для разного рода социальных коммуникаций среди пользователей.

## **1. Официальный сегмент**

В официальном сегменте Интернета удмуртский язык представлен столь же ограниченно, как и за его пределами. Следует отметить, что, согласно Закону «О государственных языках УР», удмуртский язык наряду с русским может использоваться в работе органов государственной власти Удмуртской Республики и органов местного самоуправления, однако в виртуальном пространстве эта возможность пока остаётся нереализованной. Официальный сайт главы УР и Правительства УР [ЭР 3] представлен лишь на русском языке и мы не обнаружили ни одного документа или информационного сообщения на удмуртском языке. На данный момент единственным республиканским органом государственного управления (из 22 ведомств), имеющим двуязычную (русскую и удмуртскую) версию, является Министерство национальной политики Удмуртской Республики [ЭР 4].

Необходимо отметить, что удмуртского интерфейса нет и у сайтов организаций, которые по роду своей деятельности напрямую связаны с удмуртским языком и культурой (например, Удмуртский государственный университет [ЭР 5], Национальный музей Удмуртской Республики им. Кузубая Герда [ЭР 6] и др.). Разработка двуязычных версий официальных сайтов учреждений и организаций Удмуртии представляется перспективным направлением работы по укреплению позиций второго государственного языка республики в Интернете.

## **2. Медийный сегмент**

Медийный сегмент Удмнета представлен сайтами республиканских газет и журналов: Региональная служба новостей УР [ЭР 7], Республиканская газета «Удмурт дунне» [ЭР 8], Газета «Зечбур» [ЭР 9], Журнал «Кенеш» [ЭР 10], Журнал «Инвожо» [ЭР 11].

Телепередачи и радиопередачи на удмуртском языке представлены на сайте ТРК «Удмуртия» [ЭР 12].

Появление электронных версий новостных печатных изданий позволило значительно расширить аудиторию газет, что способствует распространению и популяризации удмуртского языка.

## **3. Познавательный сегмент**

Примером познавательного сетевого ресурса в первую очередь является научно-культурный информационный портал «Удмуртология» [ЭР 1], созданный Денисом Сахарных в 2002 г. и имевшим большую популярность в свое время. К сожалению, последнее обновление на сайте проведено 2 января 2011 г. и в настоящее время большой популярностью сайт не пользуется.

В данной категории также необходимо отметить сайты Национальной библиотеки УР [ЭР 13], удмуртскоязычный раздел Википедии [ЭР 14] и сайт, на котором представлена первая версия корпуса удмуртского языка [ЭР 15], выполненного в рамках проекта в Школе лингвистики НИУ ВШЭ.

## **4. Социальный сегмент**

Основной точкой притяжения пользователей Удмнета сегодня, безусловно, являются социальные сети. Поиск по запросу «удмуртский язык» в популярных социальных сетях «Facebook», «Одноклассники» и «ВКонтакте» позволил установить, что шире всего он представлен в последней из них. На сегодняшний день популярны юмористические страницы, группы с удмуртскими словами и выражениями, сообщества о ремёслах, национальной поэзии и современной культуре, сообщества по изучению удмуртского языка и др., т.е. паблики, содержащие разнообразный контент (свыше 60-ти сообществ). Так, например, «ВКонтакте» такой средой является сообщество «Удмуртлык» [ЭР 16] – всеудмуртская общественная площадка, где «формируется актуальная повестка современного удмуртского общества».

В середине января 2018 года удмуртских юзеров социальной сети «ВКонтакте» всколыхнуло отключение удмуртского языка в интерфейсе пользователей. Причиной тому стала низкая доля перевода паблика на национальный язык республики – всего 39%. Однако усилиями активистов уже через несколько дней его удалось вернуть в строй.

Такая инициатива вполне оправдана хотя бы тем, что, согласно недавним исследованиям «Языки России в интернете» Высшей школы экономики [ЭР 17], удмуртский оказался одним из самых популярных в «ВКонтакте». Специалисты нашли 1 630 сообществ, в которых есть хотя бы один текст на малом языке народов России. Среди них около 300 удмуртскоязычных, вторым по количеству групп оказался башкирский язык, третьим – якутский. Также было найдено 379 сайтов на национальных языках.

Также в данный сегмент следует отнести сайты-блоги, насчитывающие порядка 50-ти персональных страниц, большинство из которых уже не существует. Но, как отмечают сами пользователи: «К сожалению, сейчас в «письменной» блогосфере – период декаданса».

Вероятно, упадок в удмуртской блогосфере можно объяснить с появлением в нашей жизни социальных сетей, в которых люди проводят огромное количество времени. Именно туда переключались все блогеры.



Удмуртский язык в глобальной сети представлен и за пределами Удмнета. Так, на видеохостинге «Youtube» размещены видеозаписи на русском и удмуртском языках об истории и культуре Удмуртии, различных событиях в жизни Республики; там же представлена подборка материалов по изучению удмуртского языка. Большой популярностью в «Youtube» пользуются ролики с удмуртской эстрадой.

Одним из знаковых событий является то, что в 2016 году удмуртский язык появился в Яндекс переводчике. В новостном портале, представляющем это событие отмечено [ЭР 18], что «Удмуртский язык – один из самых сложных для машинного перевода, потому что в интернете мало одинаковых текстов на удмуртском и других языках. Сейчас перевод для удмуртского работает в режиме тестирования, и возможны недочёты. Переведённые фразы пользователи могут редактировать – это поможет Яндексу улучшить качество перевода».

Таким образом, удмуртский язык в национальном Интернет-пространстве, пусть в небольшом количестве, но представлен. Наиболее существенными негативными характеристиками Удмнета являются практически полное отсутствие второго государственного языка Удмуртской Республики на официальных сетевых ресурсах, а также многократное преимущество русскоязычных источников в медийном сегменте (в том числе в подаче информации, касающейся народа, его языка и культуры). Подобный характер репрезентации удмуртского языка в национальном Интернет-пространстве во многом обусловлен объективными характеристиками языковой ситуации в Республике (асимметричный билингвизм, демографическая неравносность, коммуникативная несбалансированность), а также подчас индифферентному отношению русскоязычного большинства к миноритарному удмуртскому языку. На наш взгляд, первоочередной задачей по развитию Удмнета является разработка и внедрение двуязычных версий официальных сайтов учреждений и организаций Удмуртии, что необходимо для укрепления позиций второго государственного языка республики и повышения его статуса.

### **Прогнозы и перспективы**

Так есть ли у Удмнета перспективы для развития? Удмуртские активисты и все равнодушные уверены, что есть: будут появляться новые технологические решения, осваиваться новые каналы и среды, появятся тематические порталы и страницы по разным направлениям, т.е. необходимо уделять огромное внимание развитию многоязычия в сети Интернет.

Почему важно развивать многоязычие в Интернете?

Во-первых, представление языка в Интернете уменьшает риск его исчезновения, потому что люди, владеющие этим языком, получают возможность общаться на нем посредством электронной почты, мессенджеров, чатов и т. п., читать новости, художественную и специальную литературу, получать другую необходимую им информацию через Интернет.

Во-вторых, Интернет способствует обеспечению равноправного и всеобщего доступа к знаниям и преодолению «цифрового разрыва». Он предоставляет огромные возможности пользователям в плане получения образования, доступа к информации и обеспечения свободы выражения мнений. Однако получить информацию и услуги в Интернете можно только на сравнительно небольшом количестве языков, поэтому люди, не говорящие на языках, употребляемых в Интернете, не могут адекватно участвовать в информационном обмене. Это может привести к еще большему углублению «цифрового разрыва» и дальнейшей маргинализации отдельных слоев общества.

Кроме того, стоит отметить, что развитие языкового разнообразия в Интернете имеет важное политическое значение. Распространение многоязычной информации об истории, языках и культурах различных народов способствует укреплению толерантности и взаимопонимания, является одним из важнейших факторов обеспечения стабильного и мирного развития современной цивилизации.

### **СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

- ЭР 1: [Электронный ресурс]. URL: <http://udmurtinfo.ru/> (дата обращения: 10.11.2019).
- ЭР 2: [Электронный ресурс]. URL: <http://www.peoples.org.ru/> (дата обращения: 10.11.2019).
- ЭР 3: [Электронный ресурс]. URL: <http://www.udmurt.ru> (дата обращения: 10.11.2019).
- ЭР 4: [Электронный ресурс]. URL: <http://www.minnac.ru> (дата обращения: 10.11.2019).
- ЭР 5: [Электронный ресурс]. URL: <http://udsu.ru> (дата обращения: 10.11.2019).
- ЭР 6: [Электронный ресурс]. URL: <http://nmur.ru> (дата обращения: 10.11.2019).
- ЭР 7: [Электронный ресурс]. URL: <https://udmurt.media> (дата обращения: 10.11.2019).
- ЭР 8: [Электронный ресурс]. URL: <https://udmdunne.ru> (дата обращения: 10.11.2019).
- ЭР 9: [Электронный ресурс]. URL: <http://zechbur.com> (дата обращения: 10.11.2019).
- ЭР 10: [Электронный ресурс]. URL: <http://kenesh.org> (дата обращения: 10.11.2019).
- ЭР 11: [Электронный ресурс]. URL: <https://invozhos.u> (дата обращения: 10.11.2019).
- ЭР 12: [Электронный ресурс]. URL: <http://www.myudm.ru> (дата обращения: 10.11.2019).

- ЭР 13: [Электронный ресурс]. URL: <https://unatlib.ru/udm> (дата обращения: 10.11.2019).
- ЭР 14: [Электронный ресурс]. URL: <https://udm.wikipedia.org> (дата обращения: 10.11.2019).
- ЭР 15: [Электронный ресурс]. URL: <http://udmurt.web-corpora.net> (дата обращения: 10.11.2019).
- ЭР 16: [Электронный ресурс]. URL: <https://vk.com/udmort> (дата обращения: 10.11.2019).
- ЭР 17: [Электронный ресурс]. URL: <https://www.gumilev-center.ru/udmurtskijj-yazyk-okazalsya-odnim-iz-samykh-populyarnykh-v-vkontakte-ehksklyuziv> (дата обращения: 10.11.2019).
- ЭР 18: [Электронный ресурс]. URL: <http://fugazeta.ru/na-servise-yandeks-perevodchik-poyavilsya-udmurtskij-yazyk> (дата обращения: 10.11.2019).

## ЭЛЕКТРОННЫЕ ОБРАЗОВАТЕЛЬНЫЕ РЕСУРСЫ И МЕТОДИКА ОБУЧЕНИЯ НАЦИОНАЛЬНЫМ ЯЗЫКАМ

*Н. Абдурахмонова, Г. Хайруллаева, К. Урдишев  
г. Ташкент, Узбекистан*

### ИСПОЛЬЗОВАНИЕ КОРПУСА ДЛЯ ОБУЧЕНИЯ СИНОНИМИИ НА УЗБЕКСКОМ ЯЗЫКЕ

В данной статье рассматривается среда обучения узбекскому языку с помощью корпусного инструмента в области синонимии в лексикологии. Согласно расширяющейся исследовательской синонимии для учащихся, узбекский язык как второй или иностранный должен быть непростым объяснением в различном значении слов.

Развитие компьютерных технологий влияет на преподавание и изучение языка. Особенно язык как иностранный и второй язык (L2) требует больше аспектов языковой технологии. Одним из вкладов в проведение исследований в этой области является корпус лингвистики.

**Ключевые слова:** корпусное обучение, узбекский язык, AntConc, синонимы.

*N. Abdurakhmonova, G. Khayrullayeva, K. Urdishev  
Tashkent, Uzbekistan*

### USING A CORPUS FOR TEACHING SYNONYMY IN UZBEK

This paper deals with teaching environment for the Uzbek language by corpus tool in the scope of synonymy in lexicology. According to expanding research synonymy for learners, Uzbek as second or foreign language is to be challenge to explain in the different meaning of the words.

Development of computer technologies influence on language teaching and learning. Especially language as a foreign and a second language (L2) requires more aspects of language technology. One contribution of conducting research in this field is corpus linguistics.

**Keywords:** corpus based teaching, Uzbek, AntConc, synonyms.

A lot of research has been conducted on corpus linguistics as a teaching technique to highlight how native speakers of English use certain language forms, vocabulary items, and expressions. “The use of authentic and real-life examples with L2 learners is more beneficial than examples that are made up by the teacher and do not simulate real life use of language. Corpora can be integrated in language classrooms to teach students of various levels” [URL: [https://www.researchgate.net/publication/282186652\\_BENEFITS\\_OF\\_CORPUS-BASED\\_APPROACH\\_TO\\_LANGUAGE\\_TEACHING](https://www.researchgate.net/publication/282186652_BENEFITS_OF_CORPUS-BASED_APPROACH_TO_LANGUAGE_TEACHING)].

In the sphere of corpus-based teaching, we can see some investigations on agglutinative language [Güvenir, Oflazer, 1994] one good example applying corpus tool for Turkish morphology. Affixes adding the word modified owing to harmony of vowels the last components of the stem. According to this information, there are nominal and verbal paradigms for word formation. Hence, sequences of morphotactic combinations of the words might be difficult for L2 learners to understand derivation of parts of speech. Authors’ the main points for teaching is to explain morphological structure and part of speech ambiguity. As a result above mentioned facts that “developed POS tagger for Turkish texts based on two-level specification of Turkish morphology augmented with a multi-word and idiomatic construct recognizer, and a morphological disambiguator based on local neighborhood constraints, heuristics and limited amount of statistical information” [Oflazer, Kuruoz, 1994].

Teaching or learning the language are crucial standpoint how to acquire language with special aspect the circle of “*teacher* ↔ *resource* ↔ *student (L2 learner)*”. Hence, any kind of information (oral or written text) contains two attributes: core information (as a basic element) and periphery (given extra details about what is that) to comprehend the language. Predicate is a basic part of the sentence in Uzbek. One of typical peculiarity of agglutinative languages is nearly all components of the sentence connected and followed predicate (mood, tense, voice, person etc). Other constituencies have the role of periphery in the sentence. For instance, “*Kelmadingmi?* - Did you not come?”. If we analyze this sentence by morphologically “*kel*” - main verb, “*-ma*” - negative form, “*-di*” - past tense, “*-ng*” - 2<sup>nd</sup> person, “*-mi*” - particle for question. Here some parts of speech are dropped to give this information: “*Sen kecha kelmadingmi?*” However, the meaning is the same with the previous short one.

Corpus based teaching is handful and advantageous to acquire language competence. Many scholars discuss what is important on teaching the language by linguistic aspects (morphology, syntax or semantics). Considering grammatical relativeness of each layers of language on keeping balance of each of them is significant task for teaching and learning at the same time. Comprehending all components of the language is complex psychological process depending on extra linguistically information for L2 learners.

Due to free forms and changeable position of the words of syntactic structures in Uzbek can be difficult for L2 learners.

Corpus based teaching is suitable technology for teachers to collaborate with students giving tasks at workshops by individually. Because students by themselves tend to analyze the word at any level of grammar-morphology (word forms and grammatical features of parts of speech) and syntax (structure of word combination and sentences), lexicology (semantic meaning) as well.

Teaching the Uzbek language as foreign language (or as a second language) and native language should be renovated in order to obtain high-level qualification of learners. Although traditional teaching methods (doing exercise, revision, check grammar etc.) are useful as a foundation of knowledge for language learning, this process should comprise self-study education for obtaining language competence at different levels (A, B1, B2 or C1, C2). Corpus based teaching for the Uzbek language as a second or foreign language contains the following steps to work individually for students:

- to identify what level is an object to study;
- what acquiring vocabulary is requirement for learning move forward to study;
- what occupation vocabulary is more specialized for learning
- what is the level of knowledge

Different genres of the texts are selected by corpus tool according to level of the auditory.

### **Corpus based teaching synonymy in Uzbek**

Collocations show the learners combinations connecting grammatical and semantic attitudes of the words together. “There are a large number of miscollocations in their writings (Hill, 2000). According to Hill (2000), L2 learners often lose mark in their writings because they do not know the most common collocations of the words that are necessary in their writing. Studies on collocations in L2 writings conducted by Bahns and Eldaw (1993) as well as Youmei and Yun (2005) revealed that even advanced L2 learners have difficulties producing collocations [Hashemian, 2013]”.

In the Uzbek language there is very rich semantic expressions representing with different meaning in one group of synonyms. Observing the results of the synonyms in dictionary show that synonyms active applied in adjectives. If SOV (*Men kitob o‘qiyman – I read the book*) is general structure of the Uzbek sentence, then the adjective plays an important role to enhance the meaning of each component of parts of speech. So adjective is used for expressing extra factors of information on which listener should draw attention to what information come from what way and what peculiarities it has.

The following tags show for adjective lexical features with synonyms:

Adj_qual	chiroyli, go‘zal, husndor, husnli, xushro‘y, ko‘hlik, ko‘rkam, jozibador, barno, suluv, zebo, latofatli, sohibjamol
Adj_state	kambag‘al, qashshoq, bechora, faqir, yo‘qsil, bechorahol
Adj_colour	ko‘k, moviy, havorang
Adj_form	qiyshiq, qing‘ir, egri
Adj_size	baland, yuqori, yuksak, buyuk, osmono‘par, viqorli
Adj_taste	lazzatli, xushta>m, shirin, totli, ta>mli, xushxo‘r
Adj_smell	sassiq, badbo‘y, qo‘lansa
Adj_time	kunduzgi, ertalabki

L2 learners face problems restriction usage by genres of styles in writing in most cases. Lexical knowledge of L2 learners should clarify between collocation and concordance of the words. Today testing is available with different type of tools for free open source. For instance, Blacklab, BootCat, Sketch Engine, Crawler etc. assist to compile the texts as corpus in database. Unfortunately, currently there is not electron corpus of Uzbek language in internet for foreign learners that is available for self-study. The process of creation corpus is being developed by many scientific investigations for the Uzbek language. Nevertheless,

for teachers one available AntConc used for as a ready tool taken plain texts including concordance, n-gram, key words and key plot. In our article, we tested used AntConc in order to teach the Uzbek language as a foreign language to introduce lexical and grammatical information. Moreover, it is suitable when online is not available in teaching environment.

Using AntConc tool in the classroom is more useful for analysis the language in context. For instance concordance line which taken as a key word for looking up the lexical unit (KWIC) give essential information the surrounding the words to impress for L2 learners.

However, dictionary is also convenient for L2 learners and teachers for using lexeme with the meaning by different aspect of language, corpus can provide those words meaning in different context by concordance given semantic representations. Moreover, frequency of the words give information the exact impression what words are active in communication. In this way, L2 learners can easily know how many words and which ones used in communication effectively. Following table indicates the collection of concordances as example the word “*chiroyli-beautiful*” in synonyms.

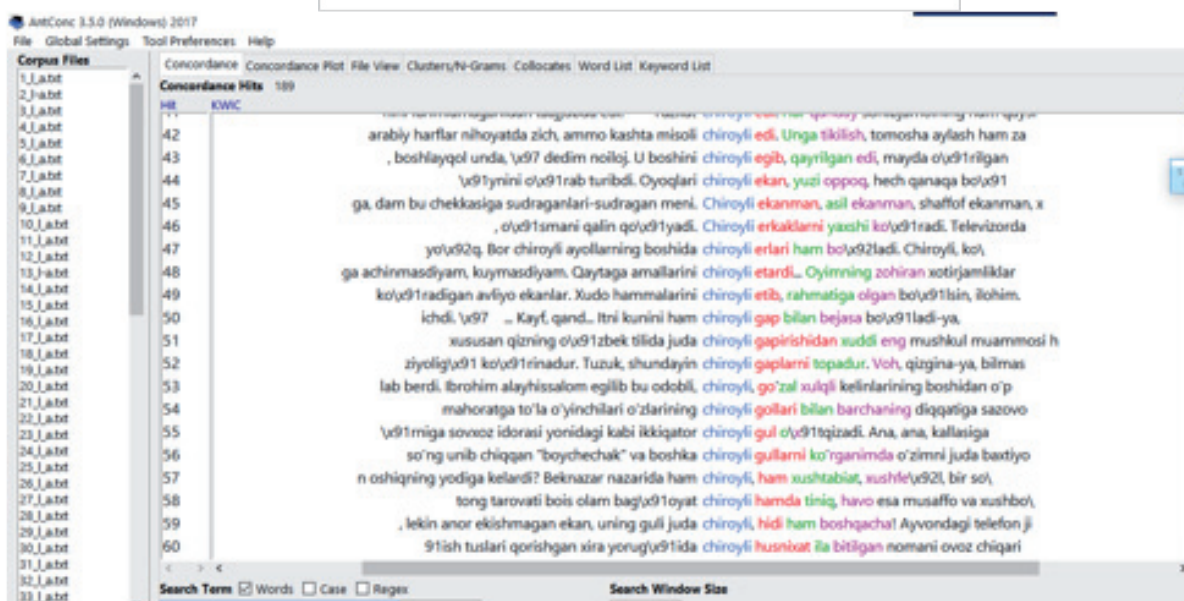
Table 1.

**The collection of concordances as example the word  
“*chiroyli-beautiful*” in synonyms**

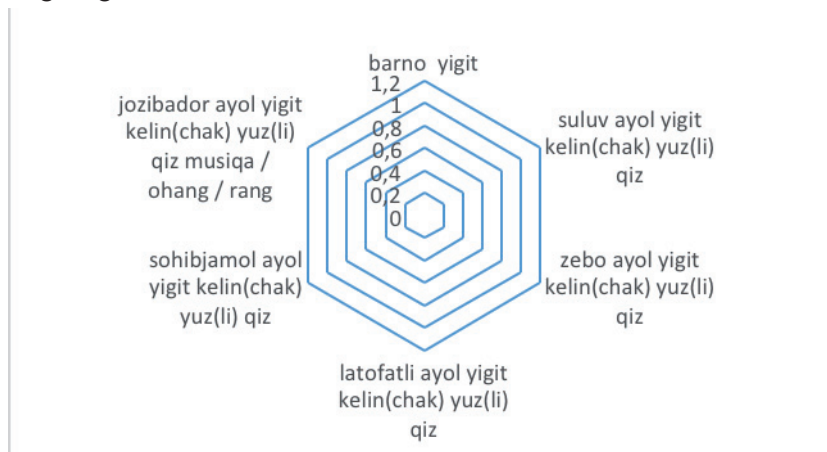
chiroyli	go‘zal	husndor	xushro‘y	ko‘hlik	ko‘rkam	suluv	zebo	lato-fatli	joziba-dor	sohib-jamol
Chiroyli amal	Go‘zal axloq	Husndor ayol	Xushro‘y kelin	Ko‘hlik ayol	Ko‘rkam yigit	Suluv ayol	Zebo qiz	Lato-fatli qiz	Joziba-dor qiz	Sohib-jamol qiz
Chiroyli ayol	Go‘zal ayol	Husndor bo‘lmoq	Xushro‘y raqqosa		Ko‘rkam ot	Suluv kelinchak	Zebo singil	Lato-fatli ayollar	Joziba-dor rang	Sohib-jamol juvon
Chiroyli bosh chayqamoq	Go‘zal qiz		Xushro‘y qomat		Ko‘rkam bino	Suluv yuz	Zebo ayol		Jozi-bador musiqa	Sohib-jamol (as a person)
Egilsangiz chiroyli chiqadi	Go‘zal makon				Ko‘rkam yuzli	Suluv qiz			Jozi-bador ohang	Sohib-jamol siymo
Chiroyli erkak	Go‘zal manzara					Suluv tong				
Chiroyli gap	Go‘zal shahar					Suluv opa				
Chiroyli kishnab	Go‘zal xulq									
Chiroyli gul	Go‘zal yigit									
Chiroyli kiyim	Go‘zal ko‘rinmoq									
Chiroyli ko‘ylak	Go‘zal hayot									
Chiroyli yigit	Go‘zal axloqli									
Chiroyli yuz	Go‘zal xulqli									

Listing all variations semantic pole of *chiroyli*, show what words related as word combination through analyzing them by concordance. Moreover, some synonyms of *chiroyli* can be as names of people: *Go‘zal*,

*Suluv, Zebo.* Additionally some of them can be as metonymy without any word: *sohibjamol, go'zal, chiroyli.* In context L2 learners see a number of examples of each of them.



One of the contribution of corpus based teaching to improve language acquisition, it gives chance to work independently to come conclusion about synonymy usage in communication. As above pointed we can see dominant word *chiroyli* has expanding diapason with combinations of relative words. Other words are not usually applied as the same meaning of a dominant word. If we compare the statistics of word usage, it can be seen the following diagram:



Shaping vocabulary via corpus-based learning in the scope of an exact semantic pole of the word L2 learner can easily image semantic pictures comparing concordances in context. Furthermore, frequency of the words helps to users in communication to express an idea in most situations.

### **Conclusion**

Conducting study the synonymy of Uzbek by corpus we come to conclusion that one of the most effective way for teaching L2 purpose is corpus based technology. Misunderstanding or misused the words made be reduced and it might be available to reach proficiency-writing skill for L2 learners. Concordance line used mainly for writing and reading classes to comprehend and expand vocabulary of the content of the text. Corpus is more efficient than dictionary to gain deep knowledge about association words. Furthermore, it helps to correct some mistakes misused homographs that occurred making sentence [Abdurakhmonova, Urdishev, 2019].

### **References**

- Abdurakhmonova N., Urdishev K. Corpus based teaching Uzbek as a foreign language // Journal of Foreign Language Teaching and Applied Linguistics (J-FLTAL). V.6, №1 – 2019, - P. 131-137.
- Balkan Distance Education Network // BADEN Newsletter, ISSN: 2334-9220 (Online), Year III, Issue 7, Jun 2015, Tempus Project: Enhancing the quality of distance learning at Western Balkan higher education institutions DL@Web, <http://www.badennet.org/>
- Guvener A., Oflazer K. Using a corpus for teaching // Turkish morphology proceedings of the seventh twenty workshop on language technology. The Netherlands 16-17 June, 1994, 28-40.
- Hashemian M. A Corpus-based approach toward teaching collocation of synonyms // Theory and Practice in Language Studies, Vol. 3, No. 1, pp. 51-60, January 2013 © 2013 ACADEMY PUBLISHER Manufactured in Finland.- P. 52.
- Oflazer K., Kuruoz I. Tagging and morphological disambiguation of Turkish text // Submitted to 4th conference on applied natural language processing, 1994.

## МАШИННЫЙ ФОНД КАК СРЕДСТВО ОБУЧЕНИЯ БАШКИРСКОМУ ЯЗЫКУ

В данной статье речь пойдет о Машинном фонде башкирского языка (МФБЯ), одном из информационных ресурсов башкирского языка, разработанного лабораторией лингвистики и информационных технологий Института истории, языка и литературы Уфимского федерального исследовательского центра Российской академии наук.

**Ключевые слова:** машинный фонд, башкирский язык, информационные ресурсы, словарь, корпус башкирского языка.

МФБЯ функционирует по адресу [mfb12.ru](http://mfb12.ru) и состоит на сегодня из 10-ти крупных разделов, каждый из которых содержит свои поисковые системы и базы данных, объемы которых исчисляются от десятков тысяч до десятков миллионов лингвистических единиц [Сиразитдинов, Бускунбаева, Ишмухаметова, Ибрагимов, 2013, 6].

Фонетика и образцы произношения в МФБЯ. Знание и практические навыки по освоению фонетической системы башкирского языка учащиеся могут получить в «Экспериментально-фонетической базе». В этой базе (подфонде) представлен фонетический словарь наиболее часто встречающихся слов объемом в 8000 единиц. Эта база представляет большой интерес для изучающих башкирский язык самостоятельно.

Словари башкирского языка в МФБЯ. Словари представлены в лексикографической базе, которая содержит 58 структурированных словарей общим объемом словарных статей порядка 500000 единиц. В базе представлены академические и учебные словари: одноязычные и двуязычные, частотные, терминологические, фразеологические, синонимические; словари-справочники (названий населенных пунктов, названий улиц городов, названий водных и горных объектов РБ) и другие. 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, поэтому электронные словари в лексикографической базе также имеют разные структуры. Приведем для примера двуязычные общие словари (башкирско-русский, русско-башкирский):

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.

Грамматическая информация в МФБЯ. Данную информацию в Машинном фонде предоставляют академическая грамматика башкирского языка в гипертекстовом формате, морфологическая база данных и корпусы. Морфологическая база реализована на основе данных выборок из текстов трех функциональных стилей языка: публицистики, прозы и науки на основе выборки объемом в 600000 словоформ. Интерфейс позволяет производить отдельный поиск по именным и глагольным словоизменениям как отдельных морфологических категорий и конкретных аллофонов аффиксов этих категорий, так и поиск по сочетаемости категорий и аффиксов.

В МФБЯ представлены три корпуса: прозаических текстов, публицистики и фольклора [Сиразитдинов, Бускунбаева, Бардыбаева, Ишмухаметова, Рафикова, 2012; Бускунбаева, Сиразитдинов, Ишмухаметова, Ибрагимов, Мигранова, 2012; Сиразитдинов, Бускунбаева., Ишмухаметова, Ибраги-



мова, 2014]. Система морфологической разметки башкирских корпусов ориентирована на представление всех регулярных словоизменительных грамматических форм, не всегда отражаемых и совпадающих с формами, принятыми в академической грамматике [Сиразитдинов, Полянин, 2014, 87]. Морфологическая информация башкирской словоформы в корпусе включает: а) частеречную характеристику; б) совокупность морфологических признаков по типу агглютинативных аффиксов словоизменения, которые подразделяются на именные и глагольные формы.

Программа поиска позволяет производить гибкий поиск по многим лингвистическим параметрам:

- поиск словоформы,
- поиск леммы,
- поиск по семантике,
- поиск грамматических категорий словоизменений,
- поиск грамматических подкатегорий,
- поиск сочетаний грамматических категорий,
- поиск сочетаний грамматических подкатегорий,
- поиск сочетаний словоформ,
- поиск сочетаний лемм.

При этом поисковая система в стандартном режиме выводит примеры употребления лексем, словоупотреблений и грамматических форм в объеме предложения с указанием конкретных источников. Возможен расширенный вывод в объеме до 40 предложений. Например, при запросе на сочетание слов «дусар» и «бул» система выдает 108 примеров употребления по произведениям 46 писателей; на сочетание же слов «дусар» и «итеу» – 139 примеров. Эта возможность МФБЯ позволяет выявлять нормы сочетания слов в башкирском языке.

Корпусы дают иллюстрацию в применении редких и устаревших слов, что, безусловно, будет способствовать возвращению их в состав активного словарного состава через систему обучения. Так, при поиске слова «киэмэт» МФБЯ выдает 11 примеров, приведем только несколько из них:

Шулай итеп халайыкты үлтерә, ти, Баштарына киэмэт килтерә, ти; Рәхимһез, шәфкәтһез Карағөлөмбәт Кыпсактарзың токомон бөтөрә, ти. (Эпос – Тормош-көнкүреш / Күсәк бей (Кыпсак версияһы) / Башкорт халык ижады. Дүртенсе том. Эпос. – Өфө: Китап, 1999. Б.319-344/186).

Киэмәт кайымы көнөнә тиклем ер йөзөндәге халыктар бына ошоллай буталып бөтөрзәр. (Өкиәттәр - Тормош-көнкүреш / Мулла менән мәзиндең бутканан колак какканы / Башкорт халык ижады. XII том: сатирик һәм юмористик әкиәттәр. - Өфө: Китап, 2015.-528 бит./11).

Корпусы могут активно применяться и с целью изучения культуры башкирского народа. Проиллюстрируем это на примере поиска по семантике. Так, в фольклорном корпусе при выборе семантики «музыка», система выдает названия музыкальных инструментов: «курай», «ез курай», «сор курай», «думбыра», «кубыз» и др. Например:

1) Туктамыш шат булды, ти,  
Илгә был зур байрам, тип,  
Һанһыз алтын тәңкәне

.....  
Сор курайзар уйнатып,  
Изеүкәйзе зурланы,  
Йәш батырзы данланы,  
Азна-ун көн туйланы.

(Тарихи эпос / Изеүкәй менән Моразым. Вариант 1 / Башкорт халык ижады: 8-се том. Эпос: иртәктәр һәм эпик кобайырзар. Өфө, 2006. 492 б./526).

2) Китеп хәсрәт, килеп шатлык, бөтөрзә тилмереүзәрзе, Уйнаһын саз вә курайзар, уйын-йыр вақыты инде (Тарихи эпос / Ләйлә менән Мәжнүн / Башкорт халык ижады. 7-се том. Язма кисса һәм дастандар. Өфө, 2004. 624 б./693).

3) Кубуз, думбыра, төрлө уйындар үгрәнде (Тарихи эпос / Файса улы Әмәт (Фасыл фи бәйән дастаны Файса уғлы Әмәт) / Башкорт халык ижады. 7-се том. Язма кисса һәм дастандар. Өфө, 2004. 624 б./190).

4) Бер көндө Бикйән йыйынып, кыны менән ез курайын алған да, ата-әсәһе менән хушлашып, илгә сығып киткән. (Тарихи эпос / Бикйән Мәргән / Башкорт халык ижады. 7-се том. Язма кисса һәм дастандар. Өфө, 2004. 624 б./42).

Сегодня корпусы становятся эффективным инструментом овладения языком, позволяя преподавателям оживить процесс обучения, как за счет замены примеров в школьных учебниках, так и творческого ввода новых методик обучения. На любое слово, конструкцию, явление учитель найдет тысячи

примеров за минуту. Преподаватели могут предложить обучающимся в виде домашней работы найти примеры употребления по корпусам из современных газет, современной литературы, из разговорного языка.

Для овладения любым языком необходимы три вещи, это – словарь, грамматика и живое пространство, где язык, собственно, и функционирует [Плунгян, 2009]. Машинный фонд башкирского языка представляет собой уникальную информационную лингвистическую систему, которая объединяет в себе все эти три источника. Поэтому создатели данной системы надеются, что и учителя башкирского языка, и изучающие башкирский язык самостоятельно, будут активно его применять и по достоинству оценят труд сотрудников лаборатории лингвистики и информационных технологий.

### **Литература**

- Башкирско-русский словарь / Под ред. З. Г. Ураксина. Уфа, 1996. – 884 с.
- Бускунбаева Л. А., Сиразитдинов З. А., Ишмухаметова А. Ш., Ибрагимова А. Д., Мигранова Л. Г. Корпус текстов периодической печати на башкирском языке // Актуальные проблемы диалектологии языков народов России. Уфа, 2012. – С. 139-141.
- Плунгян В. А. Почему современная лингвистика должна быть лингвистикой корпусов // [Электронный ресурс] URL:<http://polit.ru/article/2009/10/23/corpus/> (дата просмотра 10.12.2019).
- Русско-башкирский словарь / Под ред. З. Г. Ураксина. Том 1. Уфа, 2005.
- Русско-башкирский словарь / Под ред. З. Г. Ураксина. Том 2. Уфа, 2005.
- Сиразитдинов З. А., Бускунбаева Л. А., Бардыбаева А. Д., Ишмухаметова А. Ш., Рафикова Л. Н. О разработке национального корпуса башкирского языка // Роль и место национальной библиотеки в социокультурном пространстве. Уфа, 2012. – С. 118-126.
- Сиразитдинов З. А., Бускунбаева Л. А., Ишмухаметова А. Ш., Ибрагимова А. Д. Информационные системы и базы данных башкирского языка. Уфа, 2013.
- Сиразитдинов З. А., Полянин А. И. Об опыте разработки интегрированной корпусной системы на базе СУБД Оракл // Труды Казанской школы по компьютерной и когнитивной лингвистике TEL-2014. Казань, 2014. – С. 85-88.
- Сиразитдинов З. А., Бускунбаева Л. А., Ишмухаметова А. Ш., Ибрагимова А. Д. О создании корпуса башкирского фольклора // Урал-Алтай: через века в будущее. Материалы VI Всероссийской тюркологической конференции (с международным участием). Уфа, 2014. – С. 86-89.

## **НОВЫЕ ИНТЕРАКТИВНЫЕ ФИЛОЛОГИЧЕСКИЕ ТЕХНОЛОГИИ В СОВРЕМЕННОМ ИНФОРМАЦИОННОМ ПРОСТРАНСТВЕ**

В статье описывается опыт работы по реализации ФГОС ООО преподавателями кафедры востоковедения и башкирского языкознания Башкирского государственного университета по теме «Инновационные модели как фактор изучения и сохранения башкирского языка в поликультурном пространстве» на базе гимназии №3 имени Джалиля Киекбаева г. Стерлитамак. Раскрываются инновационные методы, позволяющие решать задачи приведения содержания образования в гимназии в соответствие с требованиями государственного образовательного стандарта, с запросами и возможностями обучающихся, перспективами их дальнейшего развития; разработать механизмы отслеживания и контроля качества образования, выявления и развития обучающихся.

**Ключевые слова:** информационные технологии, интерактивная технология обучения, башкирский язык, изучение башкирского языка, методика обучения.

Язык – душа народа, это не просто крылатое выражение. С помощью языка любой народ из поколения в поколение передает свое восприятие окружающей действительности, свое мировоззрение, ценностные ориентиры. С помощью языка мы не только общаемся, передаем свои мысли, но и выражаем образовательный, духовный и культурный уровень. И если сегодня говорят о необходимости гуманизации образования, то речь идет, прежде всего, о языковой культуре в самом широком смысле этого слова. Именно языковая культура, языковая личность развивается на уроках родного языка и литературы.

В целях реализации ФГОС ООО, на базе кафедры востоковедения и башкирского языкознания Башкирского государственного университета была открыта опытно-экспериментальная площадка по теме «Инновационные модели как фактор изучения и сохранения башкирского языка в поликультурном пространстве» на базе гимназии №3 имени Джалиля Киекбаева г. Стерлитамак. Были поставлены следующие задачи:

- привести содержание образования в гимназии в соответствие с требованиями государственного образовательного стандарта, с запросами и возможностями обучающихся, перспективами их дальнейшего развития;
- разработать механизмы отслеживания и контроля качества образования, выявления и развития обучающихся;
- сформировать дидактическую модель педагогического процесса, обеспечивающего интеллектуальное и духовно-нравственное развитие личности, ее креативных способностей;
- создать воспитательную систему, ориентированную на формирование ЗОЖ у воспитанников механизмов самоуправления и саморазвития, совершенствование личностных возможностей приобщения к социокультурным и образовательным ценностям.

Основная суть обучения заключается в том, чтобы построить в школах образовательную систему, которая способствовала бы изучению и сохранению родных языков в современном поликультурном пространстве путем использования инновационных моделей обучения.

В результате реализации Программы данной экспериментальной площадки, гимназия №3 имени Джалиля Киекбаева должна сформировать нового педагога, интегрирующего в себе учителя образовательной области знаний, воспитателя, психолога, новатора, подготовленного к работе с обучающимися.

Как известно, новые информационные технологии открывают большие возможности расширения образовательных рамок и использования инновационных моделей.

Предлагаем виды инновационных моделей, которые успешно используются педагогами нашей республики:

**1. Использование в работе платформы «Электронная игровая школа» –ePlaySchool.** Гимназия сотрудничает с Стерлитамакским филиалом Башкирского государственного университета с целью практической реализации проекта «Разработка и реализация модели геймифицированной образовательной платформы «Электронная игровая школа» – ePlaySchool» и является одной из пилотных школ данного проекта. Наши учителя активно участвовали в разработке данного проекта. Например, учитель башкирского языка и литературы Тимергалиева С. И. в республиканском конкурсе «Урок на платформе «Электронная игровая школа» заняла почетное второе место. Учащиеся с желанием обучаются на данных уроках, проходят этапы, строят мир, копят баллы и тратят их в свое усмотрение на пользу себе.

**2. Использование специальных сайтов с возможностью совместного использования персонального компьютера учителя и личных смартфонов учащихся.** При помощи специализированного сайта **Triventy** учащиеся с желанием вспоминают пройденный материал, так же сайт **Kahoot** дает возможность проведения коллективных и индивидуальных соревнований.

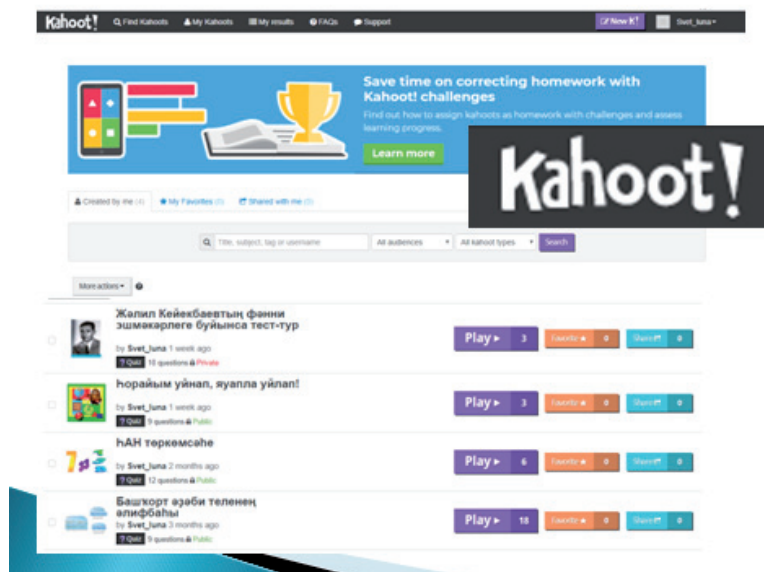


Рис.1. Использование Kahoot для проведения коллективных и индивидуальных соревнований

Перенеся упражнения с учебников в данный сайт, и визуализировав задания, можно дополнительно заинтересовать учащихся. Тем более, в каждом классе у каждого ребенка есть свой аккаунт и учитель может контролировать его самостоятельную работу на каждом этапе. Параллельная работа в тетрадях и на сайте существенно облегчила работу учащимся. Сейчас уже наблюдается ситуация, когда ребенок на перемене со своего смартфона заходит на данный сайт и решает упражнения, с целью получения дополнительных баллов.

### 3. Использование редактора тестов MyTestX.

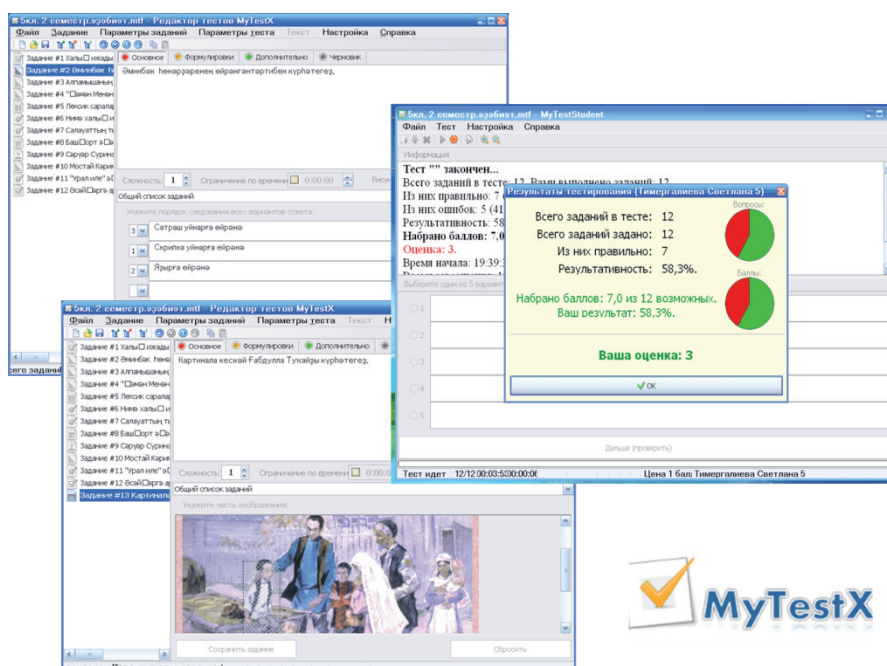


Рис.2. Использование редактора тестов MyTestX

Эта программа используется для составления и проведения тестов для оценивания знаний и умений учащихся. Для коллективной работы требуется специализированный компьютерный кабинет с централизованной системой, так же можно сделать индивидуальную рассылку.

#### 4. Использование веб-сайтов для самостоятельной или групповой работы учащихся в виде прохождения web-квестов.

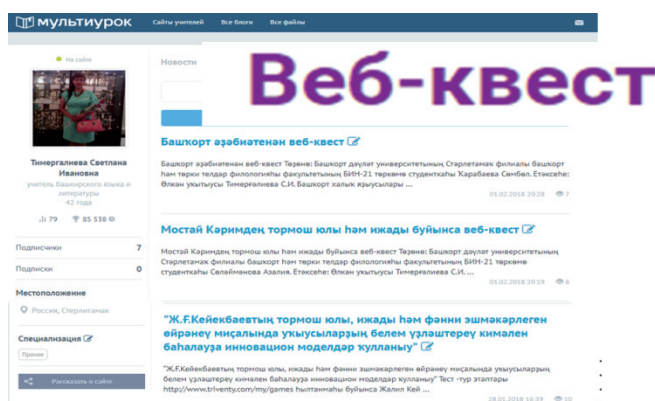


Рис.3. Использование web-квестов для самостоятельной работы учащихся

Особенность веб-квестов в том, что вся информация, расположенная в разных веб-сайтах, благодаря гиперссылкам, находится в едином виртуальном листе. Работа усложняется тем, что пока мало сайтов на башкирском языке. Учащиеся с желанием собирают материалы в Интернете, просматривают материалы канала YouTube в поисках документальных фильмов, спектаклей, решают какую-либо проблему по ссылкам, указанным учителем, или, наоборот, сами создают гиперссылки и рисуют полную картину информации по той или иной теме. Веб-квесты способствуют развитию компьютерных навыков обучающихся и повышению их словарного запаса, поощряют учиться независимо от учителя.

Таким образом, использование компьютерных, информационных технологий поможет учителю поднять уровень владения родным языком учащихся до современных требований международных стандартов. Программа экспериментальной площадки поможет построить образовательную систему, которая будет способствовать изучению и сохранению родных языков в современном поликультурном пространстве путем использования инновационных моделей обучения.

#### Литература

Валиева Г.Д. Использование технических средств на уроках башкирского языка. Учебное пособие для студентов. – Стерлитамак: СФ БашГУ, 2016. – 80 с.

Тимергалеева С.И. Инновационные модели в изучении башкирского языка в современном информационном пространстве. Современные проблемы изучения башкирского и родных языков // Материалы III Всероссийской (с международным участием) научно-методической конференции (Уфа, БашГУ, 15 февраля 2018 года). Отв. ред. Г.Р. Абдуллина. – Уфа: БашГУ, 2018. – 546 с.

## ПРОБЛЕМА ОТТЕСНЕНИЯ РОДНЫХ ЯЗЫКОВ В ВИРТУАЛЬНОМ СОЦИУМЕ

В статье рассматриваются проблемы роли родных языков в виртуальном социуме. На примере башкирского языка изучены актуальные вопросы функционирования родного языка в виртуальном пространстве.

**Ключевые слова:** родные языки, башкирский язык, виртуальное общение, сленг.

Эпоха информационного общества оказывает огромное влияние на повседневную жизнь. Языковые коммуникации выходят на качественно новый уровень. Сегодня жизнь современного человека невозможно представить без гаджетов и возможности выхода в глобальную сеть Интернет, одна из основных функций которой – услуги связи и коммуникации с другими членами социума.

Для коммуникации в виртуальном пространстве человек использует традиционные средства языка, но они, так или иначе, подвергаются влиянию условий виртуального общения. Сегодня каждый язык в мире приобретает универсальные черты языка виртуальной реальности с уникальными характеристиками интернет-коммуникации. Наблюдается тесное взаимодействие всех языков с глобальным английским языком. Для современного человеческого общества характерен масштабный процесс вымирания языков малых народов. По статистике, выявленной ЮНЕСКО, каждый месяц в мире умирает два языка [Мослей 2015, 145].

Все указанные процессы говорят об актуальности проблемы функционирования родных языков в пространстве интернета. Глобальный переход на единый язык несет в себе положительные стороны для интеграции как в культурных, так и в экономических сферах. В то же время, у данного процесса есть и негативные стороны, в частности, такая унификация ускоряет условия, приводящие к вымиранию языков и этносов, а с исчезновением языка мы всегда теряем уникальный блок знаний и понимание человеческой мысли, мировоззрения.

Сегодня проблема оттеснения языка в виртуальном социуме существует и у представителей башкирского этноса. Русскоязычная среда Рунета и факторы отставания башкирского языка в интеграции с виртуальным пространством привели к тяжелым коммуникативным последствиям. Сегодня родным языком в пространстве виртуальной реальности пользуется лишь малая часть представителей башкирской нации. Решение проблемы популяризации и интеграции башкирского языка в социальных сетях и мессенджерах требует совместных усилий общественников и представителей интеллигенции в сфере образования, СМИ и государственной политики.

Вместе с развитием сети Интернет появились новые слова, употребляемые и в бытовом общении. Популярностью среди выражений в социальных сетях пользуются слова из английского языка, ставшего международным языком для коммуникации в Интернет-пространстве. Например: юзер – «пользователь»; онлайн – «пользователь в сети»; контент – «наполнение, содержание»; лайк – «мне нравится, одобряю»; репост – «рассказать о предпочтении чего-либо, поделиться» [Максимов 2015, 106].

В виртуальном мире люди говорят «общим языком», зачастую, менее употребляя в общении слова из родного языка. Тем самым снижается потребность в общении на родном языке. Сегодня носители башкирского языка в интернет-коммуникации зачастую вынуждены обращаться к англоязычным или русскоязычным заимствованиям. Например: «хейтер» – «ненавистник»; «хейтить» – «очернять кого-либо, злословить»; «бомбит» – «сильно волновать»; «юзать» – «пользоваться»; «бан» – «внести в «черный список» заблокированных пользователей»; «изи» – «легче»; «игнор» – «игнорировать, не обращать внимания»; «паблик» – «группа или сообщество» и др. Примеры виртуального общения, взятые из социальной сети «ВКонтакте», свидетельствуют о широком использовании англоязычного сленга в языке представителей всех этносов.

В последние годы, интеграция башкирского языка в сферу виртуальной реальности позволяет компенсировать коммуникативные потери и оттеснение родного языка с пространства Интернета. В 2013 году была запущена башкироязычная версия интерфейса социальной сети «ВКонтакте». С этого времени появились первые положительные тенденции в популяризации башкирского языка в социальных сетях. Появились башкироязычные группы «Башкорт катын-кызы» («Башкирская женщина»), «Башкорт йырзары» («Башкирские песни»), «Башкортостаным, яңылыктар» («Мой Башкортостан, новости») и др.

Внедрение башкироязычного интерфейса и популяризация башкирского языка оказала большое влияние на формирование этноязычной коммуникативной культуры у носителей башкирского

языка. Сегодня наблюдаются первые положительные изменения в использовании родного языка в повседневном общении через сеть. Исследуя коммуникативные выражения носителей башкирского языка в виртуальной реальности, мы можем выявить первые результаты изменений в лингвистическом уровне. Например, многие англоязычные или русскоязычные выражения постепенно в языке пользователей социальной сети заменяются башкирскими кальками или национальным сленгом: «лайк» – «лайык», «досвидос» – «хуш»; «бомбит» – «ғәләмәт, аптырак»; «хейтить» – «төкөрөнөү» (идиоматическое выражение).

Башкирский сленг постепенно проникает в сферу виртуальной коммуникации. Но в большинстве эти выражения присутствуют в сленге старшего поколения, тогда как молодежный сленг все чаще обращается к англоязычным выражениям. В особенности, проблема актуальна для сленга геймеров (игроков), в среде которых выражение англоязычными словами игровой языковой среды является «модным» и естественным стереотипом.

Проблема сохранения башкирского языка в пространстве интернет-социума указывает на необходимость дальнейшей работы по интеграции башкирского языка с виртуальной средой. На сегодняшний день IT-специалисты, лаборатория лингвистики и информационных технологий ИИЯЛ УФИЦ РАН продолжают реализовывать ряд проектов по созданию национальных шрифтов, электронных словарей, учебных пособий по башкирскому языку [Сиразитдинов 2017, 149]. Создана основа для электронной трансформации башкирского языка – терминологическая база, существуют проекты национальных корпусов, опираясь на которые был создан десятитомный академический словарь башкирского языка.

Несмотря на это, для внедрения башкирского языка, как средства виртуальной коммуникации необходимы качественно новые шаги в области образования, создания доступных электронных модулей. В первую очередь, нужна реализация проектов разработки программ автоматического перевода с русского на башкирский язык и с башкирского на русский язык с функцией озвучивания. Современная электронная языковая среда должна уделять большое внимание пользовательским потребностям, среди них – геолокационные данные на башкирском языке и разработка соответствующих интерфейсов, создание культурно-ориентированного национального развлекательного и познавательного контента.

Популяризация башкирского языка в социальных сетях и мессенджерах невозможна без реализации активных школьных образовательных программ, деятельности педагогов по знакомству детей с электронными возможностями родного языка в виртуальном пространстве. Работа педагогов раскрывает неизвестные возможности башкирского языка, помогает внедрению повседневного общения на родном языке в пространство виртуальной реальности. На пути реализации педагогических инициатив в обучении башкирскому языку в пространстве интернет-социума не должно быть ограничений.

Таким образом, проблема функционирования родных языков в пространстве виртуального социума – один из главных актуальных вопросов в работе по сохранению родных языков, так как сегодня, язык, не функционирующий в электронной сфере, не имеет шанса сохраниться как полноценное средство коммуникации и подвергается исчезновению.

### Литература

Кристофер Мослей. Исчезающие языки // Современная библиотека. – 2015. – № 1 (51) – С. 144–159.

Максимов И. С., Дворак Е. В. Сленговые заимствования английского языка в речи современной молодежи // Молодежный вестник ИРГТУ. – 2015. – №4. – С. 105-109.

Сиразитдинов З.А. Роль и место информационных технологий в реализации государственных языков Республики Башкортостан: этапы и перспективы // Актуальные вопросы алтайского языкознания: проблемы развития литературного языка, совершенствование современной орфографии: Материалы Всероссийской научно-практической конференции, посвященной 115-летию Т. М. Тошковой. – Горно-Алтайск: Горно-Алтайская типография, 2017. – С. 135-151.

## **ЦИФРОВЫЕ ОБРАЗОВАТЕЛЬНЫЕ РЕСУРСЫ: ПРОБЛЕМЫ И ПЕРСПЕКТИВЫ**

Цифровизация затрагивает все сферы деятельности, в том числе и образование. Повсеместно во всех учреждениях образования осуществляется переход на электронные учебники, электронный классный журнал. В статье рассмотрены такие новые цифровые образовательные ресурсы, как виртуальные экскурсии и киберспорт. Подробно изучены особенности киберспорта в России, а также его основные плюсы и минусы.

**Ключевые слова:** цифровая экономика Российской Федерации, образование, виртуальные экскурсии, киберспорт.

Цифровизация – это тренд эффективного мирового развития. Переход на цифровые услуги происходит повсеместно. В России в 2018 году была принята программа «Цифровая экономика Российской Федерации». Она имеет ряд целей: создание устойчивой и безопасной информационно-телекоммуникационной инфраструктуры высокоскоростной передачи, обработки и хранения больших объемов данных, доступной для всех организаций и домохозяйств; использование преимущественно отечественного программного обеспечения государственными органами, органами местного самоуправления и организациями; увеличение внутренних затрат на развитие цифровой экономики за счет всех источников (по доле в ВВП) не менее чем в 3 раза по сравнению с 2017 г. Программа предусматривает внедрение цифровых технологий в промышленность, культуру, образование, государственное управление, здравоохранение.

В настоящее время в сфере электронного образования уже создана цифровая инфраструктура, доступная для образовательных учреждений всех уровней. Появляются цифровые образовательные платформы с большими возможностями выбора и персонализации курсов. Развивается дистанционное обучение, цифровые экзамены и аттестация. Повсеместно во всех учреждениях образования в рамках проекта «Электронная образовательная среда РФ» осуществляется переход на электронные учебники, электронный классный журнал. Например, в школах г. Дзержинска Нижегородской области в рамках проекта «Эффективная губерния» изучение ОБЖ будет вестись с использованием современных тренажеров виртуальной реальности.

Цифровизация образования ведет к его коренной, качественной перестройке. Педагог обязан научиться применять новые технологические инструменты и практически неограниченные информационные ресурсы. Технологии виртуальной реальности создают возможность применения цифровых тренажеров, не привязанных к одному рабочему месту, что расширяет круг изучаемых технологий. Технологии мобильного обучения позволяют учиться в любое время и в любом месте. В использовании цифровых технологий видятся большие возможности для образования.

Одно из направлений цифровизации современного образования – это создание новых учебных материалов и программ на основе технологий виртуальной и дополненной реальностей. К новым цифровым образовательным ресурсам можно отнести виртуальные экскурсии и киберспорт. Рассмотрим их более подробно.

Виртуальная экскурсия – это форма обучения, отличающаяся от реальной экскурсии виртуальным отображением реально существующих объектов с целью создания условий для самостоятельного наблюдения, сбора необходимых фактов [Митриковская 2016, 155].

Виртуальные экскурсии можно использовать на уроках истории, литературы, мировой художественной культуры, изобразительного искусства, иностранного языка, географии, астрономии и ряда других дисциплин. Благодаря таким экскурсиям учащиеся из сельских районов и отдаленных уголков России смогут посетить Эрмитаж, Третьяковскую галерею, ряд других мест.

Виртуальная экскурсия более проста в проведении, имеет больше возможностей и не зависит почти ни от каких факторов. Их можно использовать для посещения различных городов, стран, музеев, выставок, театров, ботанических садов, парков и многих других мест. На сегодняшний день уже существует множество готовых экскурсий в сети Интернет и их количество с каждым годом растет.

Ожидаемыми результатами от виртуальной экскурсии могут быть:

- 1) получение знаний об объектах экскурсии (памятники, музеи и т.п.);
- 2) формирование нравственных качеств личности (бережное отношение к окружающей среде, уважение к объектам, представляющим культурное наследие);



- 3) формирование положительных эмоций через визуальное погружение в красоту природы;
- 4) развитие у обучающихся чувства любви к месту, где они родились или находятся.

Таким образом, виртуальные экскурсии предоставляют огромное количество возможностей. Они могут заменить традиционные и стать одной из наиболее эффективных форм обучения.

Также хотелось бы рассмотреть еще одно нововведение, которое может стать реальностью для школьников в ближайшие годы – это киберспорт. Сегодня он стремительно развивается как в России, так и за ее пределами: проводятся соревнования различного уровня, создаются сообщества, осуществляется процесс подготовки спортсменов и т.д.

Российской Федерацией первой в мире киберспорт был признан спортивной дисциплиной на официальном уровне. В 2001 году Министерство спорта РФ подписало Приказ «О признании и включении во Всероссийский реестр видов спорта спортивных дисциплин, видов спорта и внесении изменений во Всероссийский реестр видов спорта» [Приказ Минспорта России от 29.04.2016 № 470] и Приказ «О признании и включении видов спорта, спортивных дисциплин во Всероссийский реестр видов спорта» [Приказ Минспорттуризма России от 17.06.2010 № 606], что означало придание киберспорту статуса спортивной дисциплины на федеральном уровне. Однако впоследствии в 2006 году он был исключен из Всероссийского реестра за несоответствие критериям спортивной дисциплины: развитие в более чем половине субъектов Российской Федерации и наличие зарегистрированного в установленном порядке общероссийского физкультурно-спортивного объединения. И включен туда лишь в 2018 Приказом Министерства спорта [Приказ Минспорта России от 15.06.2018 № 562]. Это значит, что спортсмены смогут получать спортивные звания и разряды. Также стали возможны соревнования и чемпионаты страны.

14 августа 2019 г. Институт развития интернета предложил Министерству просвещения ввести в качестве эксперимента в школах уроки по киберспорту. Отмечается, что занятия должны быть факультативными. Эксперимент предлагается реализовать в нескольких школах в каждом регионе в 2020–2025 годах. Было отмечено, что такой опыт уже успешно проводится в Швеции и Норвегии.

ИТ-индустрия нуждается в притоке квалифицированных кадров, которых не хватает уже сегодня, и этот дефицит будет только расти. Результаты опроса 10 тыс. студентов-программистов, проведенного интернет-сервисом HackerRank, показали, что 65% опрошенных – самоучки. Между тем, в некоторых российских вузах уже сейчас есть специальности, связанные с киберспортом.

Одним из плюсов введения киберспорта как одной из факультативных дисциплин в школе является выраженная активация правого полушария по сравнению с левым. Это было выявлено в ходе исследования влияния компьютерных игр на центральную нервную систему компьютерных игроков на основе анализа электроэнцефалограммы, регистрируемой до и сразу после игры. Анализ общей активности коры головного мозга показал более выраженную активацию правого полушария по сравнению с левым. Этот факт косвенно указывает на развивающий потенциал компьютерных игр для функций правого полушария, что редко встречается в традиционной педагогической практике. Школьное обучение, в основном, больше адресуется к левому полушарию, которое связано с вербальными функциями, счетом, логикой. Соответственно, это способствует развитию его функций. Правое полушарие, «образное», получает гораздо меньше внимания в школьной практике, и компьютерные игры могут восполнить этот пробел [Стрельникова 2014, 96].

Также ребенок может развить в себе умение работать в команде, последовательно продвигаться к победе, наблюдательность, внимательность, дисциплину, хорошую реакцию.

Хотелось бы отметить, что это один из немногочисленных видов спорта, где не имеет значения гендерная принадлежность: девушка может сразиться с противником, примерив на себя любой образ.

Киберспортсмен должен не только хорошо знать правила игры, в которую он играет, но и уметь пользоваться множеством других компьютерных программ. Большинство игроков являются (или становятся со временем) продвинутыми пользователями компьютера и интернета.

Для большинства игр необходимо знание английского языка, что тоже можно считать плюсом, так как человек может оттачивать свои навыки.

Однако вместе с плюсами существуют и значительные препятствия для введения киберспорта в школьную программу.

В настоящее время нет полного понимания физических и психо-неврологических изменений, которые происходят в организме ребенка при длительном взаимодействии с цифровыми технологиями, в частности, с компьютерной игрой. Например, Марк Д. Гриффитс – английский психолог, профессор Ноттингемского университета, один из ведущих мировых специалистов по изучению поведенческой и игровой зависимости утверждает, что как на психологическом, так и на поведенческом уровне азартные игры и видеоигры имеют много общего. Механизм получения удовольствия («вознаграждения»)

при этих играх совершенно другой, чем, например, при занятии классическими видами спорта. Он более возбуждает нервную систему и потенциально ведет к формированию зависимости.

Ряд исследований связывают злоупотребление играми с такими расстройствами, как синдром гиперактивности и дефицита внимания, тревожное расстройство, депрессия, повышенная агрессия и психоз.

Также существует множество утверждений, что киберспорт вредит зрению и приводит к искривлению позвоночника. Однако при правильных паузах и чувстве меры эта проблема не будет являться актуальной.

Таким образом, следует внимательно взвесить все «за» и «против» погружения в эту деятельность.

Вместе с этим, игры для факультативов должны подбираться совместно со школьными психологами и учителями, а между школами следует проводить соревнования для популяризации данного вида спорта и выхода на новые уровни.

Таким образом, цифровые образовательные ресурсы расширяются. Вместе с этим, увеличиваются возможности способов обучения. Но также выявляются и новые угрозы, которые цифровизация несет в себе. Поэтому необходимо не только внедрять цифровые технологии, но и учитывать риски, которые они в себе несут.

### Литература

Митриковская М. С. Музейная педагогика в школе // Молодой ученый. — 2016. — № 9. — С. 154–157.

Приказ Минспорта России от 15.06.2018 N 562 «О государственной аккредитации общероссийской общественной организации и наделении ее статусом общероссийской спортивной федерации по виду спорта «компьютерный спорт» // СПС КонсультантПлюс.

Приказ Минспорта России от 29.04.2016 N 470 «О признании и включении во Всероссийский реестр видов спорта спортивных дисциплин, видов спорта и внесении изменений во Всероссийский реестр видов спорта» // СПС КонсультантПлюс.

**Приказ Минспорттуризма России от 17.06.2010 N 606 (ред. от 21.11.2016) «О признании и включении видов спорта, спортивных дисциплин во Всероссийский реестр видов спорта» // СПС КонсультантПлюс.**

Стрельникова И. В. Развивающий потенциал компьютерных игр / И. В. Стрельникова, Г. В. Стрельникова // Компьютерный спорт (Киберспорт): проблемы и перспективы: Материалы 111-й Всероссийской научно-практической конференции (в формате интернет-конференции). – Москва: РГУФКСМиТ, 2014. – С. 95–97.

## **ЦИФРОВИЗАЦИЯ ОБРАЗОВАНИЯ И НОВЫЕ ТЕХНОЛОГИИ В ОБУЧЕНИИ РОДНЫМ ЯЗЫКАМ**

Самыми значимыми трендами в образовании в настоящее время являются перенос обучения в цифровую среду и его индивидуализация. Вопросы внедрения цифровых технологий в образование рассматриваются и решаются на самом высоком уровне: от Российского инвестиционного форума до послания Президента Федеральному собранию. Данная статья посвящена цифровизации образования как процессу, позволяющему решить проблемы, связанные с разработкой системы учебно-методического сопровождения, интегрирующую лучший республиканский, российский и международный опыт в области новых образовательных технологий.

**Ключевые слова:** цифровые технологии, цифровая образовательная среда, обучающиеся, виртуальная онлайн среда, международная дистанционная система обучения башкирскому языку.

Современная школа или вуз (и содержательно, и организационно) – это не чтение учебников и решение задач по базовым предметам, а обучение метанавыкам в сотрудничестве, использование современных технологий, смешивание форматов обучения, проектное и проблемное обучение, исследовательские методы и методы организации групповых дискуссий, игровые технологии.

Исследование воздействия современных технологий на образовательный процесс и их эффективного использования в ходе обучения рассматривается в работах отечественных исследователей М.А. Бовтенко, Р.К. Потаповой, Е.С. Полат, П.В. Сысоева, М.Н. Евстигнеева, С.И. Титовой, А.В. Зубова, Р.К. Пиотровской, Р.Г. Пиотровского, О.А. Усковой, С.В. Фадеева, А.В. Хуторского и др., а также западными исследователями G. Dudeney, M. Pegrum, D. Boyd, D. Crystal, M. Warshauer, G. Davis, P. Brett, M. Levy, D. Healey и др.

Е.А. Кашина отмечает: «Изменились требования к умениям обучающихся, поскольку необходимо не только читать, писать и считать, нужно уметь организовывать ресурсы данных, плодотворно сотрудничать, собирать, оценивать и использовать информацию» [Кашина 1997, 1].

Следует отметить, что необходимость внедрения современных технологий в процесс обучения в России признается и рядом государственных инициатив. Приоритетный проект «Современная цифровая образовательная среда в Российской Федерации» разработан Министерством образования и науки РФ и утвержден президиумом Совета при Президенте Российской Федерации по стратегическому развитию и приоритетным проектам (протокол от 25 октября 2016 №9).

В п.3 ст. 16 Федерального закона «Об образовании в Российской Федерации» от 29.12.2012 № 273-ФЗ 9 ред. от 03.08.2018) говорится о том, что «электронная информационно-образовательная среда включает в себя электронные образовательные ресурсы, совокупность информационных технологий, телекоммуникационных технологий, соответствующих технологических средств, обеспечивающих освоение обучающимися образовательных программ в полном объеме».

Технический прогресс привел к появлению более совершенных технических средств, обладающих значительным лингводидактическим потенциалом. Например, обучение языкам происходит при использовании широкого спектра технических средств, как стационарных (например, компьютер, интерактивная доска), так и мобильных (планшетные компьютеры, смартфоны), которые объединены под понятием «цифровые технологии». Цифровые технологии обеспечивают совершенно иную подачу информации, в которой используется мультимедийный формат. Способы передачи оформляются в нескольких формах одновременно: текстовой, графической, видео, аудиальной, интерактивной.

В связи с этим считаем, что необходимо дать определение понятий «цифровая образовательная среда» и «цифровые образовательные технологии», которые актуализируются в современном образовательном процессе.

«Цифровая образовательная среда является совокупностью цифровых образовательных ресурсов, средств и технологий, обеспечивающих образовательный процесс в условиях цифровизации» [Морозов 2018, 106].

Цифровые образовательные ресурсы – это информационные образовательные ресурсы, хранящиеся и передающиеся в цифровой форме, цифровые технологии – основанные на методах кодировки и передачи информации дискретной системы, позволяющая совершать большое количество разнотипных задач в короткие временные промежутки» [Морозов 2018, 106].

Интернет играет огромную роль в индивидуализации процесса получения информации, учебники и курсы, представленные на CD-дисках теряют актуальность. Цифровая образовательная среда позволяет выбирать собственную образовательную траекторию, которая ориентирована на получение необходимой информации для решения той или иной задачи.

Отечественные методисты-лингводидакты М.Н. Евстигнеев, Е.С. Полат, С.П. Сысоев, Л.А. Цветкова и др. считают, что использование мультимедийных и цифровых технологий в обучении имеют следующие преимущества:

- 1) доступность большого объема аутентичной информации;
- 2) воздействие мультимедийных технологий на каналы восприятия информации (текст, графика, звук, мультипликация, видео);
- 3) адаптивность;
- 4) нелинейность предоставления информации;
- 5) высокая включенность в учебный процесс [Гусевская 2012].

Цифровая среда требует разработки новых методических рекомендаций, совершенно иного способа и форм работы с обучающимися. Появляются новые возможности создавать и применять контент посредством цифровых технологий, включая навыки компьютерного программирования, поиска, обмена информацией, коммуникацию. Использование технологий в обучении сами по себе не делают процесс обучения более эффективным, в связи с этим возникает проблема, связанная с недостаточной разработанностью теоретических аспектов проблемы использования мультимедийных и цифровых технологий в процессе обучения языку в средней школе.

Для решения данной проблемы методистами разрабатываются методические рекомендации по использованию мультимедийных и цифровых технологий в обучении родным языкам в образовательных учреждениях.

Новые методики обучения языкам ориентированы на переход с отдельного компьютера на виртуальную онлайн-среду, так как компьютер играет роль посредника между человеком и виртуальной средой, создаваемой совместно миллионами пользователей. Например, методисты БГПУ им. М. Акмуллы совместно со специалистами It-технологий разработали дистанционную систему по обучению башкирскому языку.

Проект «Международная система дистанционного обучения башкирскому языку» стартовал в сентябре 2017 года, целью которого является обеспечение интересного изучения языка в интерактивной форме.

Данный портал находится в системе MOODL по адресу: [region.bspu.ru](http://region.bspu.ru).

Система состоит из следующих разделов (рис. 1):



Рис. 1. Начальная страница Международной системы дистанционного обучения башкирскому языку

В разделе «Башкирский язык для начинающих» обучающиеся регистрируются и могут начать обучение. По окончании обучения обучающиеся проходят тестирование, после успешного его прохождения получают электронный сертификат. Методика обучения башкирскому языку ориентирована на разговорный башкирский и интерактив: грамматика плавно соприкасается с ситуациями и диалогами. Видеоситуации созданы по определенному сценарию, где присутствует юмор и связь с жизнью народов Республики Башкортостан и Российской Федерации.

Остальные разделы носят познавательный формат, где есть шуточный тест, который может выполнить представитель любой страны. В разделе «Галерея» представлена ресурсная информация о жизни Республики Башкортостан и о его представителях.

В разделе «Уголки Башкортостана» в интерактивную карту республики пользователями размещаются фотографии прекрасных мест природы Республики Башкортостан.

Можно отметить то, что обучение языку при помощи цифровых технологий на данной платформе соответствует важнейшим принципам обучения:

1) принципу научности – за счет содержания обучающих программ, созданным согласно методическим правилам;

2) принципу сознательности – за счет создания возможности для обучающихся использовать индивидуальные стратегии достижения учебных результатов;

3) принципам активности и интерактивности – за счет изменения роли обучающегося с пассивного слушателя на инициатора работы, построенной на активном взаимодействии с компьютером;

4) принципам систематичности и последовательности – при условии интегрированности в программу обучения, компьютерные ресурсы могут обеспечить последовательность подачи учебного материала;

5) принципу прочности результатов обучения – за счет возможности многократно использовать обучающие программы и ресурсы как на уроке, так и самостоятельно в условиях внеклассной работы;

6) принципу наглядности – за счет применения обучающих средств различных форматов, включающих всевозможные виды вербальной и невербальной наглядности (картинки, видео, аудио и пр.) [Зубов 2009, 12].

Можно констатировать о том, что процесс цифровизации образования приведет к формированию цифровой (информационной) культуры обучающихся, которая позволит грамотно использовать открывающиеся возможности и органично встраиваться в цифровую образовательную среду.

### Литература

Гусевская Н. Ю., Ерёмина В.М. Обучение иноязычному общению студентов неязыковых специальностей на основе использования ИКТ и активных методов обучения // Учёные записки ЗабГУ. Серия: Профессиональное образование, теория и методика обучения. 2012. – № 6. – С.97–101.

Зубов А. В. Методика применения информационных технологий в обучении иностранным языкам / А.В. Зубов, И.И. Зубова. – Москва: Академия, 2009. – 140 с.

Кашина Е. А. Прогнозирование структуры интегрированного курса информатики : дисс. ... канд. пед. наук. — Екатеринбург, 1997. — 187 с.

Морозов В.А. Профессиональная подготовка руководителей системы образования с использованием современных цифровых технологий // Человек и образование. – 2018. – 4 (57) – С. 105–110.

Livingstone S. Online risk, harm and vulnerability: Reflections on the evidence base for child Internet safety policy / S. Livingstone // ZER: Journal of Communication Studies. – 2013d. – Volume 18. – № 35. – P. 13–28.

## ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ КОРПУСНОЙ ЛИНГВИСТИКИ В ПРАКТИКЕ ПРЕПОДАВАНИЯ УДМУРТСКОГО ЯЗЫКА

Статья посвящена анализу использования возможностей корпусной лингвистики при изучении родного языка на материале корпуса удмуртского языка. Анализируются способы изучения полисемии и синонимии в практике преподавания удмуртского языка на материале корпусного анализа.

**Ключевые слова:** корпусная лингвистика, корпус удмуртского языка, полисемия, синонимия, методика преподавания языков.

Одним из перспективных направлений современной лингвистики и лингводидактики является изучение языков через использование возможностей лингвистических корпусов. Данная статья посвящена описанию возможностей использования корпусной лингвистики в образовательном процессе на материале удмуртского языка.

Термин «*корпусная лингвистика*» используется в такой отрасли как компьютерная лингвистика, занимающейся разработкой общих принципов построения и использования лингвистических корпусов на основе компьютерных технологий [Захаров 2011, 7]. Корпусная лингвистика изучает языковые закономерности через большой объем примеров использования языковых единиц (слов, словосочетаний, предложений и т.д.), реально функционирующих в письменной или устной форме. Сегодня существуют различные виды языковых корпусов. Наиболее крупными считаются корпуса русского языка (Национальный корпус русского языка, <http://www.ruscorpora.ru>, содержит 500 млн слов русского языка), английского языка (British National Corpus, <http://corpus.byu.edu/bnc/>, содержит 100 млн слов британского варианта английского языка; Corpus of Historical American English (COHA), содержит 450 млн слов американского варианта английского языка), немецкого языка (корпус Берлинской Бранденбургской академии наук (DWDS), <https://www.dwds.de/>, содержит 106 млн слов немецкого языка).

Сегодня корпусная лингвистика используется для создания корпуса языков малочисленных народов. Так, например, существует корпус удмуртского языка, разработанный Йормой Луутоненом, Леонидом Ившиным и Кари Аро в университете г. Турку (Финляндия). В корпус входят около 11000 текстов из удмуртских газет и журналов «Удмурт дунне», «Кенеш», «Вордскем кыл», «Инвожо», «Зечбур» и «Кизили». Также существует платформа корпуса удмуртского языка, разработанная Тимофеем Архангельским, Марией Медведевой и Екатериной Георгиевой в рамках проекта в Школе лингвистики национального исследовательского университета «Высшая школа экономики» в 2014 году.

В образовательном процессе языковой корпус широко используется для формирования и развития лексико-грамматических навыков учащихся в практике преподавания языков, при изучении лингвистических дисциплин, а также для сбора фактического материала при выполнении исследовательских проектов. Опираясь на данные языковых корпусов, студенты могут определить частотность использования языковой единицы, описать ее семантическую и грамматическую структуру, дифференцировать значения синонимов и т.д.

В статье описаны возможности использования языкового корпуса, разработанного Т. Архангельским, М. Медведевой и Е. Георгиевой, в практике преподавания удмуртского языка для работы над определением семантики слов. Так, например, анализ данных, полученных в результате работы с корпусом удмуртского языка, показывает, что существительное *гур* используется в значениях «печка», «мелодия», а также в роли междометия «гур-р-р», обозначая шум, исходящий от работающей техники: удм. Сыче коркаосын вылэм муръётэм *гур*. – рус. В таких домах была *печка* (*здесь и далее – перевод авторов статьи*) без дымохода; удм. Чузьяске спортсменъёсты огазеясь *гур*. – рус. Звучит *мелодия*, объединяющая спортсменов; удм. *Гур-р-р!* потто моторъёс, ёшето станокъёс. – рус. *Гур-р-р!* шумят моторы и станки.

Интересны случаи использования различных синонимов в предложениях. Так, например, синонимы *трос*, *уно*, *тачак*, *лачак* в значениях «много, полно, очень много» могут заменять друг друга в предложениях: удм. Каргам война быдтїз туж *трос* калыкез, *уно* семьялэсь тарказ шудо улонзэс. – рус. Проклятая война погубила очень *много* людей, у *многих* семей не сложилась счастливая жизнь; удм. Котыр гурезьёс, узьез но, борьез но, губиез но, эмезез но *лачак* отын. – рус. Вокруг горы, и земляники, и клубники, и малины там *много/полно*; удм. Башкортостанын удмурт кылын вераськисьёс *тачак*. – рус. В Башкортостане *очень много* людей, говорящих на удмуртском языке. Студенты могут использовать эти синонимы в своей речи, делая ее более разнообразной и интересной.

В то же время анализ значений этих слов с опорой на корпусный анализ показывает, что имеются нюансы, дифференцирующие их значения. В первую очередь, необходимо отметить, что слова *трос*, *уно* встречаются чаще, чем *лачак* и *тачак*. *Трос* встречается в 24984 предложениях и примерно в 4563 документах, *уно* – в 3776 предложениях примерно в 1603 документах, *лачак* – в 471 предложениях в 367 документах, *тачак* – в 153 предложениях в 134 документах. Использование стилистически нейтральных *трос* и *уно* в большей степени продиктовано тем, что родной диалект автора текста может влиять на его письменную литературную речь. Необходимо также отметить, что слова *лачак* и *тачак* имеют более экспрессивную окраску, в связи с чем чаще используются в разговорной речи; *тачак* характерно для некоторых диалектов. Корпусный анализ показывает, что *трос* и *уно* могут употребляться с наречиями *туж* «очень» или *укыр* «слишком» в значении «очень много», «слишком много»: удм. Синмаськымон интыос *туж трос* ке но, Удмурт шаермылэсь дуноез номыр ёвёл кадъ. – рус. Есть *очень много* красивых мест, но дороже нашей Удмуртской земли, наверное, ничего нет; удм. Герман Алексеевич улытояз *туж уно* гуртьёсы, школаосы, нылпи садъёсы вуылйз. – рус. Герман Алексеевич на протяжении всей своей жизни посетил *очень много* деревень, школ, детских садов; удм. *Укыр трос* юанъёс сётъяськоды... – рус. *Слишком много* вопросов задаете. Выражения *туж лачак* и *туж тачак* являются некорректными, и конкорданс (перечень всех употреблений заданного слова или выражения в контексте) подтверждает это – не выдает ни одного примера.

Корпусный анализ отчетливо показывает различия между значениями слов *вераны* «говорить, сказать» и *шуыны* «сказать». Глагол *вераны* чаще используется в значении «сообщать информацию, рассказать»: удм. Со пуке диван вылын, маке *вераны* тырше. – рус. Он сидит на диване, пытается что-то *сказать*; удм. Дорама бригадир лыктйз но вормон сярись ивор *вераз*. – рус. К нам подошел бригадир и *рассказал* про победу. Основной функцией глагола *шуыны* является введение прямой речи: удм. Иисус солы *шуиз*: «Вай Мыным юыны». – рус. Иисус *сказал* ему: «Дай Мне попить»; удм. Таће шулдырьяськонэз вань пиналъёс чидатэк витё, – *шуиз* нылпи садэн кивалтйсь Надежда Баранова. – рус. Такого праздника с нетерпением ждут все дети, – *сказала* руководитель детского сада Надежда Баранова. Для введения прямой речи может использоваться и глагол *вераны*: удм. Зэм, та югдурен нюръяськыны сокем ик капчи ёвёл, – мылкыдзэ *вераз* В. Путин. – рус. Да, с этой проблемой бороться не так уж и легко, – *сказал* В. Путин.

Глагол *шуыны* также используется в значении «произнести»: удм. Туннэ ушъяськеме потэ тани ма сярись, – журналистъёсын пумиськылыкуз, *шуиз* Андрей Макаров. – рус. Сегодня хочу похвастаться вот чем, – *произнес* Андрей Макаров на встрече с журналистами.

В удмуртском языке имеется ряд устойчивых выражений с глаголами *вераны* и *шуыны*. В некоторых контекстах эти слова взаимозаменяемы и используются в таких выражениях, как *шуыны луоз/вераны луоз* (можно сказать), *шуыны кулэ/вераны кулэ* (надо сказать): удм. Соку мон нырысьсэ тодматски, *шуыны луоз*, зэмос джазэн. – рус. Тогда я впервые познакомилась, *можно сказать*, с истинным джазом; удм. *Вераны луоз*, тросэз кырзасьёс – егитъёс. – рус. *Можно сказать*, большинство певцов – молодежь. При этом выражение *шуыны луоз* (встречается в 1186 предложениях примерно в 784 документах) превалирует над выражением *вераны луоз* (108 предложениях примерно в 102 документах). Противоположное явление наблюдается с выражениями *шуыны кулэ/вераны кулэ*, в этом случае выражение *вераны кулэ* встречается в 1116 предложениях и примерно в 915 документах, тогда как *шуыны кулэ* – в 48 предложениях и примерно в 46 документах.

В других случаях границы коллокаций четко очерчены, и одно слово не может быть заменено другим. Так, например, глагол *вераны* используется в выражениях *шуг вераны* (сложно сказать) и *вераны луонтэм* (нет таких слов, чтобы сказать) в таких примерах, как – удм. Маин со герзаськемын, *шуг вераны*. – рус. С чем это связано, *сложно сказать*; удм. Со залъёс *вераны луонтэм* чебересь. – рус. *Нет таких слов, чтобы сказать*, насколько красивы эти залы. Корпусный анализ показывает, что в этих выражениях глагол *вераны* не может быть заменен глаголом *шуыны*.

Отличия между семантикой и сочетаемостью этих слов наблюдаются и в таких коллокациях, как *тау вераны* и *тау шуыны* (сказать спасибо): удм. Удмурт кылын дышетйсьлы Елена Герасимовналы *тау вераны* кулэ. – рус. Учителю удмуртского языка Елене Герасимовной надо *сказать спасибо*; удм. Анаелы *тау шуыны* кулэ: со монэ крезьгур дуннее кыскиз. – рус. Моей маме надо *сказать спасибо*: она меня привлекла в мир музыки. Частотность использования вышеуказанных выражений, безусловно, отличается: выражение *тау шуыны* встречается в 51 предложении и примерно в 44 документах, в то время как *тау вераны* обнаружено только в 3 предложениях примерно в 3 документах.

Таким образом, использование возможностей лингвистического корпуса в образовательном процессе совершенствует процесс изучения удмуртского языка, позволяя студентам работать с аутентичными материалами. При работе со студентами могут быть использованы различные

упражнения, в том числе упражнения, направленные на определение значений многозначного слова в контексте и дифференциацию значений синонимов. Упражнения этого рода могут быть использованы при работе с различной целевой аудиторией, однако наибольшая практическая ценность этих упражнений заключается в возможности их использования в практике преподавания удмуртского языка как неродного (второго, иностранного), ибо корпусный анализ позволяет зримо увидеть нюансы значений слов, незаметные для не носителей языка и не всегда отражаемые в словарях и другой учебной литературе.

### **Литература**

Захаров В.П., Богданова С.Ю. Корпусная лингвистика: учебник для студентов гуманитарных вузов. – Иркутск: Изд-во Иркут. гос. лингв. ун-та, 2011. – 161 с.  
<http://udmurt.web-corpora.net> (дата обращения: 25.11.2019)



## ИЗУЧЕНИЕ ФРАЗЕОЛОГИЗМОВ В ПРОЦЕССЕ ОБУЧЕНИЯ ИНОСТРАННОМУ ЯЗЫКУ В НАЧАЛЬНОМ КЛАССЕ

В данной статье рассматривается роль фразеологических единиц при обучении начальных классов иностранному языку. Особое внимание уделяется фразеологизмам как одному из показателей уровня владения иностранным языком. Автором предлагается перечень фразеологических единиц, которые могут быть использованы в ходе учебного процесса.

**Ключевые слова:** иностранный язык, фразеологизм, лексический запас, обогащение речи.

Перед учителем иностранного языка в начальной школе стоит такая огромная задача, как пробудить интерес у обучающихся к языку и формировать базовые знания, которые станут опорой для дальнейшего изучения языка.

По требованиям Федерального государственного образовательного стандарта начального общего образования, обучение иностранному языку в начальной школе должна быть ориентирована на применение знаний, умений и навыков в элементарных учебных ситуациях и реальных жизненных условиях и отражать формирование иноязычной коммуникативной компетенции на элементарном уровне владения (А1) в совокупности ее составляющих – речевой, языковой, социокультурной, компенсаторной, учебно-познавательной. Так же имеет первостепенное значение освоение начальных лингвистических представлений, необходимых для овладения на элементарном уровне устной и письменной речью на иностранном языке, расширение лингвистического кругозора; формирование дружелюбного отношения и толерантности к носителям другого языка на основе знакомства с культурой других стран...» [URL: <http://nova56.ru/secondary/phrazeduc>]

Для достижения этих результатов необходимо владение устойчивыми сочетаниями изучаемого языка, так как в процессе возникновения и формирования языка определенного народа огромное значение имеют его менталитет, традиции, устои. Следовательно, обучение иностранному языку предполагает проникновение в культурные особенности языка. В этом смысле особую роль играют фразеологические единицы, в которых отражается специфика, самобытность, жизненный уклад, культура, многовековая история определенного народа. Каждая фразеологическая единица нередко олицетворяет определенные события. Например, история возникновения фразеологизма *Give a cold shoulder* (буквально: дать холодное плечо) которое интерпретируется как «*быть неприветливым с кем-то, вести себя грубо по отношению к человеку*» связана с обычаями, присущий средневековой Англии. В те времена там существовал весьма необычный способ намекнуть гостю, что он, слегка задержался, и ему следовало бы пойти домой. Ему давали холодное мясо с плеча мутона или другого животного. При этом, данный намек считался вежливым [URL: <https://tryeng.ru/1820>]

Фразеологизмы это – сокровища языка. Их ценность определяется еще тем, что в фразеологических единицах резервируется информация об определенных традициях, которые остались в глубине веков и возможно уже и не существуют и сохранились лишь в языке.

Однако следует отметить, что фразеологизмы – значительно более сложные образования, чем слова и словосочетания. Их смысл не складывается из значений составляющих его слов. Возникает необходимость умения абстрагировать, а абстрагирование в своем роде играет важную роль в умственной деятельности детей, т.к. на её основе происходит обобщение и формирование понятий. Для преодоления барьеров, связанных понятием значения устойчивых сочетаний, необходимо постепенно начать их усваивать уже в самом раннем этапе изучения языка. Кроме того, формирование речевых умений по фразеологии требует систематизированной работы. Поэтому в процессе обогащения речи обучаемых младших классов должна соблюдаться определенная последовательность. После того, как обучающиеся начнут использовать эти устойчивые словосочетания в своей речи, откроются большие возможности для расширения их лексикона, а также будет развиваться их культура речи.

Фразеологизмы могут быть использованы на различных этапах урока. Например, при прохождении таких тем, как

### 1. «Спорт»:

- Make a come back – взять реванш
- Second wind – второе дыхание, новый прилив энергии
- The hat trick – тройной успех, три забитых мяча (футбол, крикет, скачки)
- Toe the line – подчиняться требованиям, строго придерживаться правил

- Front runner – лидер гонок; опережающий своих конкурентов и др.
- Even the score – сравнять счёт; занять равное положение
- From start to finish – от начала и до конца
- Send smb. to grass – повалить кого-то, сбить с ног

#### 2. «Дом»:

- Make yourself at home – почувствуйте себя как дома
- Home is where the heart is – дом там, где твое сердце
- Make oneself at home – чувствовать себя как дома
- Keep house – хозяйничать, вести домашнее хозяйство
- Keep one's house – сидеть дома, не выходить из дома

#### 3. «Семья»:

- Shoulder to shoulder – плечо в плечо (вместе жить)
- Head of the family – глава семьи
- Wooden wedding – деревянная свадьба
- A family skeleton – семейная тайна

В целом, поле использования фразеологических единиц достаточно широкое. Только важно учитывать тот факт, что они должны быть подобраны в соответствии с уровнем владения языком обучающихся. И не менее важным является то, что обучаемые могли бы использовать эти подобранные фразеологические единицы и в дальнейшем, не только в конкретном уроке. Например, в разных коммуникативных ситуациях, при составлении диалогов, написания эссе и т.д. Ну и безусловно, несли воспитательный характер.

Итак, мы провели обзор о необходимости изучения фразеологизмов в процессе обучения языкам в начальном классе и сделали следующие выводы: во-первых, усвоение фразеологических единиц способствует обогащению лексики, разговорной речи обучаемых. Во-вторых, устойчивые словосочетания играют важную роль в лингво-культурологическом аспекте, то есть фразеологические единицы тесно связаны с определенными историческими нравами и традициями конкретного народа и помогают ознакомиться с культурой изучаемого языка. И в-третьих, фразеологизмы носят воспитательный характер и способствуют развитию абстрактного мышления у обучающихся начальных классов.

### Литература

Кунин А.В. // Курс фразеологии современного английского языка: Учеб. для ин-тов и фак. иностр. яз.– 2-е изд., перераб.– М.: Высш. шк., Дубна: Изд. центр «Феникс», 1996 – 381 с.

Происхождение английских фразеологизмов [Электронный ресурс]. URL: <https://tryeng.ru/1820> (дата обращения 06.10.2019).

Работа с фразеологизмами на уроках в начальной школе как средство овладения национальной спецификой языка [Электронный ресурс]. URL: <https://infourok.ru/rabota-s-frazeologizmami-na-urokah-v-nachalnoy-shkole-kak-sredstvo-ovladieniya-nacionalnoy-specifikoy-yazika-538332.html> (дата обращения 06.10.2019).

Роль изучения английских идиом в расширении лексического запаса школьников. [Электронный ресурс]. URL: <https://cyberleninka.ru/article/v/rol-izucheniya-angliyskih-idiom-v-rasshirenii-leksicheskogo-zapasa-shkolnikov> (дата обращения 24.10.2019).

Федеральный государственный образовательный стандарт начального общего образования (новая редакция) [Электронный ресурс]. URL: <http://www.zhg-zresurs.ru/index.php/fgos-fgos-noo/fgos-noo-norms/item/2153-> (дата обращения 23.10.2019).

Фразеология при обучении иностранным языкам [Электронный ресурс]. URL: <http://nova56.ru/secondary/phrazeduc> (дата обращения 06.10.2019).

## МЕЖДИСЦИПЛИНАРНЫЕ ОБРАЗОВАТЕЛЬНЫЕ ПРОЕКТЫ КАК УСЛОВИЕ РАЗВИТИЯ НАЦИОНАЛЬНЫХ ЯЗЫКОВ

В статье представлены механизмы организации и проведения образовательных проектов, инициированных Институтом удмуртской филологии, финно-угроведения и журналистики Удмуртского государственного университета и направленных на развитие удмуртского языка и культуры.

**Ключевые слова:** удмуртский язык, институт удмуртской филологии, финно-угроведения и журналистики, интеллектуальный челлендж, научно-исследовательская работа.

Удмуртская Республика – один из многонациональных субъектов Российской Федерации, где проживает более 130 различных национальностей. Удмуртский язык как язык титульной нации имеет статус государственного. Несмотря на это, согласно официальным данным, происходит ежегодное снижение количества носителей удмуртского языка, удмуртский язык остается востребованным в узкой среде профессионального сообщества [Кондратьева, Юмина 2019]. С учетом вышеперечисленного, а также еще целого ряда критериев функционирования удмуртского языка, в 2009 году организацией ЮНЕСКО удмуртский язык был отнесен к языкам, подверженным угрозе исчезновения.

Учитывая вышесказанное, Институт удмуртской филологии, финно-угроведения и журналистики Удмуртского государственного университета в качестве одной из стратегических задач своего развития определил повышение имиджа удмуртского языка среди подрастающего поколения через формирование научно-исследовательского и олимпиадного движения в области удмуртского языка и литературы. основополагающими принципами деятельности института в рамках указанного направления стали системность, преемственность и междисциплинарный подход [Кондратьева 2018]. Указанные принципы, на наш взгляд, являются наиболее оптимальными в условиях развития современного образования [Артемьева, Мильруд 2005; Ковалева 2019].

В качестве основных задач указанного направления деятельности были определены следующие:

- отработка механизмов взаимодействия школ с вузами с целью модернизации и интеграции учебного процесса;
- повышение уровня знаний школьников;
- информационно-профориентационная деятельность, направленная на осознанный выбор обучающимися выпускных классов специальностей, связанных с удмуртским языком и культурой;
- повышение мотивации школьников к обучению по социально-гуманитарному блоку дисциплин и др.

Независимо от содержания проектов, любые формы взаимодействия школы и вуза требуют определенных ресурсов. Для реализации наших проектов большую помощь оказывают студенты. Рассмотрим механизмы организации студенческой активности на примере проведения Интеллектуального челленджа «Удмуртия: мой путь к успеху» (от англ. *challenge* «вызов», «задача», «сложная проблема»). Указанное мероприятие – это интеллектуальный турнир среди учащихся среднего звена: на базовой площадке собираются команды (обычно – 7-9 команд из разных школ). Каждая команда состоит из 2-х учащихся 5-х классов, 2-х учащихся 6-х классов и 2-х учащихся 7-х классов. Такой состав команд позволяет обеспечить преемственность развития интеллектуального движения и формирует навыки soft-skills.

Очный раунд организуется по форме и технологии игры «Что? Где? Когда?» и включает следующие разделы:

### А. Географический диктант

Командам предлагается письменно ответить на 10 вопросов из предметной области «География Удмуртии» (12 минут).

### Б. Экономическая задача

Командам предлагается решить 6 экономических задач (12 минут)

### В. Экология и биология

Командам предлагается выполнить практическое задание (например, определить следы животных, определить названия деревьев и др.) (12 минут)

Все задания выполняются в команде. Также оценивается выполнение домашнего задания. В частности, это может быть составление стенгазеты, тематика которой связана с популяризацией знаний об известных людях Удмуртии или района (политиков, бизнесменов, экономистов, спортсменов, ученых, людей творческих специальностей и др.). Главная задача этого этапа – показать

подростающему поколению возможности для творческого роста в Удмуртской Республике, т.к. в последние годы молодежь уже со школьной скамьи стремится уехать из региона.

В условиях загруженности профессорско-преподавательского состава вуза, вся организация мероприятия переложена на студентов трех институтов: института удмуртской филологии, финно-угроведения и журналистики (составление сценария, подготовка площадки проведения, работа в жюри, техническое сопровождение, подготовка концертных номеров, фото- и видео-сопровождение – задействовано не менее 18 человек), института естественных наук (составление вопросов к раунду «Географический диктант» и «Экология и биология Удмуртии», проведение своего раунда – задействовано 3-4 студента), института экономики и управления (составление экономических задач, проведение своего раунда – задействовано 3-4 студента).

Чтобы данное мероприятие имело большой охват, ежегодно выбираются базовые площадки. В частности, в 2019 году такими площадками стали Малопургинский Центр детского творчества, Карамас-Пельгинская СОШ и Гимназия № 56 г. Ижевска. Задача базовых площадок – подготовка площадки для проведения районного этапа игры, набор команд, обеспечение безопасности на время игры, взаимодействие с УдГУ. Как показал опыт, главным фактором успеха I сезона интеллектуального челленджа стало оптимальное распределение функций и ресурсов между организаторами мероприятия.

Необходимо подчеркнуть, что это не единственный проект Института удмуртской филологии, финно-угроведения и журналистики в деле сохранения и популяризации удмуртского языка. Уже более 10 лет успешно работает межрегиональный конкурс исследовательских и проектных работ школьников в области удмуртской филологии, фольклора, краеведения, журналистики, экономики и предпринимательства «Мои первые шаги в науку», олимпиада по английскому и немецкому языкам для учащихся, изучающих удмуртский язык. Второй год организуется интеллектуальное многоборье имени Юрия Семеновича Перевощикова (олимпиада для учащихся 7-11 классов в трех предметных областях: удмуртский язык, экономика, обществознание).

Все указанные мероприятия проводятся при активном включении студентов университета. Несомненно, подобная активность способствует формированию у молодежи организаторских и коммуникативных компетенций.

Считаем, что в реализации разработанных моделей особую значимость приобретает взаимодействие вуза и школы, предполагающее следующие научно-практические основания в сопровождении одаренных детей: поддержка творческого потенциала подрастающего поколения (как у школьников, так и студентов), учебно-методическая и научно-исследовательская работа с учителями; научное консультирование; разработка и апробация совместных проектов и др. Именно взаимодействие ВУЗа и школы, формируя систему ценностей, способствует взаимоактивизации, интегрирует положительное влияние на единое образовательное пространство.

Следующим этапом развития указанных выше проектов, несомненно, должно стать их включение в процессы цифровизации: создание образовательной электронной среды с размещением выполненных конкурсных заданий на удмуртском языке; создание и размещение на этой площадке видеороликов на удмуртском языке в рамках тематических конкурсов; возможности проходить он-лайн олимпиады и квизы на удмуртском языке. Именно развитие цифровых технологий является необходимым условием в деле сохранения и развития национальных языков Российской Федерации.

### Литература

Артемьева О. А., Мильруд Р. П. Методология организации профессиональной подготовки специалиста на основе межкультурной коммуникации. – Тамбов : Изд-во Тамб. гос. техн. ун-та, 2005. – 160 с.

Ковалева – Ковалева А. Г., Анчугова О. В., Дымова Е. Е., Курманова Д. И. Методика междисциплинарных проектов на иностранном языке // Педагогическое образование в России. – 2019. № 7. – С. 76–83 [Электронный ресурс]. URL: <http://elar.uspu.ru/bitstream/uspu/12768/1/povr-2019-07-10.pdf> (дата обращения: 03.12.2019).

Кондратьева Н. В. Об успешных практиках реализации языковой политики в Удмуртской Республике // Языковая политика: общероссийская экспертиза. – М.: Федеральное агентство по делам национальностей, 2018. – С. 122–128.

Кондратьева Н. В., Юмина Т. Ф. Реализация этнокультурного компонента в условиях городской образовательной среды // Приоритеты стратегии научно-технологического развития России и обеспечение воспроизводства инновационного потенциала высшей школы – Ижевск: Изд. дом «Удмуртский ун-т», 2019. – С. 268–271.

Мартынова Т. М. Использование проектных заданий на уроках английского языка // Иностранные языки в школе. – 1999. – № 4. – С. 19–21.

## ИЗУЧЕНИЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ В ВУЗЕ ДЛЯ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ УЧИТЕЛЯ РОДНОГО ЯЗЫКА И ЛИТЕРАТУРЫ

В статье раскрыты особенности изучения информационных технологий на примере дисциплин бакалавриата и магистратуры кафедры коми филологии, финно-угроведения и регионоведения Сыктывкарского государственного университета им. Питирима Сорокина.

**Ключевые слова:** бакалавриат, магистратура, информационные технологии.

На сегодняшний день изучение информационных технологий входит во все программы высших учебных заведений. Своя специфика имеется в каждом направлении бакалавриата и в программах магистратуры. Некоторые особенности информационных технологий для гуманитария описаны в книге А.Т. Хроленко [Хроленко 2007]. В нашей работе мы попытаемся раскрыть опыт работы кафедры коми филологии, финно-угроведения и регионоведения Сыктывкарского государственного университета им. Питирима Сорокина. В последние пять лет кафедра ведёт обучение студентов по направлению подготовки 44.03.05 Педагогическое образование (с двумя профилями), профили Родной язык и литература и иностранный язык (английский). Квалификация выпускника бакалавр. А также осуществляется обучение по направлению подготовки 45.04.01 Филология. Направленность (профиль) программы «Финно-угорские языки в поликультурном пространстве» (магистратура). Первый выпуск бакалавров данного направления ожидается в 2020 году, поэтому можно признать инновационность, экспериментальность всего периода обучения, сопровождающееся мобильными изменениями в учебном плане. Изучение современных информационных технологий проводится на 4-5 курсах бакалавриата (дисциплина «Новые информационные технологии в филологическом образовании») и на I курсе магистратуры (дисциплина «Информационные технологии в профессиональной деятельности»). Также в магистратуре параллельно проходит курс «Финно-угорские языки в электронном информационном пространстве» [Федина 2015, а]. Основную цель данных дисциплин можно определить как формирование педагога-филолога, владеющего знаниями и умениями в области новых информационных технологий. Задачи курса: формирование четкого представления о технологических процессах сбора, хранения, передачи и обработки информации; знакомство с профессиональным программным обеспечением – системами обработки текстов, графики, автоматизированными информационными системами; приобретение практических навыков по обработке текстовой и числовой информации, использованию мультимедийных технологий, представления информации, гипертекстовых способов хранения и представления информации. В ходе освоения дисциплин по изучению НИТ студенты знакомятся с понятием ИТ, электронными библиотеками, национальным корпусом текстов, электронными словарями, цифровыми ресурсами по коми языку и литературе, с коми версией *Wiki* (изучение структуры сайта, регистрация участника, создание личной страницы), программой АBBYY FineReader. Все изложенные ресурсы либо находятся на портале *Fu-lab.ru*, либо тесно с ним связаны (*Wiki*). Портал *Fu-lab.ru*, созданный и разрабатываемый в течение ряда лет усилиями Центра инновационных языковых технологий ГАУ РК «Дом дружбы народов Республики Коми», стал инновационной и базовой площадкой в разработке совершенно новых ресурсов на коми языке. Эта тема наиболее полно раскрывается руководителем центра М.С. Фединой во многих интервью и статьях [Федина 2015]. Студенты с удовольствием знакомятся с новыми ресурсами, принимают посильное участие в их создании (озвучивание, набор и редактирование). Пожалуй, самыми востребованными на портале являются онлайн словари (русско-коми и коми-русский). Самой же сложной операцией, требующей усидчивости и грамотности, является редактирование текстов программе АBBYY FineReader. Среди заданий к зачету или экзамену предложены:

1. Регистрация участника коми версии *Wiki*, создание личной страницы, добавление 20 текстов.
2. Сканирование и редактирование текстов (100 стр.) в программе АBBYY FineReader.

Для магистрантов – создание учебных тестов в программе ADTester 2.88.4. (50 вопросов) по выбранной дисциплине в соответствии с направлением магистратуры. Данный комплекс тестирования состоит из трех условно независимых частей: программы тестирования знаний, программы создания тестов и программы администрирования. Программа тестирования позволяет определить уровень знаний ученика в той или иной области. После прохождения теста ученику выставляется оценка, вся информация о прохождении теста записывается в журнал. Программа администрирования позволяет учителю создавать, изменять или удалять группы пользователей (аналог классов), редактировать

карточки пользователей, просматривать результаты тестирования и проводить анализ ответов учащихся. Программа создания тестов служит для подготовки новых тестов. С ее помощью можно создать тест, включить в него критерии оценки, установить временные рамки, – установить все параметры тестирования. В тесте могут встречаться вопросы следующих четырех типов: 1) одиночный выбор – пользователь может выбрать только один вариант ответа; 2) множественный выбор – пользователь выбирает один или несколько вариантов ответа; 3) ввод ответа с клавиатуры – тестируемый вводит ответ с клавиатуры; 4) соответствие – тестируемому необходимо сопоставить элементы 2-х списков. К сожалению, кроме явных преимуществ, данная программа имеет большой недостаток – она представляет собой русскую версию, даёт возможность работать только кириллицей.

Весь курс изучения НИТ сопровождается самостоятельной подготовкой аннотаций, конспектов, докладов, презентаций, рецензий на сайты, написанием эссе на сформулированную самим студентом тему по дисциплине, заканчивается защитой собственного проекта по использованию или созданию НИТ.

Информационные технологии характеризуются постоянным количественным и качественным обновлением, пользователю невозможно «объять необъятное». Сотрудничество кафедры с ЦИЯТ даёт возможность использовать активные методы освоения студентами дисциплин по изучению новых информационных технологий, почувствовать социальную значимость проделанной работы. Сочетание данной деятельности с заданиями, развивающими научно-исследовательские компетенции студентов и магистрантов, позволяют создавать самим студентам новые, необходимые, по их мнению, ресурсы на родном языке.

### **Литература**

Федина М.С. Финно-угорские языки в электронном информационном пространстве: учеб. пособие / М.С. Федина, А.В. Чемышев, В.А. Степанов. – Сыктывкар: ГОУ ВО КРАГСиУ, 2015. – 96 с.

Федина М.С. Центр инновационных языковых технологий: настоящее и будущее, перспективы и проблемы // Арт. №3. 2015. – С.130 – 134.

Хроленко А.Т., Денисов А.В. Современные информационные технологии для гуманитария: практическое руководство. – М.: Флинта, 2007. – 128 с.

## МУЛЬТИМЕДИЙНЫЕ УЧЕБНЫЕ ПОСОБИЯ ПО ИСТОРИИ РОССИИ И ОБЩЕСТВОЗНАНИЮ НА КОМИ ЯЗЫКЕ

В статье описывается структура и содержание мультимедийных учебных пособий по истории России и обществознанию на коми языке, созданного в программе Macromedia Flash MX. Пособия содержат учебный текст, переведённый на коми язык и озвученный автором пособий, обучающие и контрольные тесты с ответами, рабочие тетради, словарь, наглядный материал. Данные образовательные продукты позволяют расширить функционирование коми языка в гуманитарно-образовательной сфере.

**Ключевые слова:** история, обществознание, мультимедийное, тесты, рабочая тетрадь.

Цель данной статьи – описать авторские мультимедийные образовательные продукты по истории России и обществознанию на коми языке для общеобразовательных школ. Оба ресурса подготовлены по современным программам для школ России. Мультимедийное учебное пособие по истории России предназначено для 6 класса. Мультимедийное учебное пособие по обществознанию представляет собой полный курс для основной школы.

Первым шагом к их подготовке стала разработка и издание в 2011 терминологических словарей по истории и обществознанию для общеобразовательных школ в рамках программы Совета Европы и Европейского Союза для РФ «Национальные меньшинства в России: развитие языков, культуры, СМИ и гражданского общества» [Остапов 2011], [Остапов 2011].

Следующим важным шагом стала разработка совершенно новых, уникальных образовательных ресурсов на коми языке по предметам «История России. 6 класс» и «Обществознание». Они помещены на сайте <https://fu-lab.ru/uchposobiya>.

Образовательные продукты по истории и обществознанию создавались в программе MacromediaFlashMX и опубликованы в папках как совокупность определённым образом связанных друг с другом swf-файлов с запускающим История.exe и Обществознание.exe файлами. Общепризнано, что достоинства таких файлов следующие:

1. При работе с swf-файлами можно не беспокоиться об установке специальных программ для работы из-за наличия повсеместно распространённого сегодня на персональных компьютерах флеш-плеера для работы с такого рода продуктами. Файлы с таким расширением уже давно широко используются для создания рекламных баннеров, анимации, игр, а также воспроизведения на веб-страницах видео- и аудиозаписей.

2. Flash-файлы, как правило, небольшого размера, поэтому не требуют к компьютерам особых требований к производительности.

3. Материалы с расширением .swf, могут быть размещены в Интернете и просмотрены с помощью большинства интернет-браузеров на любой компьютерной платформе. Такой файл можно легко отправить по почте, парой щелчков мыши может быть развёрнут на весь экран.

4. Разделённые на отдельные файлы Flash-материалы несложно использовать как в предложенной нами системе, так и отдельно друг от друга, запуская каждый файл самостоятельно. При изучении материала у обучающегося появляются альтернативы: либо следовать заданному автором пособия алгоритму последовательных учебных действий при изучении тем, построенных на основе проблемно-хронологического принципа, либо, по своему выбору, знакомиться фрагментарно с отдельными историческими периодами, персонажами, проблемами.

В предлагаемом образовательном ресурсе заложена как линейная, так и нелинейная навигация. С файлов распределителей по запросу обучающегося можно открыть необходимый элемент напрямую в зависимости от возникших потребностей и вернуться к исходному положению для изучения локальной проблемы (например, карты страны), повторения даты события или получения кратких сведений о деятеле. Нами намеренно не были включены в материалы видеофильмы и анимации, так как они значительно увеличили бы размеры созданного ресурса и могут привести к существенному повышению требований к системным требованиям персонального компьютера пользователей.

Мультимедийное учебное пособие по истории России создано в 2014 г. Пособие по обществознанию подготовлено в 2018 году, является инструментом обучения, содержащим краткое систематическое изложение всего обязательного материала по обществознанию для 5-7 классов. В обоих ресурсах представлены в определённой системе учебно-дидактические материалы. Учебные тексты,

исторические карты, терминология, хронология, исторические портреты, разработанные задания по основным умениям и видам познавательной деятельности должны создать условия для самостоятельных занятий по предмету. Дополнительная сложность заключалась в том, что подобного рода ресурсы по истории и обществознанию на коми языке вообще отсутствуют в образовательном поле РК. В то же время за плечами был немалый опыт коллективной работы в проектах по созданию электронных пособий по коми литературе. Подробнее данный процесс освещается в статье Остаповой Е.В. [Остапова 2014, 267-274] и коллективной статье сотрудников кафедры финно-угорской филологии и национального образования СГУ им. Питирима Сорокина [Попова 2014, 323-332].

Собранные нами материалы являются вспомогательным инструментом обучения, содержащим краткое систематическое изложение событий по истории России с VIII века до XIII века и обществознанию.

В образовательном продукте содержатся:

I. Организованные текстовые материалы, в которых концентрированно изображены все ключевые факты, явления, процессы. Предложенные тексты ученику можно преобразовать в структурно-логическую схему, конспект, опорные знаки, что позволяет без помощи учителя понять и воспроизвести материал, уяснить взаимосвязи между фактами, событиями для получения представления общей динамики исторического процесса. При знакомстве с текстами для обучающегося доступны сопутствующие справочные или дополнительные материалы (исторические карты, термины, карты, даты событий, портреты деятелей). Обучающийся имеет возможность в любой момент переключиться на определённое время к сопутствующим сведениям справочного характера и обратно вернуться на нужную страницу простым перемещением или щелчком компьютерной мыши. Представленный материал позволяет варьировать темпы усвоения учебной информации в соответствии с индивидуальными особенностями учеников, а также обращаться к расширяющим, углубляющим знания дополнительные ресурсы.

II. Рабочие тетради для самостоятельной работы содержат множество разнообразных заданий с терминами, картой, портретами с возможностью фиксации полученных результатов каждой страницы с помощью принтера. Эти задания можно использовать при закреплении учебного материала по теме или разделу. Все задания можно классифицировать следующим образом:

а) группа заданий, которая предполагает сопоставления сведений из двух столбиков перемещением контента с помощью компьютерной мыши;

б) вторая группа – это самостоятельное определение с помощью изображения соответствующей исторической личности, города, события в специально выделенном поле для ввода текста;

в) третья разновидность заданий – это заполнение пустот в тексте словами (названиями, датами, фамилиями, понятиями), чтобы сохранилось с точки зрения смысла и логики первоначальное состояние авторского текста;

г) четвёртая группа – это небольшие кроссворды по темам с возможностью заполнения на странице и фиксации результата.

III. Небольшие по объёму контрольные образовательные тесты для самопроверки по темам с автоматической обработкой результатов. Каждый тест содержит 7 вопросов, правильные ответы на которые необходимо указывать щелчком мыши. Тесты содержат ключевые позиции по каждой теме и позволяют при ответах закрепить, уточнить, корректировать полученные знания, однако они не могут быть использованы как средство объективного контроля знаний по предмету.

На наш взгляд, выбор языка для изучения истории и обществознания находится сегодня в плоскости научно-практической реализации. Изучение истории и обществознания на коми языке на сегодняшний день находится в самом начале практической реализации. По нашему глубокому убеждению, подготовленные образовательные ресурсы создают условия для расширения сферы функционирования коми языка, что является благодатной почвой формирования личности.

### Литература

Остапов Н.В. Словарь терминов по истории на коми языке для общеобразовательных школ. Сыктывкар – Ижевск – Йошкар-Ола – Саранск – Бадачоньтомай: Издательство Мордовского университета, 2011. – 78 с.

Остапов Н.В. Словарь терминов по обществознанию на коми языке для общеобразовательных школ. Сыктывкар – Ижевск – Йошкар-Ола – Саранск – Бадачоньтомай: Издательство Мордовского университета, 2011. – 38 с.

Остапова Е.В. Мультимедийные литературно-образовательные ресурсы на коми языке // Родной язык в современных условиях двуязычия: сборник статей / Под общей ред. Е.В. Остаповой и С.Н. Терентьевой. — Сыктывкар: ООО «Издательство «Кола», 2014. – С. 267–274.



Попова Р.П., Лудыкова В.М., Остапова Е.В., Пунегова Г.В., Молчанова В.А. Кафедра коми и финно-угорской филологии Сыктывкарского государственного университета в условиях меняющейся образовательной парадигмы // Родной язык в современных условиях двуязычия: сборник статей / Под общей ред. Е.В. Остаповой и С.Н. Терентьевой. – Сыктывкар: ООО «Издательство «Кола», 2014. – С. 323-332.

## ПРОГРЕСС КОМПЬЮТЕРНОГО ПЕРЕВОДА И ПОЛЕМИКА О ПРЕПОДАВАНИИ АНГЛИЙСКОГО ЯЗЫКА В ЯПОНИИ

В то время как качество компьютерного перевода продолжает оставаться нестабильным в тех ситуациях, когда входящий текст содержит семантические или синтаксические неоднозначности, отныне выходящий текст вполне пригоден для многих практических целей. Люди, которым нужно читать или писать что-то на языке, которым они не владеют, всё чаще напрямую прибегают к компьютерному переводу, что чревато масштабными последствиями в различных областях и направлениях деятельности.

**Ключевые слова:** компьютерный, машинный, обучение, язык, перевод, культура.

В конце 2016 года качество переводов с/на японский и английский языки, выполняемых Google онлайн бесплатно, непредсказуемо внезапно улучшилось. До этого времени машинный перевод с/на японский и английский языки нередко производил непонятную тарабарщину. Низкое качество машинного перевода длилось на протяжении десятилетий. Многие сомневались в том, что машинный перевод когда-либо получит широкое применение, особенно для таких непохожих языков, как японский и английский. Недавние достижения в области машинного перевода [Chan 2018, 113] свидетельствуют о том, что пессимизм был неуместен. В то время как качество компьютерного перевода продолжает оставаться нестабильным в тех ситуациях, когда входящий текст содержит семантические или синтаксические неоднозначности, отныне выходящий текст вполне пригоден для многих практических целей.

Люди, которым нужно читать или писать что-то на языке, которым они не владеют, всё чаще напрямую прибегают к компьютерному переводу, что чревато масштабными последствиями в различных областях и направлениях деятельности, в частности речь будет идти об обучении английскому языку в Японии. Японский язык является национальным языком Японии и единственным языком, употребляемым в повседневной жизни японцев. Английский язык преподаётся всем японским детям в школах, причём каждый ребёнок посвящает сотни часов занятиям английским языком сперва до поступления в среднюю школу, затем дети продолжают изучать язык в средней школе, колледжах и университетах [Chapla 2012, 61]. Многие взрослые японцы также изучают английский язык самостоятельно либо на курсах для взрослых, либо стремятся общаться на английском языке с людьми из других стран.

Вероятно, отныне мотивация к изучению английского языка значительно уменьшится, ибо машинный перевод позволяет сегодня легко и эффективно осуществлять общение на многих языках, помимо английского. Разумеется, стремление к практическому общению не является единственной причиной массового изучения английского языка в Японии, некогда бывшей изолированной от всего мира. Японцы стремятся при помощи английского языка расширить свои представления о других странах и культурах [Kenny 2017, 116]. Процесс изучения английского языка считается в Японии полезным умственным упражнением, особенно в отношении грамматических штудий и экзерсисов по переводу с английского на японский и обратно [Ningsih, Emzir, Hidayat 2018, 66].

В качестве иллюстрации, обратимся к спорам о преподавании английского языка, возникавшим в Японии в 1920-х, но сегодня приобретших неожиданный резонанс в связи с новейшей версией учебной программы для младших школ. Поскольку усовершенствования в сфере машинного перевода [Yang, Lepage 2018, 119], по-видимому, повлекут за собой изменение отношения к необходимости изучения других языков, постольку этому документу суждено стать полезной отправной точкой для дебатов в кругах педагогов, политиков и среди широкой общественности на тему о том, что следует сделать в ближайшие годы, чтобы приспособиться к этой новой и быстро развивающейся технологии. Вопрос ставится по существу: зачем учить английский?

Дело в том, что английский язык преподаётся всем японским школьникам уже настолько долго, что представляется излишним интересоваться причинами преподавания английского языка. На самом деле, причины не вполне очевидны, особенно для самих детей, ибо в возрасте двенадцати лет большинство из них никакой практической потребности в английском языке не испытывают. Фактически, для подавляющего большинства единственным языком, используемым в общении, является японский язык. Возможны исключения, и окружающие детей взрослые в Японии общаются на другом языке, но это скорее китайский, корейский либо вьетнамский, но ни в коем случае не английский. Поскольку

непосредственная практическая польза не является убедительным оправданием для обучения всех японских детей английскому языку, приходится искать ответ на поставленный вопрос среди множества других причин, как очевидных, так и неявных.

Эти причины можно условно разделить на две категории. Первая категория по-японски называется кёё, что можно перевести как общие знания и развитие характера. Аргументация в поддержку пользы кёё как причины для обучения всех японских детей английскому языку акцентирует личную пользу, которую систематическое изучение английского языка якобы сулит детям как в краткосрочной, так и в долгосрочной перспективе. Вторая категория причин, называемая дзицуё, относится к использованию языка в практических целях. Поскольку детям школьного возраста английский язык по большей части не нужен сразу, то аргументация обращена к потенциалу детей на будущее с учётом владения английским языком. Следует отметить, что эти две позиции не являются взаимоисключающими, ибо сторонники кёё также склонны воспринимать дзицуё в качестве дополнительной веской причины, и наоборот.

Исторически сложившейся причиной изучения английского языка в Японии, по-видимому, была дзицуё в силу необходимости учиться у иностранцев. Первым языком, который сыграл когда-то подобную роль, был китайский язык около пятнадцати столетий назад. Китайский язык в своё время служил в Японии инструментом юридической кодификации и отправления религии. В период Эдо 1600–1867 гг., когда Япония была отрезана от большей части мира, и внешняя торговля ограничивалась купцами из Китая и Голландии, голландский язык изучался как способ получения знаний из Европы о науке, технике, медицине, географии и других предметах. Голландский язык был также средством, с помощью которого японские лидеры узнавали о последних событиях в мире.

Открытие Японии миру во второй половине XIX века привело к массовому изучению иностранных языков, особенно немецкого, французского и английского, поскольку страна спешила догнать западные державы в экономическом, военном и культурном отношении. Основная причина изучения иностранных языков была, таким образом, изначально утилитарной, или дзицуё. Однако к тому времени преподавание английского и других языков уже прочно вошло в образовательную систему, и появилось дополнительное, оправдание кёё, утверждавшее, что изучение иностранных языков, особенно английского, благотворно влияет на личностное и интеллектуальное развитие учащихся.

Примером такого аргумента, основанного на кёё, представляется часто цитируемая публикация полемикаческого характера в майском номере журнала «Гэндай» за 1927 год. Написанная литературоведом и профессором токийского Императорского университета Цукуру Фудзимурой статья называлась «Срочность ликвидации курсов английского языка». Ц. Фудзимура сетовал на то, что огромное количество времени посвящается изучению иностранных языков в средней школе. Поскольку большинству учащихся английский язык в будущем фактически не понадобится, и они не смогут применить его на практике, автор выступал за отмену обучения языкам и призывал вместо этого создать государственное ведомство, которое возьмёт на себя выполнение переводческих обязанностей.

Сегодня научно-технический прогресс ставит вопрос о том, стоит ли требовать от большинства детей, чтобы они тратили годы на изучение английского языка, которым большинство японцев в повседневной жизни практически не пользуются [Youdale 2018, 166]. Поэтому последствия усовершенствования машинного перевода глубоко отразятся на преподавании английского языка в Японии в первую очередь в аспекте практических целей, или дзицуё.

### Литература

- Chan S. The Human Factor in Machine Translation. – New York: Routledge, 2018. – 256 p.
- Chapla E. A. Statistical Machine Translation of Japanese. – Houston, TX: BiblioScholar, 2012. – 100 p.
- Kenny D. Human Issues in Translation Technology. – New York: Routledge, 2017. – 178 p.
- Ningsih E., Emzir E., Hidayat R. S. Dynamic equivalence assessment of Japanese comic (manga) “garasu no kamen” into Indonesian language translation “topeng kaca” (the glass mask) // International Journal of Linguistics, Literature and Culture, 2018. – Volume 4. Issue 4. – P. 53-68.
- Yang W., Lepage Y. Improving automatic Chinese–Japanese patent translation using bilingual term extraction // Transactions on Electrical and Electronic Engineering, 2018. – Volume 13. Issue 1. – P. 117-125.
- Youdale R. Using Computers in the Translations of Literary Style. – New York: Routledge, 2018. – 242 p.

## **ЦИФРОВИЗАЦИЯ КАК УСЛОВИЕ ФОРМИРОВАНИЯ НОВЫХ КОМПЕТЕНЦИЙ ПЕДАГОГОВ**

Статья посвящена цифровизации образования. Учителем ставится задача рассмотреть эффект от использования цифровых образовательных ресурсов (ЦОР) на уроках. Показано, что ЦОР играет важную роль в процессе усвоения новых знаний, развивает коммуникативные умения и навыки, привыкает работать в команде, реализовывать совместную деятельность, снятие нервной нагрузки. Благодаря цифровизации возрастает уровень использования наглядности на уроке. Главное достоинство цифровизации – открытость для диалога и коммуникации, возможность самореализации. Цифровизация позволяет «войти» педагогу в новую систему управления учебным процессом, поведения, культуры. Таким образом, речь идет о формировании у педагога качественно новых компетенций. Отечественная система повышения квалификации должна проектировать новые ресурсы для формирования этих компетенций.

**Ключевые слова:** обновленное содержание образования, компетентность, цифровизация, методические ресурсы.

*B.A.Torekeyev, R.S.Torekeyeva, Z.Shadkam  
Almaty, Kazakhstan*

## **DIGITALIZATION AS A CONDITION FOR NEW TEACHERS COMPETENCES FORMATION**

The article is devoted to the digitization of education. The teacher is tasked to consider the effect of the use of digital educational resources (DER) in the lessons. It is shown that DER plays an important role in the process of assimilation of new knowledge, develops communicative skills and skills, accustoms to work in a team, realize joint activities, and relieve nervous tension. Thanks to digitalization, the level of use of clarity in the lesson increases. The main advantage of digitalization is openness for dialogue and communication and the possibility of self-realization. Digitalization allows the teacher to log onto a new system of educational process, behavior, culture management. Thus, we are talking about the formation of qualitatively new teachers competencies. National Professional Development System should design new resources for formation these competencies.

**Keywords:** updated content of education, competence, digitalization, methodical resources.

“The teacher always unwittingly strives to choose the most convenient way of teaching for himself. The more convenient the teaching method for the teacher, the more uncomfortable it is for students. Only that way of teaching is true with which students are satisfied. We will seek and try on the ways of learning that students will be pleased with.” The model of education of the twenty-first century throughout the world is an educated society. Continuing education - should be the basis of each person’s life, a condition for his professional growth, mobility, development of internal potential, creativity over the course of life. At the center of the quality of modern education is the quality of its content, the quality of human education.

Over the past decades, the world has been rapidly moving to a new type of economy, where digital technologies are becoming the main tool for its formation. [[https://digitalkz.kz/upload/iblock/020/44010\\_mik\\_rk\\_pp\\_2017\\_proekt\\_pp\\_rk\\_10\\_04\\_2017\\_1\\_.pdf](https://digitalkz.kz/upload/iblock/020/44010_mik_rk_pp_2017_proekt_pp_rk_10_04_2017_1_.pdf)]. In modern conditions, information technology and the integration of digital technologies in all areas of business are the main factor in technological change and a condition for ensuring the competitiveness of the state. According to a study by Boston Consulting Group, placed in the article “Kazakhstan on the way to the digital economy” [Сычева], in terms of digitalization, Kazakhstan is ranked 50th out of 85 countries and is in the group with an emerging digital economy. To maintain the competitiveness of our economy, the development of the digital component through the joint efforts of the state and the education system is necessary. The first step towards digitalization of Kazakhstan was the comprehensive Digital Kazakhstan program, adopted on January 31, 2017. In his 2018 Address “New Development Opportunities in the Conditions of the Fourth Industrial Revolution,” N. Nazarbayev identified the challenges ahead for us in the development of digital technologies [<https://uchet.kz/news/poslanie-prezidenta-2018-goda-novye-vozmozhnosti-razvitiya-v-usloviyakhchetvertoy-promyshlennoy-rev/>]. Modern digital technology has stepped forward so quickly, that it became possible to print not just

sheets of documents on printers, and create complex three-dimensional projection models using 3D printers, which are used in various fields.

The student's personal interest is a decisive factor in the education process. One of the main tasks is to increase the pedagogical skill of a teacher by mastering modern teaching and educational technologies. Pedagogical technology – the design of the educational process, based on the use of a set of methods, techniques and forms of organization of training and educational activities that increase the effectiveness of training, the use of which has a clearly defined result.

With the mastery of any new technology, a new pedagogical thinking of the teacher begins: clarity, structure, clarity of the methodological language, the emergence of a sound norm in the methodology. Applying new pedagogical technologies in the lessons in the process of teaching English can achieve better results. To increase the effectiveness of the educational process when conducting English lessons, we use elements of the following educational technologies, taking into account the age characteristics of students:

The project method is aimed at developing active independent thinking of the students and teaching them not only to remember and reproduce knowledge, but to be able to put it into practice. It is important that in the work on the project, children learn to cooperate, and training in cooperation brings them mutual assistance, desire and ability to empathize, creative abilities and activity are formed. It should be remembered: in order to solve the problem that underlies the project, students must have certain intellectual, creative and communicative skills. These include the ability to work with text, analyze information, make generalizations, conclusions, the ability to work with a variety of reference material, the ability to conduct a discussion, listen to and hear an interlocutor, to defend one's point of view, and the ability to express thoughts concisely.

Thus, the competent use of the project method requires considerable preparation, which is carried out in an integrated training system, and it is not necessary that it precedes the students' work on the project.

Such work should be carried out continuously. At the first stage, a topic is selected and a project work plan is developed. The second stage provides the actual work on the project:

- putting forward hypotheses for solving the problem (“brainstorming”), discussing and substantiating each of the hypothesis;
- discussion of methods for testing accepted hypothesis in small groups (in each group one hypothesis), possible sources of information to test the hypothesis put forward; presentation of results;
- work in groups on the search for facts, arguments, confirming or refuting the hypothesis.

If there are errors in the final version of the project, we correct them with a pencil or write notes on a separate sheet of paper. Then the students themselves can decide how to correct mistakes in the final version of the work.

The third stage is the defense and discussion of projects. Since the methodical content of the modern lesson of a foreign language is communication, the teaching of a foreign language is organized in such a way that it is similar to the process of natural communication. Protecting projects is great for this. Students enjoy discussing projects presented in training sessions, both in pairs and in groups; They're not afraid to express their point of view in front of the class, to ask questions of interest. An important point is the debriefing. In order to maintain and stimulate the activity of students during the discussion and to be able to objectively evaluate the participation of each of them in work, the teacher can use various ways of encouraging. So, during the lesson, students can earn points for a successful question or answer, the use of active vocabulary, the use of communicative cliches. At the end of the lesson, points are summed up, and grades are given based on these results.

In addition, the practice of using the project method shows that – to learn together is not only easier and more interesting, but also much more effective. Helping each other, solving any problems together, sharing the joy of success or the bitterness of failure is as natural as laughing, singing, enjoying life. Therefore, it is safe to say that the student develops creative competence, as an indicator of communicative knowledge of a foreign language at a certain level. The project creates maximum conditions for the manifestation and formation of the student's creative abilities.

We are sure that today we need to quickly navigate the world of discoveries of new technologies, catch modern trends, open school scientific and STEM laboratories, robotics classrooms, nano-technology classrooms; design, model, design with students more.

This is the priority of Kazakhstan education. For the successful implementation of the technical area, it is necessary to provide the education system with qualified personnel. Today it is necessary that every teacher in any school discipline can prepare and conduct a lesson using digital technology. The use of digital educational resources in the educational process is one of the ways to increase learning motivation, aimed at developing the creative personality of not only the student, but also the teacher, helps to realize the main human needs – communication, education, self-realization.

However, in the practice of teaching, there are also disadvantages that were identified as a result of a survey of teachers of educational institutions:

- 62% of teachers consider that the use of digital educational resources in the educational process is not a priority for them;
- 50% of teachers say that they don't have enough time to prepare lessons using digital educational resources.

Questioning of students showed that in some subjects digital educational resources are used only 3-4 times a month. To eliminate the identified problems, a pedagogical council «Methods of developing informational competence of teachers» was held at the educational institutions. One of the decisions of the teacher's council was to intensify self-education on the use of digital educational resources through participation in distance learning courses. Teachers were invited to create a school bank of their video lessons and video lectures for their subsequent publication on the Internet.

Transformation is happening everywhere. Automation is changing entire industries. Artificial intelligence is on the rise. The work environment – how, where and when we work, changes a lot. The future of the work environment has already come, and the cause of change is digitalization.

- 87% of teachers agree that digitalization is a priority in their educational institutions;
- 66% of teachers are wondering what impact digitalization has on current and future work roles.

Technologies of the “digital doubles” family are the main driver of ongoing changes in enterprises, leading to significant changes in the structure of demand in the labor market. On the one hand, specialists who are able to “assemble” products and technological processes from ready-made digital models – components will be in demand. On the other hand, there is a demand for the creators of libraries of ready-made models – components. These areas will require teachers to develop fundamentally different competencies.

“Teacher of the future: a new role for teachers in the digitalization of education”, note that the national project «Education» is a key in creating a modern educational environment with a developed network infrastructure that allows the widespread use of social educational networks, innovative technological solutions that provide personalized training in a diverse interaction of participants in educational relationships using varied, adaptive digital content and digital educational technology, identified the priorities of the professional community:

- The active participation of the pedagogical community in updating the content of education and introducing new methods of training and education, digital educational technologies;
- Creation of digital professional communities with the involvement of teachers, the parent community, partners and business representatives to motivate them to learn about their involvement in the educational process aimed at self-determination and professional orientation;
- Creating conditions for the introduction of a national system of teacher growth, ensuring the participation of at least 50 percent of teachers of educational institutions and the formation of a new system of personal and professional development of teachers;
- Modernization of vocational education, including additional education, through the introduction of practice-oriented educational programs for continuous updating by teachers of professional knowledge, acquisition of new skills, including mastery of competence in the field of “digital pedagogy”;
- Supporting the initiative of the Academy's scientists to continue the scientific research “Kazakhstan teacher”, “Graduate of the Kazakhstan school” in the context of the digitalization of the school and conducting scientific discussions on urgent problems of professional and personal development of the teacher;
- Creating conditions for the development of a mentoring system, supporting initiatives and projects of young teachers, providing opportunities for personal professional and career growth.

One of the main ways to improve the quality of students' knowledge is the organization of the educational process. Transition to updated educational content high demands are made. But we cannot achieve their fulfillment if we treat the lesson as a fragment of life and turn it into a spontaneous process. The lesson begins in time, the organization of class space, the clear organization of all stages of the lesson, the interaction between the teacher and students, the teacher's reaction to certain student actions, the selection of educational material and methods of its presentation, the use of modern teaching methods and technologies, the use of visualization and digital educational resources are all this affects the final educational outcome of students. Modern teaching of a foreign language is a combination of facts of thinking and independence – qualities so valuable in developmental learning technologies. Today, the required minimum content foreign language education of school graduates. Gradually, an idea develops of what to do in the lessons of a foreign language, what to teach, how to teach and teach, how to learn and learn. The main methodical innovations are associated with the use of active or, as they are also called,

interactive teaching methods that allow you to interact or be in a mode of conversation, dialogue with something (for example, a computer) or someone (person). Interactive learning is a special form of organization of cognitive activity. Its essence lies in the organization of the educational process, in which almost all students are involved in the cognitive process, they have the ability to understand and reflect on what they know and think. As a result, conditions are created under which the student feels his success, his intellectual viability, which makes the learning process productive. Digital educational resources are able to stimulate cognitive interest in a foreign language, give a problematic, creative character to academic work, to a large extent contribute to the renewal of the content side of subjects, to individualize the learning process and to develop independent activity of schoolchildren. The joint activity of students in the process of cognition, development of educational material means that everyone makes their own individual contribution to this process, that there is an exchange of knowledge, ideas, ways of activity. Moreover, this happens in an atmosphere of goodwill and mutual support, which allows not only to obtain new knowledge, but also develops the cognitive activity itself, translates it into higher forms of cooperation. During the interactive training, students learn to think critically, solve complex problems based on an analysis of circumstances and relevant information, weigh alternative opinions, make informed decisions, participate in discussions, and communicate with other people. For this, individual, pair and group work, research projects, role-playing games, work with documents and various sources of information are organized in the lessons. Thus, interactive learning allows you: to develop communication skills, train you to work in a team, provide students with the necessary information, without which it is impossible to implement joint activities.

The systematic use of a digital educational resources in a lesson leads to a number of important consequences:

- sharply increases the level of use of visibility in the lesson;
- the productivity of teachers and students in the lesson increases;
- interdisciplinary communications are established with the basics of computer science and computer technology;
- it becomes possible to organize students' project activities to create curricula under the guidance of teachers of foreign language;
- a teacher who creates or uses information technology is forced to pay great attention to the logic of teaching material, which has a positive effect on increasing the level of students' knowledge;
- teacher relations with students far from history, especially those who are passionate about DER, are changing for the better;
- these students begin to see a "soulmate" in the teacher;
- changes, especially among middle-level students, the attitude to the DER as an expensive, exciting toy.

The students begin to perceive it as a universal tool for working in any area of human activity.

Using a digital educational resources in a lesson is a sphere of optimizing the teacher's work, it helps students increase their interest in subjects, expands the possibilities of using various visual aids, develops students' computer skills, makes it possible to diversify the form of the lesson, and increases the ability to organize students' project activities. When learning a new topic, we give a lesson with a multimedia presentation. This allows students to focus on the significant points of the information presented. The combination of oral lecture material with a slide show allows you to focus visual attention on particularly significant moments of the educational work. After this lesson, we invite students to complete a booklet on the topic. The use of digital educational resources makes it possible: to increase the motivation of learning; individual activity; focus on the personality of the student; the formation of information competence; freedom of creativity; interactive learning. The joint work of the teacher and student in the lesson makes this lesson interactive. When the student's personality comes first, his individual abilities and inclinations. We cannot ignore the fact that new educational technologies are rapidly entering our lives. And the faster we turn to them, evaluate their value and develop a methodology for their application, the more complete, informative, exciting the educational process will be, both for students and for us.

Thus, the effective implementation of the tasks of high-quality updating and digitalization of school education in the Kazakhstan is possible subject to the following mandatory conditions: 1) the diffusion of new digital technologies in education, including by providing access to mobile technologies; 2) creating a digital educational environment that can ensure the development and dissemination of new digital technologies that is safe for students, including the development of modern digital educational content; 3) ensuring the quality professional development of the teacher, motivation to master new digital skills, the use of digital educational content through the transformation of the system of continuous teacher education.

### References:

Государственная программа «Цифровой Казахстан» [Электронный ресурс]. URL: [https://digital.kz/upload/iblock/020/44010\\_mik\\_rk\\_pp\\_2017\\_proekt\\_pp\\_rk\\_10\\_04\\_2017\\_1\\_.pdf](https://digital.kz/upload/iblock/020/44010_mik_rk_pp_2017_proekt_pp_rk_10_04_2017_1_.pdf).

Казахстан на пути к цифровой экономике. Екатерина Сычева, директор BCG [Электронный ресурс]. URL: [http://image-src.bcg.com/Images/BCG-Review-May-2016\\_tcm27157057.pdf](http://image-src.bcg.com/Images/BCG-Review-May-2016_tcm27157057.pdf).

Послание Президента 2018 года «Новые возможности развития в условиях четвертой промышленной революции». [Электронный ресурс]. URL: <https://uchet.kz/news/poslanie-prezidenta-2018-goda-novye-vozmozhnosti-razvitiya-v-usloviyakhchetvertoy-promyshlennoy-rev/>



## ЛИНГВОКУЛЬТУРОЛОГИЧЕСКИЙ ПОДХОД В ПРЕПОДАВАНИИ

Изучение параметрической пары «близко-далеко» в грамматике относится к наречиям и прилагательным. Наречия подразделяются на группы. Данная пара относится к наречию места, которое характеризует действие с позиции места или направления этого действия и отвечает на вопросы: где? куда? откуда? В статье данная параметрическая оппозиция исследуется с акцентом на лингвокультурологический подход в изучении и преподавании разных языков.

**Ключевые слова:** лингвокультурология, преподавание, разноструктурные языки.

Комплексная научная дисциплина, изучающая взаимосвязь и взаимовлияние культуры и языка в его функционировании и отражающая этот процесс как целостную структуру в единстве их языкового и внеязыкового (культурного) содержания называется лингвокультурологией [Дерюгина 2007, 52–54].

Наряду с влиянием на сознание индивидуума коллективного опыта, который включает в себя традиции, образ жизни, обычаи, индивидуальный опыт, также оказывает влияние специфическое восприятие окружающего мира, свойственное данному человеку.

Соответственно язык, мышление и культура тесно взаимосвязаны и «практически составляют единое целое». Другими словами, язык отображает реальность, «от настоящего мира к мышлению и от мышления к языку» [Терминасова 2000, 39]. Для того, чтобы правильно представить культуру изучаемого языка, необходимо ознакомиться с культурным фоном для изучения его глубинной сути.

В качестве эмпирического материала в своем исследовании мы использовали фразеологизмы, поговорки, пословицы. Основное внимание было сосредоточено на концептуальных единицах, которые представляют собой совокупность определенных лексических единиц. Раскрытие и реконструкция концептов позволяет донести до учащихся учебных заведений знания как о родной лингвокультуре, так и страны изучаемого языка, ведь именно в сравнении культур происходит глубокое понимание значимости своего языка и культуры. Предметом нашего исследования являются единицы языка, получившие символическое, эталонное, образно-метафорическое значение в культуре и которые обобщают результаты собственно человеческого сознания

Традиционно пословицы и поговорки изучались в фольклористике как жанровые тексты, сейчас они активно изучаются в лингвистике. С прагматической точки зрения цель пословиц размыта: одна и та же пословица может быть упреком, утешением, нравоучением, советом, угрозой. Но не все пословицы могут изучаться в лингвокультурологии. Есть определенный пласт пословиц, который имеет общее для некоторых происхождение, а значит, у них нет конкретной истории народа или этноса.

Фразеологический фонд языка – ценнейший источник сведений о культуре и менталитете народа, т.к. в них отражены мифы, обычаи, обряды, ритуалы, морали, поведение, привычки и т.д. определенного народа.

Так, необходимо составить диалог с использованием наречий/прилагательных «близко-далеко», участие в котором могут принять англичане, русские и другие представители культур. В вышеупомянутом диалоге пользуемся выражениями, как *“How do I get to ...? Is it far? It is near the office. It takes 5 minutes”*.

Для реализации частично-поискового и исследовательского методов обучения, учащимся можно дать задания на нахождение фразеологизмов, пословиц и поговорок подбор русских/башкирских аналогов и анализ их с культурологической точки зрения. Так как фразеологизм связан со стереотипом (образом, лежащим в основе индивидуального и общественного сознания), то именно он является средством выражения этого стереотипа, который связан с определенным представлением или неким образом, выраженном в данном фразеологизме.

Первой подгруппе необходимо найти пословицы (proverbs) как на английском языке, так и на башкирском языке с использованием слов «близко-далеко», второй – поговорки (sayings), а третьей – устойчивые словосочетания. Эти исследовательско-поисковые задания могут быть заданы в виде домашней работы, так как требуют времени для поиска и анализа заданий. Для сравнения те же работы могут быть даны, опираясь на башкирскую литературу.

Преподавателем могут быть поставлены различные вопросы, например, в пословицах, поговорках, песнях, загадках, скороговорках нужно уделить особое внимание объектам и темам носителей

английской культуры, а также тому, какие темы и объекты важны для носителей башкирской культуры и почему.

Важнейшим условием формирования и совершенствования умений студентов по осуществлению межкультурной коммуникации является формирование вторичного когнитивного сознания студентов. Следствием этого явилось выделение в качестве основного компонента коммуникативной компетенции студентов языковых вузов лингвокультурологической компетенции.

Лингвокультурологический подход или его элементы могут быть положены в основу методик обучения другим аспектам языка в рамках языкового вузовского образования.

#### **Литература**

Дерюгина И. В. Педагогика искусства как новое направление гуманитарного знания. 2007. – 233 с.

Зубкова Н. Ю. Пространственные категории «далеко-близко» во французской лингвокультуре // Теория и практика общественного развития. № 7 (2011).

Маслова В. А. Лингвокультурология. – М.: Изд. Центр «Академия», 2001. – 208 с.

Степанов Ю. С. Константы. Словарь русской культуры, 2000. – 824 с.

Терминасова С. Г. Язык и межкультурная коммуникация. – М.: Слово, 2000. – 624 с.

## СПОСОБЫ СОХРАНЕНИЯ КУЛЬТУРНОГО НАСЛЕДИЯ

В данной статье рассматривается возрождение и охрана историко-культурного наследия Республики Татарстан. Особо отмечается работа республиканского Фонда «Возрождение», учрежденного в 2010 году на основании указа Президента РТ М. Ш. Шаймиева. На данный момент фондом достигнуты две главные цели – включение в список Всемирного наследия ЮНЕСКО историко-архитектурного комплекса г. Болгар (в 2014 году) и Успенского Собора и монастыря острова-град Свияжск (в 2017 году).

**Ключевые слова:** историко-культурное наследие, духовное наследие, элементы культуры.

Историко-культурное наследие является ценнейшим достоянием человечества. Оно включает в себя сферу не только материальной культуры, но и духовной. Испокон веков наследие наших предков передается при помощи памятников культуры. Памятники культуры и истории имеют несколько значений. Первое заключается в том, что памятник является объектом, который представляет собой богатое наследие и историю определенного этноса или страны. А, второе, довольно широко употребляемое его значение, заключается в том, что памятник является произведением, которое было создано для того, чтобы увековечить память о каких-либо событиях или исторических деятелях.

Не стоит забывать, что к памятникам истории и культуры не всегда относились с должным вниманием. С давних пор они подвержены разным опасностям. Во-первых, памятники культуры гибнут и исчезают по причине вмешательства человека. Человек может способствовать этому намеренно или же случайно. Яркими примерами здесь могут послужить пожары, войны, неосторожные реставрационные работы, недостаточное внимание к признакам историчности, и актуальные в последние десятилетия – террористические акты. Во-вторых, памятники культуры могут погибнуть из-за неуправляемых сил природы, которые неподвластны человеку. Землетрясения, наводнения, ураганы и многие другие явления природы могут раз и навсегда уничтожить богатое наследие наших предков. Культурное наследие немыслимо без духовной ее составляющей. Оно очень тесно взаимосвязано с материальной стороной, т. к. богатства духовного наследия обычно имеют материальное воплощение. К сфере духовного наследия мы обычно относим следующие доминирующие элементы культуры – это язык, религия, мифология, искусство, наука, философия, праздники, танцы, песни и многое другое. Таким образом, мы выявили следующее определение этому явлению: культурное наследие представляет собой широкий спектр ментального и телесного освоения опыта предков, включающий в себя памятники духовной материальной культуры, способствующие просвещению и обогащению молодого поколения.

Культурное наследие имеет ряд функций – экзистенциальную, репродуктивную, аксиологическую, этическую, креативную. Именно эти функции обеспечивают стабильное развитие и передачу историко-культурного наследия через определенные механизмы. Механизмы передачи культурного наследия способствуют сохранению непрерывной межпоколенной связи. Соблюдение и актуализация этих механизмов в совокупности – нормативно-правового, ритуально-сакрального, морально-аксиологического, воспитательно-обучающего и художественно-творческого, дает положительный результат для всего государства в виде сознательного и ответственного гражданского общества.

Сфера охраны и защиты культурного наследия имеет довольно долгую историю. Проблема сохранения ценных для государства памятников культуры волновала человечество практически на протяжении всей истории. Особая заслуга в этой деятельности непременно принадлежит историкам. Но именно на законодательном и государственном уровнях прогресс был выявлен только в 20 веке. Решающую роль в этом сыграл пакт, составленный великим русским гуманистом Н. К. Рерихом – «Пакт о сохранении и защите культурных ценностей», который был подписан и официально признан 21 страной и провозглашал охрану культурных ценностей выше военной необходимости. Данная идея и сам пакт получили широкую популярность во всем мире, развернулась масштабная деятельность по сохранению наследия культуры, названная впоследствии рериховским движением. С момента появления пакта идеи и ценности Н.К. Рериха прошли длительный и сложный период своего существования. Было выявлено 5 этапов его функционирования, которые имели характер либо подъема интереса к его ценностям и деятельности, либо спада. Пакт Н. К. Рериха является предтечей такой крупной организации как ЮНЕСКО. Роль ЮНЕСКО в сфере охраны и возрождения историко-культурного наследия действительно колоссальна. Объединив своей программой, целями и актуальными

задачами почти все страны мира, она сделала огромный шаг по установлению стабильного и цивилизационного межкультурного контакта между странами. Деятельность этой организации можно разделить на два важнейших периода. Первый период связан с принятием в 1972 году Конвенции об охране природного и культурного наследия, после которого был предложен список объектов, которые имеют особую историческую и культурную ценность и находятся под охраной мирового сообщества. Второй период связан с принятием в 2003 году Конвенции об охране нематериального культурного наследия. Это стало огромным прорывом для всего человечества, т. к. документ закрепил за собой важность истории и культуры каждого народа, а также, важность сохранения культурного многообразия в мире. ЮНЕСКО является на данный момент самой значимой организацией в этой области, она способствует не только охране, защите, возрождению и приумножению памятников культурного наследия, но и обеспечивает стабильность цивилизационного развития, способствует усилению межкультурного диалога, и объединяет все страны-участницы общими программами и целями по актуализации современных проблем человечества.

Процессы, предшествующие охране и возрождению культурного наследия, можно проследить в таком феномене, как поворот к историческим корням. Наша современность немислима без знания истории, именно поэтому человеку свойственно к ней обращаться, а иногда и жить по ее правилам. Данный феномен обуславливает тот факт, что история циклична, и многие явления, которые были популярны в прошлом, могут вновь стать актуальными уже в современное время. Существуют следующие факторы, которые способствуют к интересу возвращения к истокам:

1) Стремительный процесс глобализации и культурной универсализации. Впоследствии чего, постепенно стала размываться культурная самобытность и разрушаться этническая идентичность. Именно эти процессы способствовали деятельности по сохранению и возрождению памятников культуры, т. к. появились опасения, что уникальная культура отдельного народа растворится в мировой культуре. Возвращение к истокам дает человеку чувство уверенности и позволяет почувствовать твердую почву под ногами;

2) Назревшие актуальные вопросы и нестабильная геополитическая обстановка. Этот фактор наиболее важен при исследовании феномена поворота к корням, т.к. обращение к истории и попытка возродить события и порядок из прошлого чаще всего происходит при каких-либо актуальных конфликтах или при неуверенности людей в завтрашнем дне;

3) Неудовлетворенность стремительным ростом технического прогресса и бегство от современных реалий жизни. Прогресс в сфере технологии пронизывает всю нашу жизнь, мы с каждым годом убеждаемся, что без нее мы не справимся и все вокруг находится под ее контролем. Многих людей это может не устраивать, именно поэтому они прибегают к данному феномену – к возврату к историческим корням.

Эту деятельность с успехом выполняет республиканский Фонд «Возрождение», учрежденный в 2010 году на основании указа Президента РТ М. Ш. Шаймиева. Фонд имеет главной целью сохранение, охрану, реставрацию и возрождение историко-культурного наследия, находящегося на территории РТ. Эта деятельность должна содействовать совместному усилию государства и общества для решения проблем, связанных с сохранением и приумножением наследия предков. Также не менее важная ее задача была популяризация важности культурного наследия для молодого поколения. Это осуществляется через всевозможные проекты, гранты, мероприятия и т. д. На данный момент результаты деятельности фонда действительно впечатляют, например, были достигнуты две ее главные цели – включение в список Всемирного наследия ЮНЕСКО историко-архитектурного комплекса г. Болгар (в 2014 году) и Успенского Собора и монастыря острова-град Свияжск (в 2017 году). Болгар и Свияжск стали изюминкой Татарстана, как толерантная и дружественная республика, способствующая мирному сосуществованию и процветанию православной и исламской культуры. Такая культурная политика способствовала развитию туризма, что благоприятно влияет на экономику Республики. Также организация занимается реставрированием и возрождением памятников историко-культурного наследия находящихся на всей территории РТ.

## **ИСПОЛЬЗОВАНИЕ КОРПУСНЫХ ДАННЫХ В ИЗУЧЕНИИ ЛЕКСИКИ**

В последнее время у многих учёных-филологов интерес вызывает вопрос о перспективах развития лингвистического познания. Не исключено, что одной из причин его возникновения стало содержание лекции известного российского лингвиста В.А. Плуногяна, в которой он достаточно категорично высказывает мысль о том, что «современная лингвистика должна быть лингвистикой корпусов. Корпус позволяет понять, каков язык на самом деле, а не каким мы хотим, чтоб он был. Корпус произвёл колоссальную революцию. Корпус вернул лингвистике её настоящий, полноправный объект» [Плуногян 2009]. Позиция В.А. Плуногяна нашла большой отклик, особенно среди методистов и преподавателей иностранных языков. Появилось достаточное количество научных статей, сообщающих о важности и нужности использования корпусов в обучении [Евстигнеев 2011, 119-125].

Проблема отбора содержания, поиска средств и разработки технологий самостоятельного изучения иностранного языка особенно широко обсуждается лингводидактическим сообществом последние двадцать лет. Проблема чаще всего рассматривается в контексте перехода на продуктивную образовательную деятельность, которая будет способствовать развитию у учащихся «реальной самостоятельности, креативности и автономии в языковой и образовательной среде» [Коряковцева 2002, 5].

Т. Джонс, стоявший у истоков корпусного обучения, отмечал задачу обучаемого как явление «совершать открытия» (discovery learning – эвристический метод) в иностранном языке, а задачу преподавателя – обеспечить контекст, в котором стратегии обучаемого могут развиваться.

Корпус развивает навыки студента к самообучению, автономности, также помогает преподавателю выстроить демократичную обучающую среду. Данный переход имеет следствие снижения традиционной роли грамматики как неких правил, уровень абстракции которых зачастую делает их неэффективными и препятствует пониманию [Bernadini 2004].

Работа с корпусом текстов может осуществляться как во время аудиторных занятий в небольших группах и индивидуально, так и в виде самостоятельной работы. Главным преимуществом работы с корпусами является организация творческой работы студентов (как самостоятельной, так и под руководством преподавателя), формирование навыков работы с информационными ресурсами и навыков исследовательской работы. «Сам процесс обучения в рамках традиционного подхода подразумевает, что учитель (и учебник) выступает в роли источника готового знания: студенту даются сформулированные правила, готовые к использованию. Вместе с тем, многими лингвистами-практиками отмечается, что знание, в приобретении которого студент принимает активное участие, гораздо лучше откладывается и быстрее усваивается. Использование корпуса текстов в обучении (подход, который получил в англоязычной традиции название Data-Driven Learning) основывается на постулате о том, что студент должен в процессе своего обучения выступать в роли исследователя, делать самостоятельные выводы на основе непосредственно эмпирических данных, представленных в корпусе, не прибегая к интерпретациям грамматистов и не будучи ограничен неполнотой лексикографических ресурсов» [Язык. Культура. Образование 2008, 148-163].

В условиях, когда теоретические правила не работают (например в случае с артиклями), корпус обеспечивает исходными данными обучающего и подталкивает его к самостоятельным выводам. Несмотря на то, что эти умозаключения могут и не быть полными и абсолютно верными, важно, что они будут соответствовать языковым навыкам и интуиции обучающегося, т.е. соответствовать промежуточной языковой системе обучаемого [Вайнрайх 1979].

Упражнения, разработанные на базе того или иного корпуса текста, могут быть следующими:

1. найти с какими словами сочетается конкретная лексическая единица;
2. проанализировать контекст двух (или более) слов, принадлежащих одному синонимическому ряду, и определить оттенки значения каждого;
3. при предъявлении новой лексической единицы можно предложить студентам карточки с примерами для анализа контекста с целью определения значения предъявляемой единицы;
4. подготовить раздаточный материал и предложить студентам восстановить в примерах пропущенные слова, которые связаны с ключевым словом;
5. то же самое задание, только студенты должны определить пропущенное ключевое слово;
6. на примерах определить особенности функционирования того или иного грамматического явления.

Существует значительная разница между простой помощью преподавателя и выстраиванием опор. Авторы, работающие над идеей обучения языку на профессиональном контенте (CLIL, Content and Language Integrated Learning), считают, что выстраивание опор в этой области особенно необходимо. Вместо предоставления готового ответа преподаватель может направить обучаемого к поиску ответа (например, по корпусу) и, тем самым, создать опоры, благодаря которым в следующий раз ученик решит подобную задачу самостоятельно [Dale L 2012].

Таким образом, корпус может стать эффективным помощником как обучаемому при изучении иностранного языка, так и преподавателю, благодаря аутентичной и всегда актуальной корпусной базе и различными функциональными разметками в корпусе.

В области лексики, например, заданием может быть выявление значения незнакомого слова, изучение полисемии, определение различий в использовании близких синонимов. С.М. Богатова [Богатова 2007, 156-164] описывает возможности использования корпусов на продвинутом этапе обучения, где «значительное внимание при введении новых или закреплении ранее изученных единиц уделяется их лексической сочетаемости (collocation)». Подборка контекстных строк корпуса является более полезным инструментом для данных целей, нежели словарь, потому что словарные примеры (предписательные), а примеры из реальной жизни (описательные).

Интересным приложением для изучения языка является использование программ-конкордансов параллельных текстов (Parallel Corpora). Такие разработки используются для сопоставительного перевода. Е.П. Соснина отмечает, что «в практическом смысле перевод должен ориентироваться на возможности постредактирования, сравнения и оценку различных стратегий и интерпретаций в рамках контекста» [Соснина]. Параллельные корпуса предоставляют образцы профессионального перевода и позволяют значительно сократить временные затраты.

Не следует также недооценивать значение учебных корпусов (Learner Corpus). Под учебным корпусом понимается электронный корпус текстов группы лиц, изучающих иностранный язык. Такого рода корпуса, например, могут быть использованы для лингвистического анализа на предмет выявления разного рода ошибок при изучении иностранного языка. Это помогает преподавателю скорректировать учебный план и выработать методические приемы путем анализа частотности тех или иных типов языковых ошибок и их характерного окружения. Наиболее известным является международный англоязычный корпус ICLE (International Corpus of Learner English). Он содержит эссе студентов продвинутого языкового уровня. В основном этот корпус используется для статистического анализа вокабуляра учеников, дискурсивного анализа, а также для сопоставительных исследований.

Таким образом, используя корпуса, преподаватель получает множество примеров как грамматической, так и лексической формы слова, левого и правого контекста лексической единицы. Студент, в свою очередь, получает естественные примеры демонстрации тех или иных явлений, может самостоятельно проводить лингвистические исследования, заниматься НИР.

Возникнув как инструмент лингвистического анализа, корпус сегодня активно осмысливается в плоскости обучения языку, что позволяет говорить о становлении нового направления в гуманитарном образовании – корпусного преподавания.

Использовать корпус также возможно для осуществления контроля уровня усвоенного материала. Здесь можно применить корпус для составления лексико-грамматических заданий, тестов, создания проектов.

Говоря о собственном опыте, автор, например, организовал на своем рабочем компьютере некий аналог простого корпуса с помощью одного из программных продуктов с открытым исходным кодом «DocFetcher» (<http://docfetcher.sourceforge.net/>), позволяющий совершать контекстный поиск по содержимому файлов на домашнем компьютере. В описании приложения сказано: «Это как Google, но только для локальных файлов». Индекс позволяет программе быстро определить, в каких файлах содержится определённый набор слов.

Другим часто используемым инструментом в личной практике можем назвать мультязыковую платформу Glosbe (<https://glosbe.com>). Хотя изначально он и позиционировался как простой словарь, но постепенно перерос в полноценный корпусный инструмент.

Здесь не лишним будет отметить, что, отыскав нужное слово в словаре, очень полезно не просто прочитать первое попавшееся значение, а дочитать словарную статью до конца. Это нужно для того, чтобы познакомиться с прямыми и переносными значениями слова, с самыми распространенными фразеологизмами, в состав которых оно входит, а также с примерами типичного употребления. Так запоминание будет направлено не на конкретный перевод, а на представление слова в целом, или, иначе говоря, будет запоминаться не перевод слова, а его значение и «образ»; пользователь будет более свободен от конкретных предложенных словарём переводов.

Теория и практика перевода является ещё одним направлением, в развитии которого корпусная лингвистика играет существенную роль. Эмпирические данные компьютерного массива текстов служат основой достоверности выводов в исследовании трудностей перевода.

Подводя итоги проведенным исследованиям, можно отметить, что использование методов корпусной лингвистики во время обучения иностранному языку является довольно эффективным, так как позволяет обучающимся самостоятельно извлекать необходимую информацию, проводить исследования, делать выводы и выбирать для себя необходимый темп работы. По результатам анкетирования [Меркулова, Галкина 2019, 125-132], одним из ведущих критериев профессионализма является индивидуально-деятельностный подход, который очень хорошо обеспечивается благодаря использованию корпусной лингвистики в процессе обучения. Кроме того, обучающиеся приветствуют инновационные методы в обучении, а также инициативу со стороны преподавателя в области разработки ориентированных на отдельно взятого учащегося заданий, создания новых методик в обучении.

Таким образом, роль корпусного подхода в такой области, как компьютеризированное обучение иностранным языкам центральна.

### Литература

Богатова С.М. О некоторых возможностях применения конкорданса при изучении лексики // Информационно-коммуникационные технологии в обучении иностранным языкам: материалы II Междунар. семинара, Омск, 5-6 июня 2007 г. – Омск: Омский гос. ун-т им. Ф.М. Достоевского, 2007.

Вайнрайх У. Языковые контакты. – Киев: Вища школа, 1979. – 264 с.

Евстигнеев М.Н. Структура ИКТ-компетентности учителя иностранного языка // Язык и культура. 2011. № 1.

Коряковцева Н.Ф. Современная методика организация самостоятельной работы изучающих иностранный язык: пособие для учителей. – М.: Аркти, 2002.

Меркулова Т.К., Галкина А.А. Оценка профессионализма педагога по результатам использования корпусной лингвистики на занятиях по иностранному языку. – Наука и школа. 2019. № 2.

Плунгян В.А. Лингвистика в XXI веке: проблемы, перспективы, точки роста // Слово.ру: балтийский акцент. 2018. Т. 9. № 1. – С. 7–12.

Плунгян В.А. Зачем мы делаем Национальный корпус русского языка? / В.А. Плунгян // Отечественные записки. – 2005. – № 2.

Плунгян В.А. Почему современная лингвистика должна быть лингвистикой корпусов: [Электронный ресурс]. URL: <http://www.polit.ru/lectures/2009/10/23> (дата обращения: 10.02.2019).

Соснина Е.П. Корпусная лингвистика и корпусный подход в обучении иностранному языку. [Электронный ресурс]. URL: [http://ling.ulstu.ru/linguistics/resources/literature/artides/corpus\\_linguistics\\_language\\_teaching/](http://ling.ulstu.ru/linguistics/resources/literature/artides/corpus_linguistics_language_teaching/) (дата обращения: 14.01.2019).

Соснина Е.П. Параллельные корпуса в обучении языку и переводу, 2015. – URL: <http://spr.fld.mrsu.ru/2015/09/e-p-sosnina-parallelnye-korpusy-v-obuch-enii-yazyku-i-perevodu/> (дата обращения: 10.12.2019).

Язык. Культура. Образование: монография / Под общ. ред. Е.В. Харченко. – Челябинск: Изд-во ЮУрГУ, 2008.

Bernadini S. Corpora in the classroom: An overview and some reflections on future developments // Sinclair J. McN. (ed.) How to use corpora in language teaching. – Amsterdam [u.a.]: Benjamins, 2004. – 308 p.

Dale L. CLIL Activities: A resource for subject and language teachers / L. Dale, R. Tanner. – S.I.: Cambridge Univ. Press, 2012. – 284 p.

*Ш. Г. Абдирова  
г. Ақтобе, Казахстан*

### ОРГАНИЗАЦИЯ ТЕКСТОВ ПРИЗЫВОВ

В статье рассматриваются особенности организации текстов призывов на казахском языке. Анализируются собранные автором тексты сегодняшних вывесок городов Ақтобе, Туркестан, Шымкент, Хромтау, Кандагаш Республики Казахстан. Автором определены приёмы организации текстов, выявлена специфика организации текстов призывов на национальном языке.

**Ключевые слова:** призыв, тексты призывы, организация текста.

*Ш. Г. Абдирова  
Ақтобе қ., Қазақстан*

### ҮНДЕУ МӘТІНІНІҢ ҰЙЫМДАСУЫ

Мақалада үндеу мәтіндерінің ұйымдасуы қарастырылып, үндеу мәтіндерінің ұйымдасу амалдары мысалдар арқылы талданады. Автор үндеу мәтіндерінің ұйымдасуына тән ерекшеліктерді анықтайды.

**Түйін сөздер:** үндеу, үндеу мәтіндері, мәтін байласымы, мәтін тұтасымы.

Үндеу мәтіндерінің уәжі мен мақсаты – адресатты сендіру, иландыру. Үндеу мәтіндерінің өзіндік құрылымы бар. Біз мақаламызда осы мақсатқа байланысты үндеу мәтіндерінің құрылымдық және мағыналық ұйымдасу заңдылықтары мен байласым, тұтасым мәселелерін талдаймыз.

Мәтіннің байласымы бір сөйлемнің екіншісімен, екіншісінің үшіншісімен байланысуын білдіреді, ал тұтасым мәтін бөліктерінде, мысалы күрделі фразалық тұтастықта, абзацта, бөлімде, тарауда т.б. қамтылған мазмұнның бірімен екіншісінің байланысып жатуы.

Мәтін байласымы (когезия) күрделі бірліктерді құрайтын тілдік құралдардың белгілі әдіс-тәсілдермен ұйымдасуынан көрінеді. «Мәтіннің байласымы мен тұтасымын айыра білген жөн. Байласым тұтасымның шарты болып табылады, ал тұтасым байласым арқылы толықтай айқындала алмайды. Енді бір жағынан, байласым сипатына ие мәтін үнемі тұтасым қасиетін иеленбейді» [Садирова 2019,127]. Егер мағыналық-тұлғалық жағынан сәйкеспейтін сөйлемдерді бір мәтін бойына жисақ, тұтасымы жоқ дүние құралады. Мәтіннің негізін, оның тұтасымын осы бірліктер арасындағы тұлғалық-мағыналық байланыс түзеді. Соның негізінде байласымның екі тәсілі көрінеді: 1. Эксплицитті – мәтін құрамындағы сөйлемдердің өзара формалды, тілдік бірліктердің қолданысы арқылы байланысуы; 2. ИмPLICITТІ – мәтін құрамындағы сөйлемдердің өзара мағына бірлігі арқылы байланысуы. Мәтін байласымы – тек екі сөйлем аралығын ғана қамтымайды, бұл күрделі фразалық тұтасымдарды, абзацтарды, тарауларды да біріктіретін категория. Ендеше мәтіндегі сөйлемдер арасында екі түрлі байласым түрі анықталады. Оның бірі іргелес байласым, яғни өзара іргелес сөйлемдер арасында орнаса, екіншісі – алшақ байласым, араларында бір не бірнеше сөйлем орналасқан синтаксистік бірліктердің арасында жүзеге асады [Садирова 2019,128].

Сайлау алды дискурсы мәтіннің ұйымдасуы мәселесіне қатысты зерттеуінде ф.ғ. докторы, профессор К.Қ. Садирова мәтіндегі байласымды жүзеге асырушы лексика-грамматикалық амал-тәсілдерге есімдік, тәуелдік жалғау, үстеу, шылау, қыстырма, синоним, антоним сөздерді етістіктің бірыңғай шақ тұлғасын, синтаксистік параллельдер мен түрлі қайталауларды жатқызады [Садирова 2019,126].

Мәтін лингвистикасында мәтіннің құрылымдық бірліктерінің байласымын жүзеге асыруға қатысатын құралдар субституттар мен дейксистер, скреп сөздер немесе шегендеуіштер деп аталады.

Субститут (орынбасар) – бір сөздің де сөйлемнің де не сөйлемдер тізбегінің де орнына қолданылып, зат, мекен, мезгіл т.б. атаулардың мағынасында қолданылады. Бұл орынсыз қайталауды болдырмау үшін қолданылады.

Дейксис – (грекше – deixis – сілтеме) тілдік құралдармен берілетін сілтеме, орнына жұмсалып отырған тілдік бірліктің мағынасына немесе қызметіне нұсқайды. Оларға есімдіктер, үстеулер, шылаулар мекендік, мезгілдік, ұғымды білдіретін сөздер жатады. Дейксистер мәтіннің алдыңғы бөлігінде қамтылған хабарға нұсқайды.



Біз қазіргі кезде қала көшелерінде, ғимарат, мекемелердің маңдайшасында ілінген үндеу мақсатында жазылған мәтіндерді үндеу мәтіндері деп қарастырдық. Талдау нысанына Ақтөбе, Түркістан, Шымкент, Хромтау, Қандыағаш қаларының көшелерінде ілулі тұрған үндеу мәтіндері алынды. Ел Президенті Нұрсұлтан Назарбаевтың сөзін талдасак: *«Бүгінгі қазақ халқы сонау есте жоқ ескі замандарда-ақ тұлпарларының тұяғымен дүниені дүр сілкіндірген көне сақтардың, ежелгі ғұндардың, байырғы түріктердің ұрпағы, үлкен үйдің қара шаңырағын атажұртта сақтап қалған халық. Бұл ескіден қалған көне жәдігерлер, сақ заманының куәгері және Түркі қағанатының тарихын танытқан Орхон-Енисей мен Талас жазбалары дәлелдеген шындық»* (Түркістан қаласы).

Келтіріліп отырған мәтін 2 сөйлемнен құралған. Бұл мәтін үзінде мәтіннің байласымы тізбектеле орнығып, мәтін логикалық құрылымы анық жеткізілген. Ел Президентінің айтпағы – қазақ халқының атажұртты, қара шаңырақты сақтап қалған халықекендігі. Автордың ойын жеткізу **амалы** – *Бүгінгі қазақ халқы сонау есте жоқ ескі замандарда-ақ тұлпарларының тұяғымен дүниені дүр сілкіндірген көне сақтардың, ежелгі ғұндардың, байырғы түріктердің ұрпағы*, яғни қазақ халқының тарихына шолу болса, **дәлелдеу** – *Бұл ескіден қалған көне жәдігерлер, сақ заманының куәгері және Түркі қағанатының тарихын танытқан Орхон-Енисей мен Талас жазбалары дәлелдеген шындық*. Бұл үндеу мәтіні үзіндегі автор бірінші сөйлемді, тарихи шегіністі, информаттың – ақпарат алушының – патриоттық сезімін ояту, ұлттың тарихынан ақпарат беру мақсатында қолданып, екінші бөлігінде өз ойын тарихи жәдігерлермен, яғни нақты дерекпен дәлелдеген. Мәтін байласымының сәтті орнығуына сөйлем ішіндегі тіркестер де әсерін тигізеді. Бірінші сөйлемде қазақ халқының ұлттық танымына сәйкес *есте жоқ ескі замандардан-ақ, дүниені дүр сілкіндірген тұрақты* тіркестері мен тұлпарлардың тұяғы, көне сақтар, ежелгі ғұндар, үлкен үй, қара шаңырақ секілді бірдей тіркестер қолданылған. Бұл сөйлемде қазақ халқының танымында концептілік деңгейде танылған үлкен үй, қара шаңырақ тіркесімдері мен атажұрт сөзі байласымның жатық құрылуына септігін тигізіп алдыңғы тіркестерді өзіне үйіріп, бірінші сөйлемнің мәйегі – мәтін тұтасымын – ұстап тұр. Үндеу мәтінінің баяндауыштары да есімдерден бірінғай жасалған. Қ.Жұбанов атындағы Ақтөбе өңірлік мемлекеттік университетінің Жастар сарайының маңдайшасында ілінген Қ.Қ.Жұбановтың үндеу мәтініне талдау жасасак.

*«Бүгін жас мамандарды Қазақстан жұртшылығына тапсыра отырып, олар алған білімдерін іске асыра алар, көптің күткеніндей жеміс берер, ұлт мәдениетін құру жұмысында бұлар үмітті ақтап, шын қайраткер болар деп сенемін. Қ.Жұбанов»* Мәтіндегі үндеу амалына талдау жасасак: Ғалымның үндеу мәтіні аралас құрмалас сөйлемнен құралған. Үндеу мәтінінің баяндауыштары кезектесе етістіктің көсемше, есімше категорияларынан жасалған. Нақтырақ көрсетсек, *отырып, берер, ақтап, болар*. Есімше тұлғалы баяндауыш пен көсемше тұлғалы баяндауыш өзара кезектесіп мәтін байласымына әсер беріп тұр. Ғалымның идеологиялық мақсаты сенемін бірінші жақтағы баяндауышы мен мәтіннің бойындағы білімін іске асыру, жеміс беру, үмітті ақтау, жас маман тұрақты тіркестері және шын қайраткер тіркестері арқылы беріледі. Үндеу мәтіндерінде қаратпа сөз мәтін байласымын құраушы амалдың бірі. Бұл амал ғалымның өз стилдік ерекшелігіне сәйкес – *Бүгін жас мамандарды Қазақстан жұртшылығына тапсыра отырып – деген сөйлемге кіріктірілген. Мұндағы Қазақстан жұртшылығына тапсыруы – Қазақстандықтар, Қазақ жұрты деген қаратпа сөздің орнына жұмсалып тұр*. Ғалымның үндеу мәтінінің бойына қазақстандық жастарға міндеттелген жауапкершілік пен күтілетін нәтиже де қатар тізбектеліп, ғалымның қазақстандық жас мамандардың білім алуы, білімдерін іске жарату, жеміс беру, яғни ізгілендіру, дамыту, әрі қарай ұлт мәдениетін құру, қайраткер статусына жетуі. Бұл үндеу мәтінінде жас маманға сенім арта отырып, ғалым жастардың кәсіби өсуінің негізгі идеологиялық моделін құрып берген. *«Адамды адам еткен – кітап, адамзат еткен – кітапхана» Ә. Кекілбаев*. (Үндеу мәтіні Қ.Жұбанов атындағы Ақтөбе мемлекеттік кітапханасының холынан алынды). Бұл үндеу мәтінінің байласымы қайталанатын бірліктер арқылы жүзеге асқан. Адам үш түрлі тұлғада қолданылып тұр. Олар: адамды – табыс септігінде, адам – түбір күйінде, адамзат – туынды формада. еткен – есімше формасының 2 рет қайталануы, түбір тұлғадағы – кітап сөзі мен, туынды формадағы – кітапхана сөзі. Сөйлем негізінен 3 түбір сөздің тұлғалық өзгеруінен құралған. Бірақ мағыналық байласымы жұп-жұмыр. Бұл сөйлемде сызықша сөйлемнің игерілуіне ықпалын тигізіп тұр. Ол ой екпінін анықтап, сөйлемнің мазмұнының танылуына мүмкіндік береді.

Мәтін байласымын жүзеге асыруға қызмет ететін тәсілдің бірі – шылаулар. Шылаулар – лексикалық мағынаға ие болмағанымен, мәтінде ерекше қызмет атқарады. Соның ішінде жалғаулықты шылаулар өзара байланысқан сөйлеудің бірліктері арасында орнайтын түрлі логика-семантикалық қатынастарды эксплицитті түрде білдіретін тілдік құрал әрі олардың арасындағы байласымды жүзеге асырудың бірден-бір тәсілі болып саналады. Жалғаулықтар жекелеген сөйлемдердің арасындағы

мағыналық байланысты күшейте түседі. Ол мағыналық байланыстың түрі шылаудың білдіретін грамматикалық мағынасымен (қарсылықты, талғаулықты, себептік т.б.) сипатталады. Сондай-ақ мәтіндегі бірліктердің арасындағы байласымды жүзеге асыруда қыстырма құрылымдардың атқаратын қызметі зор. Қыстырма құрылымдар арқылы, біріншіден, сөйлемдер арасындағы байласым нығая түссе, екіншіден, мәтіндегі ойға автордың қатынасын білдіріп, мәтіннің модалділігін айқындайды. Бұл ретте автордың осы элементтер арқылы оқиғаға т.б. жағдайға әртүрлі қатынасын білдіруге мүмкіндік туады [Садирова 2019, 136]. *«Біздің жастарымыз ғылым мен білімге ұмтылып, елдің болашағын ойлайтын болса, келешегіміз кемел, тәуелсіздігіміз баянды болады» Н.Ә.Назарбаев.* Бұл үндеу мәтінде мен шылауы ғылым мен білім сөзін ғана емес, сөйлемнің байласымына әсерін тигізіп тұр. Егер ғылым және білімді десек немесе үтір арқылы айрықшаласақ, мағыналық байласымы жатық болмас еді. Ол ғылым мен білімді ұштастырып, сөйлемдегі жастардың рухани құндылығын анықтап тұр. Үндеу мәтінде байласым амалдарына талдау жасасақ: Үндеу мәтіні 3 құрамнан тұратын сабақтас құрмалас сөйлем. Сөйлемнің бойына біздің жастар, ғылым мен білімге ұмтылып, елдің болашағын ойлау, кемел келешек, баянды тәуелсіздік тіркестері енген. *«Тілі жоғалған жұрттың өзі де жоғалады» А.Байтұрсынов, «Өмір сүру үшін өзгере білу керек» Н.Ә.Назарбаев.* Бұл үндеу мәтіндеріндегі шылаулар мәтіндегі әрекеттің нәтижесінің анықталуына мүмкіндік тудырып тұр. Қорытындыласақ, қазіргі қала көшелерінде, мекеме маңдайшаларындайлулі тұрған насихаттау мазмұнындағы мәтіндерді үндеу мәтіндері деп қарастыруға толық негіз бар деп есептейміз. Себебі олардың негізгі мақсаты – мәтінді игерушіні сендіру, иландыру, үндеу мазмұнында жазылған. Үндеу мәтіндерінің өзіндік құрылымы бар, яғни олардың мәтін байласымы жатық, тиянақты болып келеді.

#### **Әдебиеттер:**

Садирова К.К. Дискурсология мәселелері. – Ақтөбе: Колибри, 2019. – 200 б.

## **ОСОБЕННОСТИ КАЗАХОЯЗЫЧНОГО ИДЕОЛОГИЧЕСКОГО ДИСКУРСА**

В статье рассматриваются существующие определения и разновидности дискурса в казахстанском языкознании, дано понимание авторов дискурса как текста, задействованного в процессе коммуникации. На основе анализа тестов выступления Н.А. Назарбаева выявлены особенности регионального идеологического дискурса Республики Казахстан, определены отличия текстов данного дискурса от текстов политического дискурса по своему структурному содержанию.

**Ключевые слова:** казахстанское языкознание, исследование дискурса, идеологический дискурс.

Рассмотрение языка в познавательном отношении в казахстанском языкознании стало давать социально значимые, отвечающие насущным потребностям современности результаты. Увеличились ряды ученых, занимающихся этим научным направлением, зародившимся в Европе, вышли в свет исследовательские работы казахстанских ученых.

Результатом исследования языка в познавательном аспекте является также и дискурс. Проблемы теории дискурса исследуются казахстанскими учеными с разных сторон: даются различные определения дискурса, выделяются его специфические признаки.

В статье д.ф.н., профессора К.К. Садировой «Дискурс теориясы мәселелерінің казахстандық лингвистикада зерттелуі» («Исследование проблем теории дискурса в казахстанской лингвистике») дается подробное рассмотрение вклада казахстанских ученых во всестороннее исследование дискурса. По словам исследователя, Ж. Кеншинбаева называет дискурсом используемый текст, профессор Уали же считает дискурс коммуникативным «событием», включающим понятия «фрейм», «автор», «словесный акт», «адресат», «словесная ситуация» (употребление слова в определенной ситуации). Г.Г. Буркитбаева под дискурсом понимает совокупность текстов участников интеракции и коммуникативной ситуации. По исследованиям А.А. Кибрик, дискурс – речевое взаимодействие и его результат, где процесс речевого взаимодействия – коммуникация, результат речевого взаимодействия – текст, т.е. дискурс – совокупность коммуникации и текста. В исследованиях д.ф.н., профессора К.К. Садировой дискурс определяется как переход в сознание и обозначение в языковой единице явления отдельного отрезка реальной действительности через речевое действие [Садирова 2011, 42].

В нашем понимании, текст остается просто текстом, зафиксированным на бумаге или в аудиозаписи, пока не будет освоен человеком. Коммуникация – процесс речевого действия. А дискурс – текст, задействованный в процессе коммуникации.

В языкознании исследуются и специальные разновидности дискурса (диалогический дискурс – А.Е. Саденовой, газетный дискурс – Ш.А. Сабитовой, кинодискурс – Ш.А. Нурмышевой, политический дискурс – К. Кенжеккановой и др.).

Политический дискурс, относящийся к институциональному виду дискурса, – основа политической коммуникации. Политический дискурс – система действий, целенаправленно используемая социальными институтами, группами, отдельными личностями в обществе. Он реализуется в виде текстов либо реального речевого действия. К политическому дискурсу относятся политические статьи, тексты, послания, программы, доклады и др. У идеологического дискурса, до настоящего времени рассматриваемого в рамках политического дискурса, есть свои, специфические особенности. Идеология – это идеи, преследующие определенные цели. Термин «идеология» впервые введен в научное обращение французским ученым, просветителем Антуаном де Тарси. В работе «Элементы идеологии» под идеологией он определяет учение о мнениях, обеспечивающих стабильность в обществе, играющих важную роль в воспитании деятельных личностей. Идеология – основное средство управления государством. Идеология находит выражение в политическом пространстве. В науке данное явление называется идеологическим дискурсом. По высказываниям зарубежных исследователей, идеологическому дискурсу присущи определенные специфические особенности на региональном или государственном уровнях. В странах с устойчивыми демократическими традициями основными объектами идеологического дискурса являются текущая политика и общественные проблемы.

С целью определения структурной модели идеологического дискурса рассмотрим статью в соответствии с содержанием и задачами частей: вводная часть – этап психологической подготовки, несущая основную нагрузку идеологического дискурса в формировании позитивного настроения адресата,

влияющая на формирование личной точки зрения адресата, т.е. подготовительный этап для введения адресата в дискурс. Основная часть – этап освоения идеологического дискурса, заключительная часть – этап завершения. Проанализируем текст статьи Н.А.Назарбаева с целью выявления отличий идеологического дискурса от политического дискурса [Назарбаев 2017]. Вводная часть содержит обзорную информацию о прошлом, нацеленную на привлечение, убеждение адресата. Можно сказать, что эта часть – этап психологической подготовки адресата. Здесь дается аналитический обзор претворенных в жизнь деяний и их итогов. Эта часть заканчивается словами, нацеленными на мотивацию адресата: «... мы готовы решать новые задачи».

Вторая часть посвящена программным задачам в конкретных сферах деятельности. Здесь рассматриваются поставленные цели. В политическом дискурсе наблюдается частое употребление в идеологических целях лексики с политическим значением.

Основное отличие идеологического дискурса от политического заключается в использовании слов, в которых закодированы многовековая история и будущее государства. К примеру, в тексте часто используются слова и выражения *нация, духовное возрождение, независимость, национальное сознание, культурное наследие, национальные традиции*. К идеологическому дискурсу можно отнести и статью, основанную на идеологических интересах.

В статье Н.А.Назарбаева «Взгляд в будущее: модернизация сознания» говорится: «В 2013 году мы приняли программу «Халық тарих толқынында», позволившую нам системно собрать и изучить документы из ведущих мировых архивов, посвященные истории нашей страны. А сегодня мы должны приступить к более масштабной и фундаментальной работе» [Назарбаев 2017]. Текст передан от лица Лидера нации, воспринимается как посыл в будущее в интересах государства в целом и каждого отдельно взятого казахстанца. Это происходит благодаря тому, что Лидер нации связывается в сознании народа с независимостью, будущим нации.

Тексты политического дискурса нацелены на осуществление в политике интересов отдельных лиц либо определенных групп и отличаются от текстов идеологического дискурса по своему структурному содержанию. В них часто употребляются выражения *выполним свои обязательства, партия оправдает ваше доверие* и т.п. Идеологический же дискурс нацелен на соблюдение всеобщих интересов и закладывает основы будущего всей нации. К примеру, «Само понятие *духовной модернизации* предполагает изменения в *национальном сознании*. Здесь есть два момента.

Во-первых, это изменение в рамках *национального сознания*.

Во-вторых, это сохранение внутреннего ядра *национального «Я»* при изменении некоторых его черт.

Наши *национальные традиции и обычаи*, язык и музыка, литература и свадебные обряды, – одним словом, *национальный дух*, должны вечно оставаться с нами.» [Назарбаев 2017]. В этом фрагменте слово *национальный* повторяется несколько раз в различных формах и сочетаниях (*национальном сознании, ядра национального, национальные традиции и обычаи, национальный дух*). Слово *национальный*, повторяясь и видоизменяясь, достигает наивысшей эффективности в формировании национального самосознания участника дискурса. Кодифицированные слова не только оказывают влияние на национальное самосознание, но и способствуют формированию бережного, уважительного отношения к духовным ценностям, восприятию новых путей духовной модернизации. Это является показателем многогранности идеологического дискурса.

Анализ частотности стержневого для идеологического дискурса слова *национальный* показал, что оно встречается в тексте статьи в различных сочетаниях: *национальный код (2), национальная традиция (1), национальное сознание (2), национальная память (1), национальные традиции и обычаи (2), национальный прагматизм (1), национальное развитие (4), национальное самосознание (1), национальная самобытность (1), национальный дух (1), национальная культура (2), национальная гордость (2), общенациональный (1), язык межнационального общения (1), национальный информационный проект (1)*. Несмотря на то, что идеологический дискурс осуществляется в политическом пространстве, наличие в нем политической терминологии не является обязательным условием. В данном тексте, написанном в жанре статьи, изредка использованы политические термины (*политическая реформа, коммунизм, либерализм, индустриализация и др.*)

Таким образом, на наш взгляд, идеологическим дискурсом можно считать текст в процессе коммуникации, имеющий в своем составе кодифицированные слова и словосочетания, фразеологические единицы, определяющие национальные ценностные понятия, отвечающие национальным интересам, способствующие формированию патриотических чувств участника дискурса.

Наряду с этим мы попытались показать на основе наблюдений отличительные особенности идеологического дискурса. Во-первых, любой текст, содержащий политические термины (т.е. поли-

тический дискурс) не может быть идеологическим дискурсом без соответствующей привязки идеологическим интересам. Во-вторых, политический дискурс включает тексты, пропагандирующие политическую информацию, программы прохождения во властные структуры, и политические термины в них подчинены этим задачам. Идеологический дискурс выражает общенациональные интересы в текстах государственных программ, посланий, статей, нацелен на формирование патриотических чувств участника дискурса. В-третьих, в идеологическом дискурсе национальные коды выражают национальные ценностные ориентиры, служат формированию национального сознания.

#### **Литература**

Назарбаев Н.А. «Взгляд в будущее: модернизация сознания» (12.04.2017) [Электронный ресурс]. URL:// [www.akorda.kz](http://www.akorda.kz) (дата обращения 15.11.2019).

Садирова К.Қ. «Дискурс теориясы мәселелерінің қазақстандық лингвистикада зерттелуі» // «Дискурс теориясының өзекті мәселелері» халықаралық ғылыми конференция материалдарының жинағы. – Ақтөбе, 2011. – 42-43 б.б.

## ОСОБЕННОСТИ ПОСЛОВИЦ И ПОГОВОРК ГЕНДЕРНОГО ЗНАЧЕНИЯ (на материале казахского и башкирского языков)

В данной статье рассмотрены гендерные пословицы и поговорки, характеризующие женщину в казахском и башкирском языках в сравнительном аспекте. Большинство пословиц и поговорок в языке выражают образ женщины. На основе большого количества гендерных пословиц и поговорок двух сравниваемых языков выявлены особенности, характеризующие образ женщины, ее статус в семье.

**Ключевые слова:** гендер, гендерная лингвистика, когнитивная лингвистика, концепт, женщина, женский образ, пословицы и поговорки, казахский язык, башкирский язык.

Понятие концепта открывает новую парадигму лингвистических исследований. Особо ценный опыт в понимании и применении термина концепт был накоплен в отечественной лингвокультурологии рубежа XX–XXI веков. Философские основания этому были заложены еще в 70-80-е годы Р. Павленисом.

В научный обиход отечественного языкознания термин *концепт* был введен в 20-е годы XX века. Однако лишь в 80-е годы в связи с переводами работ англоязычных авторов понятие «концепт» как термин адаптировалось на отечественной почве [Бабушкин 1996, 14], а с начала 90-х начинает активно использоваться в лингвокогнитологии.

В когнитивной лингвистике термин «концепт» является базовым. Этот термин заимствован когнитивной лингвистикой из логики. Проблема определения концепта в целом не является новой для лингвистики. Еще в средние века ставился вопрос о существовании языковых универсалий. Проблема концептуализации затрагивалась в трудах Л. Вейсгербера, В. Фон Гумбольдта, А. А. Потебни.

Концепция В. Фон Гумбольдта – это прежде всего концепция человека-языкотворца, концепция своеобразия «духа народа», поэтому закономерно, что к данной концепции восходят такие направления современного языкознания, как исследование языковых картин мира (ЯКМ), гендерная лингвистика, когнитивная лингвистика, лингвокультурология, тесно связанные между собой. В концепции В. Фон Гумбольдта противопоставление языка и речи, системы и ее реализации представляют антиномии, диалектически отражающие сущность языка как «великого духовного богатства человеческого рода» [Гумбольдт 1984]. «Несомненно, что семантическая система каждого языка отражает специфическую концептуализацию объективной действительности и субъективного мира его носителей. Это требует своего объяснения, и в поисках его, как правило, обращаются к экстралингвистическим факторам, которые трактуются как культурные явления данного народа» [Шелякин 2005]. Паремологию, с нашей точки зрения, целесообразно рассматривать достаточно широко, как «внешний ореол» фразеологии, включающий не только пословицы и поговорки, но и афоризмы и «крылатые слова» [Федоренко 1990, 3].

Для каждой вещи, окружающей человека в повседневной жизни, в языке имеются какие-то соответствия. Поэтому все, что люди видят и о чем думают, они могут с большим или меньшим успехом передать словами, особенно через пословиц и поговорок.

Пословицы и поговорки – это особые словесные знаки. Пословицы любого языка (в данном случае казахского и башкирского) представляют собой продукты языкового народного сознания как материализация опыта поколений. Эта форма обращается в сфере внутринационального общения, то есть, как отмечает В. А. Аврорин, «в рамках употребления языка, равно как и той или иной языковой формы его существования, совпадающих с границами соответствующего этнического единства (народа)» [Аврорин 1975, 275].

Пословицы являются отдельными познавательными актами, выраженными в изолированных языковых структурах. Поэтому, как отмечает Г. В. Колшанский, их интерпретация может быть адекватной лишь при учете не только самого языкового выражения, но и знания о действительных ситуациях, приобретенных башкиром, казахом или турком на основе как своего индивидуального опыта, так и общественного опыта, закрепленного и переданного в языковых формах (текстах) [Колшанский 1990, 108].

Концептов, выражающих гендерные типы, не очень много. Наиболее часто встречающиеся среди них это концепт «мужчина» и концепт «женщина». Причем первый появляется гораздо чаще, чем второй.

Благодаря структурным анализам языка можно изучать место генев в мировых языках и культуре, особенности поведения женщин и мужчин в различных отношениях, изменения гендерных норм во времени и пространстве, языковые, стилистические признаки, отличающие женский и мужской пол, роль генев в освоении языка и др. По мнению Б. Хасанулы, «мужчина и женщина встречаются в жизни общества, в употреблении в нем языка. Вот эта ситуация послужит основой для постановки вопроса о возможном развитии гендерной лингвистики в социальной лингвистике» [Хасанулы 2003, 47].

А. В. Кирилина в своих трудах выделила следующие понятия «гендер» и «пол»: «Гендер (социальный или социокультурный пол) не является языковой категорией, но его содержание может быть раскрыто путем анализа структур языка, что объясняет востребованность лингвистической компетенции для изучения культурной репрезентации пола» [Кирилина 2005, 133].

В фразеологизмах содержится большое количество гендерных образов, символизирующих женщину и мужчину. Как бы ни было, народная мудрость для гендеров, употребляющих слова мужчин и женщин, служит основой в пословицах и поговорках. Большинство пословиц и поговорок в нашем языке выражали гендерный образ женщины. В нашем языке достаточно гендерных пословиц и поговорок, характеризующих социальную роль женщины в ее жизни, семье [Шоқым 2007, 194].

Однако это обстоятельство ни сколько не уменьшает важность концепта «женщина». Можно сказать, что он является одним из ключевых концептов картины мира. Женщина-мать, вот тот первый человек, с которым будет взаимодействовать новорожденный, и можно сказать, что именно ее глазами маленькое существо смотрит на окружающий его мир, т.е. воспринимает его через призму материнского восприятия. Иными словами, мать закладывает в сознание своего ребенка основы картины мира, которые характерны для его народа и нации. И именно она способствует тому, чтобы у ребенка сформировалось правильное отношение к женщине. Потому как именно с нее начинается жизнь на земле.

В тюркских языках женщина как категория, абстрактное понятие обозначается словом *катын-кыз* (башк.), *эйел-кыз* (каз.), первая составляющая этого понятия *катын* (башк.), *эйел* (каз.) – служит для обозначения жены, супруги, а также женщины. Причем по проведенным исследованиям для тюркских народов более существенным становится подчеркивание значение жены, а не женщины. *Кыз* (башк.), *кыз* (каз.) же имеет значение «девушка, девушка, дочь», то есть обозначает существо женского пола до брачных отношений.

Роль женщины в мире сложна и многообразна. И это отображено в пословицах и поговорках, которые являются результатом многовекового опыта человеческой цивилизации. В этой статье мы рассмотрим образ женщины казахского и башкирского народов.

**Образ девушки.** Наибольшее количество языковых единиц с положительной коннотацией посвящено психологическому аспекту характера. В свою очередь, самые многочисленные единицы этого раздела отражают черты женского характера: на казахском языке: *Қыздың қырық жаны бар; Ақылды қыз білімге жүгірер; Қыздың қабағы күлімдесе – құт келеді; Қызда қырық көліктік бақыт бар; Қыз ат байлатар; Қыз мінезді келсін, Ұл өнерлі келсін; Қызға қырық үйден тыйым; Отырған қыз орнын табады; Жақсы қыз – жағадағы құндыз, жақсы жігіт – аспандағы жұлдыз* [Малайсарин 2011, 65]; на башкирском языке: *Кыз байлығы – күркәм холок, ир байлығы – тәуәккәл; Кыз баланың бирнәһен етеһенән йыйһаң – етмәй торор, алтыһынан йыйһаң – артып калыр; Кыз егеткә кушыла – ебәк булып ишелә; Кыз кемгә бармаҫ ине, егет кемде алмаҫ ине калым малы булмаһа; Кыз күңелендә – балалы бишек, ир күңелендә – эйәрле ат* [Нәзершина 2006, 86].

**Образ и статус женщины в семье.** В социологическом плане подчеркивается значимость женщины в семье: в казахском: *Әйел – үйдің көркі, еркек – түздің көркі; Әйел – үйдің ажары; Әйел – үйдің ұйытқысы; Әйел – үйдің қазығы; Ақылды әйел күйеуінің ойнап айтқанын, ойнап талдайды; Жақсы әйел жарының жақсылығын асырар, жамандағын жасырар* [Малайсарин 2011, 81]; в башкирском: *Катының – гүмерлек юлдашың; Катын гәйрәтле булһа, ир гибрәтле була; Катын теләһә, ишәктән ир яһар, теләһә, ирзән ишәк яһар; Катынды йөзөк кашы иткән дә, муса ташы иткән дә – ир.* [Нәзершина 2006, 105];

**Образ мудрой, всезнающей женщины, хранительницы очага.** Образ мудрой, всезнающей женщины, хранительницы очага воссоздается благодаря отведенной ей роли в семье быть рачительной, пекущейся о благосостоянии семьи, мужа, детей. Женщина в руках, у которой спорится любая работа, от которой веет неповторимой и непередаваемой аурой уюта, тепла, аурой красивого, ухоженного дома. Такую же окраску имеет и следующий пример: *Анасыз үй – панасыз; Анаға баланың алалығы жоқ; Ана алдында құрмет, ата алдында қызмет; Ананың басқан жері пейіш; Анасыз өмір сөнген көмір* [Малайсарин 2011, 76]; в башкирском: *Әсә янында бала етем булмай; Эшләгәнең атаң*

*менән әсәң өсөн булһа, өйрәнгәнең үзең өсөн; Әсә хақы – Тәңре хақы; Әсәй тыузырыусы, Атай туйзырыусы* [Нәзершина 2006, 102].

В казахском и башкирском языках образ женщины находит отражение во многих пословицах и поговорках, отражающих такие качества, как красота, мудрость, уважение и заботливость к мужу, детям и другим членам семьи, хозяйственное и гостеприимство. В гендерно-лингвистических исследованиях большое значение имеет сбор таких пословиц и поговорок, поскольку они отражают языковые особенности в менталитете определенной нации. Основной целью исследования казахских и башкирских пословиц является определение мирового образа, национального сознания и культуры каждого народа, принципов, обычаев и традиций, сложившихся в процессе общественного развития, места мужчин и женщин в них, а также сбор всесторонних сведений об их социальном статусе.

### Литература

- Аврорин В. А. Проблемы изучения функциональной стороны языка. – Л: Наука, 1975. – 275 с.
- Бабушкин А. П. Концептуальные типы значений // Контрастивные исследования лексики и фразеологии русского языка. – Воронеж, 1996. – С.10–14.
- Гумбольдт В. Фон. Избранные труды по языкознанию. М.: Прогресс, 1984. – 397 с.
- Кирилина А. В. Язык и гендер. – М.: Языки славянской культуры, 2005. – 624 с.
- Колшанский Г. В. Объективная картина мира в познании и языке. – М: Наука, 1990. – 108 с.
- Малайсарин Ж. Қазақ мақал-мәтелдері. – Алматы, 2011.
- Нәзершина Ф. Башкорт халык ижады. – Өфө, 2006.
- Самситова Л. Х. Культурные концепты в башкирской языковой картине мира. – Уфа: БГПУ, 2014. – С. 34.
- Федоренко Н. Т. Афористика / Н. Т. Федоренко, Л. И. Сокольская. – М.: Наука, 1990.
- Хасанулы Б. Проблемы развития гендерных исследований в казахском языкознании и вопросы женского языка // Язык и культура: функционирование и взаимодействие. Материалы Международной конференции. – Шымкент, 2003. – С. 47–58.
- Шелякин М. А. Язык и человек: к проблеме мотивированности языковой системы. – М.: Флинта: Наука, 2005. – 279 с.
- Шоқым Г. Т. Қазақ тіл білімінің гендерлік тұғыры. – Алматы, 2007. – 312 б.



## НЕКОТОРЫЕ ОСОБЕННОСТИ ЛИНГВИСТИЧЕСКОЙ СИТУАЦИИ В РЕСПУБЛИКЕ КАЛМЫКИЯ

Республику Калмыкию можно назвать одним из наименее известных субъектов Российской Федерации. Однако Калмыкия и её титульная нация заслуживают внимания ученых, по крайней мере, с целью сосредоточить взгляд на весьма своеобразные культурные и религиозные особенности калмыцкого народа, а также на его языковое наследие и другое. С целью исследовать языковую ситуацию в этом регионе России, настоящая статья представляет собой попытку представить некоторые особенности языкового законодательства Республики Калмыкия и некоторые меры, предпринимаемые различными органами, а также осветить основные проблемы.

**Ключевые слова:** язык, лингвистическая ситуация, языковая политика, Калмыкия, Россия, калмыцкий язык, русский язык.

L. V. Darzhinova  
Hong Kong, China

## SOME FEATURES OF THE LINGUISTIC SITUATION IN THE REPUBLIC OF KALMYKIA

The Republic of Kalmykia may be described as one of the least known constituent regions of the Russian Federation. However, Kalmykia and its major ethnicity deserves attention of scholars, at least, for the purpose of focusing on highly original cultural and religious traits of the Kalmyk people, followed by their linguistic heritage and so on. In order to explore the linguistic situation in this region of Russia, the present paper is an attempt to provide some features of the Republic of Kalmykia's language legislation and some measures taken by different bodies, also to highlight its major issues.

**Keywords:** language, linguistic situation, language policy, Kalmykia, Russia, Kalmyk, Russian

There are more than 200 languages spoken in Russia, over 130 languages are in danger of extinction and 20 of them have already been declared extinct, according to UNESCO (Moseley, 2010). There are 30 languages, which are considered official languages in various regions of Russia, along with Russian: Altay, Bashkir, Buryat, Chechen, Chuvash, Ingush, Khanty, Mansi, Mari, Moksha, Ossetic, Tatar, Tuvin, Udmurt, Yakut, etc. [Constitution of the Russian Federation, 1993]. Among them is the Kalmyk language, which is spoken in the Republic of Kalmykia.

The Republic of Kalmykia is a polyethnic region which lies at the south-eastern corner of the European Russia. The representatives of more than 90 nationalities and ethnic groups live in Kalmykia. The Kalmyks constitute more than 50% of the population, 37% is presented by the Russians, the Dargins, the Chechens, the Kazakhs, the Ukrainians, the Belarusians, the Germans, the Estonians, the Koreans, and others [Esenova, 2015].

Historically, the Kalmyks are Western Mongolians, whose language is related to Mongolian, and their main religion is Tibetan Buddhism, which distinguishes them as one of the three Buddhist peoples of the Russian Federation, along with the somewhat more numerous Buriat and Tuvan peoples, and as the only Buddhist people in Europe [Goryaev & Omakayeva, 1999]. Consequently, the Kalmyk language is a member of the Kalmyk-Oirat subgroup of Mongolic languages.

At present, the Russian and Kalmyk languages are the state languages of the Republic of Kalmykia. The teaching of the Kalmyk language alongside Russian has been made compulsory for all schoolchildren. However, in practice, they differ in the sphere of usage and the level of functional development. It became a result of interaction of different reasons. Different scholarly works point out an inrush of Russian-speaking people, Russian-teaching schools, poor motivation for learning the Kalmyk language, generation gap, many others [Baranova, 2009; Bitkeeva, 2006].

Among these reasons is also the following historical explanation: in 1648, a Kalmyk Buddhist monk called Zaya Pandita Namka Djamts'o created the Kalmyk alphabet, or *Todo Bichig* (literally *Clear Script*), by adapting the vertically written Mongolian script. This alphabet was used until 1924, when it was replaced by the Cyrillic alphabet. Cyrillic was then replaced by the Latin alphabet but was later reintroduced and has been in use to this day. Despite the attention devoted to the choice of a writing system, Kalmyk has only a

limited written language and literary tradition [Tishkov, 1997]. By the way, *The Clear Script* is still used by Kalmyks in China [Avery, 2003].

The most powerful blow to the Kalmyk language though was the historical occasion that is the genocide Kalmyks (1943–1956) suffered at the hands of Stalin’s dictatorship. The entire Kalmyk ethnic group was rounded up and sent to the most remote and inhospitable parts of Siberia and Kazakhstan. Many died from the hardships, while the rest, allowed to return only 15 years later, struggled to reclaim their land and language [Gouchinova, 2013]. Overall, the fact that the languages of indigenous peoples in the Russia’s regions are decreed as the state languages makes no difference. Also, more than 400 years of living with the Russian population in the region and the competition with the Russian language left its deep mark on the Kalmyk language, which demonstrates many explicit linguistic transformations at different levels, such as phonological, morphological, syntactic and lexical-semantic, about which can be written a separate article.

In accordance with the 1999 Language Act of the Republic of Kalmykia, the texts of legal documents of the Republic of Kalmykia are published in the state languages. In addition, newspapers and magazines, TV and radio programs are maintained in both Kalmyk and Russian. The Language Act regulates all spheres of language functioning apart from the use of the languages of the peoples of Kalmykia in interpersonal communication [Kornousova, 2001].

In terms of status planning, some progress has been made as regards the official status of the Kalmyk language, due to the Law of the Republic of Kalmykia “On the State Languages of the Republic of Kalmykia and other languages in the Republic of Kalmykia” [Law of the Republic of Kalmykia 2014]. However, given the explanations aforementioned and some other factors, the status reached is of symbolic rather than practical value, as Russian prevails nearly all major areas of communication [Bogrdanova, 2015].

The Kalmyk language has a standardized grammar, phonetic, lexis and stylistic norms. According to Boschaeva, the compilers of the explanatory dictionary of the Kalmyk language should meet the requirement of language purism [Boschaeva, 2000]. To address the need to purify the language, new terminology was introduced through the local newspaper in Kalmyk entitled as “Khalmig Unin” (literally *The Kalmyk Truth*), several concise Kalmyk-Russian dictionaries in different spheres were published. In the mass media the language is used only in the newspaper “*Khalmig Unin*” and the magazine “*Teegin Gerl*” (literally *Light in the Steppe*); and spoken on some very brief radio and TV programs.

Kalmyk is a part of the core curriculum at the local schools, moreover, compulsory exams in Kalmyk in the ninth and eleventh grades have been introduced, new teaching technologies and materials are being developed. The teachers of Kalmyk and those working in national schools have special 30% bonus to their salary.

Despite the described measures, the conducted field research by Kornousova showed that the proportion of speakers fluent in Kalmyk did not exceed 6% among the younger generation. About 98% of Kalmyk pupils entering school at the age of seven do not really speak Kalmyk, thereafter most of school graduates leave schools without gaining any proficiency [Kornousova, 2001].

To address the above-mentioned challenges, new trends in the revival of the Kalmyk language are gradually being introduced into practice in the Republic of Kalmykia.

At the level of private initiatives, for example, in 2011, the employees of the Federal Treasury began to study the Kalmyk language collectively. Later, the Kalmyk language courses for municipal employees in the Lagan and Tselinny districts of the Republic of Kalmykia were launched, as well as for officials of Elista, including the representatives of the Ministry of Economics, but not limited to.

In addition to existing school textbooks, different books on the Kalmyk language have been written, and new methods for studying have been tested, such as the Kalmyk language learning as a foreign language.

Not so long ago, the two applications, which are free of charge, for smartphones have been developed and launched, the Russian-Kalmyk dictionary and “*Halmg Keln*” (*Kalmyk Language*) for both Android and iOS platforms.

Many enthusiasts open clubs for the purpose to maintain lively Kalmyk speech, in other words, establishing language centers for those who want to hear and produce speech, thus, to learn the Kalmyk language [Nuksunova, 2014].

Nevertheless, it must be concluded that the measures taken by amateur people alongside those native speakers, wholeheartedly loving their language, are not enough to revitalize the use of the Kalmyk language to a great extent.

The linguistic situation in the Republic of Kalmykia might be described as a skewed bilingual environment, with Russian having conspicuously greater prevalence than Kalmyk. Kalmyks for the most part use Russian, whether they are young, middle-aged, or elderly. Even though both languages exist in many parts of life, press, radio, and television, but as gauged by the number of functions for which one or the other

language is used, Russian far exceeds Kalmyk. In addition to all, as of now, Kalmyk language teachers do not have enough experience in teaching Kalmyk as a foreign language – how the language should be taught in the current conditions.

Overall, it is nearly impossible to agree more with the idea of the development of a science-based program with the aim to support the development of the Kalmyk language, which is endorsed by the local government with generous funding and different tools of motivation and encouragement for people. If the situation remains the same, then it can be assumed that the Kalmyk language will die out.

### References

- Avery M. *The Tea Road: China and Russia meet across the Steppe*. – Beijing: China Intercontinental Press, 2003.
- Baranova V. *The language situation in Kalmykia: Sociolinguistic study. // The research of the Kalmyk grammar. Vol. V*. – Saint Petersburg: Nauka, 2009.
- Bitkeeva A. *Social Functions of the National Language in the Modern World: The Oirat and Kalmyk Language (Unpublished Master's Thesis)*. – Moscow: Institute of Linguistics of Russian Academy of Sciences, 2006.
- Bogrdanova T. *Kalmyk-Russian Bilingualism and Translation Issues. AGON*. 2015.
- Boschaeva N. *On the role of translation and dictionaries in language promotion, Acts of the International Seminar «Minority Languages in Russia: Perspectives for Development», 2000.*
- Constitution of the Russian Federation, 25 December 1993.
- Esenova T. *Russian Speech in Kalmykia: History and Current State. Mediterranean Journal Of Social Sciences*. 2015, 6(4), 456.
- Goryaev A. & Omakayeva E.. *On Problems of Studying Minority Languages in Russia. Case Study of the Kalmyk Language. Minority Languages in Russia: Perspectives for Development*, 1999.
- Gouchinova E.-B. *The Kalmyks*. Routledge, 2013.
- Kornousova B. *Language Policy and Minority Language Planning in Russia: The Case Study of the Kalmyk Language*. 2001.
- Law of the Republic of Kalmykia № 93-V-3, December 15, 2014, “On the State Languages of the Republic of Kalmykia and other languages in the Republic of Kalmykia”.
- Moseley C. *Atlas of the World's Languages in Danger*, 3rd ed. Paris: UNESCO Publishing, 2010.
- Nuksunova A. *Revival of everyday Kalmyk language and Buddhism as factors of ethnic self-identity of contemporary Kalmyk youth (based on sociological survey). Theory and Practice of Social Development*. 2014.
- Tishkov V. *Ethnicity, Nationalism and Conflict in and After the Soviet Union*. – Oslo: International Peace Research Institute, 1997.

## ФОНЕТИЧЕСКАЯ СИСТЕМА СЛОВАРЯ «КИТАБ АД-ДУРРА»

В этой статье рассматриваются особенности фонетической системы старокыпчакского письменного памятника «Китаб ад-Дурра». Данный памятник написан в XIII–XIV веках в государстве Мамлюков, которые правили в Египте и Сирии. В ходе исследования изучены графические, фонетические, археографические и орфографические особенности арабско-кыпчакского словаря и даны оценки с точки зрения современного языкознания. Результаты исследования дают представление не только об особенностях старокыпчакского языка, но и проливают свет на некоторые факты истории, этимологии, фонетики, орфографии, графики современного казахского языка.

**Ключевые слова:** мамлюкско-кыпчакский язык, старокыпчакские письменные памятники, рукопись, словарь, вокализм, консонантизм, графика, орфография, фонетика.

«Китаб ад-Дурра ал-мудаййа фил-лугати-т-туркиййа ‘ала-т-тамам уа-л-камалу» – بأشك<sup>1</sup> «*Полностью завершённая книга сверкающей сокровищницы тюркских языков*» была исполнена в виде словаря на арабском языке, арабской графикой в XIII–XIV вв. в государстве Мамлюков, охватывающий территории Египта и Сирии. Раздел «Словосочетания и предложения», посвященный особенностям устной речи старокыпчакского языка. «Известно, что как бы не была совершенна орфография, но, она точно не передаст звуковые особенности разговорной речи» [Әбілқасымов 2001, 52]. По этой причине, очевидно, что в составленной на арабской графике артефакте могут встречаться сложности при написании слов на кыпчакском.

О сложных моментах при написании тюркских слов на арабской графике, А.М. Щербак пишет следующее: «Насколько недостаточны графические средства арабского алфавита, для передачи тюркских звуков вообще, можно судить хотя бы по тому, что три знака для гласных, один знак для **к** и **г**, не всегда передается различие между **ч** и **ј**, **п** и **б**» [Щербак 1962 64]. Такое суждение поддерживается и П.М. Мелиоранским, который пишет: «Арабский алфавит непригоден для отражения звуков тюркского языка. Это обстоятельство всегда была и будет являться большой помехой для исследований тюркоязычных памятников, дошедших до наших дней с лингвистической, особенно с фонетико-этимологической точки зрения» [Мелиоранский 1900, 29]. Ученый-кыпчаковед С. Боранбаев считает, что: «Какой бы буквой не были записаны старокыпчакские слова, можно и разобрать, и выбирать использованные в них чистые фонемные звуки» [Боранбаев 2004, 19]. Тем не менее, арабские буквы бедны для обозначения гласных звуков и в зависимости от контекста, один знак выполнял роль различных звуков, который читался и так, и эдак».

В научной среде принято, что в старокыпчакском языке имеется 8 гласных звуков. Если В.В. Радлов их считает «**коренными** звуками», то немецкий ученый А.М. фон Габэн называет «**главными** звуками» [Боранбаев 2004, 19]. По мнению А.К. Курышжанова, в кыпчакском языке есть 8 **полных** гласных звуков и 2 **полугласных** (**й**, **у**) звука. Полные гласные звуки по особенностям их произношения, делятся на «долгие гласные звуки» и «монофтонги». Несмотря на то, что долгие гласные звуки влияют на фонетический состав слова, они не меняют его лексическое содержание. То же самое относится и к звукам-дифтонгоидам. При постоянном сочетании некоторые гласные составляют «парные звуки».

Некоторые зарубежные тюркологи придерживаются мнения, что в кыпчакском языке количество гласных звуков больше восьми. Например, фон Габэн к 8 звукам добавляет еще два. Один из них, более узкий, чем звук **е**, и более открытый, чем **і**, который произносится в составе слов *eşit* (есит), *ber* (бер), *bey* (би). Вторым звуком, фон Габэн указывает на неизвестный звук, «произносящийся» между звуками **а** и **е**. Здесь автор отмечает: «по предположению некоторых европейских исследователей».

О. Притцак считает, что «среди лингвальных согласных имеется звук средний между **а** и **о**. Обычно, он передается посредством знаков дамма и алиф арабского письма, встречается в составе слов *qal* (кал), *yağurt* (сүзбе, күрт)» [Pritsak 1959, 78]. Таким образом, в итоге получаем 11 гласных звуков в кыпчакском языке. Кыпчаковед А.К. Курышжанов утверждает, что «опираясь на заключение В.В. Радлова, мы можем быть абсолютно уверены в том, что в кыпчакском языке есть 8 гласных звуков» [Айдаров, Құрышжанов, Томанов 1971, 199]. Классификация гласных звуков кыпчакского языка представлена в таблице 1.

## Гласные звуки

По участию губ	Лингвальные звуки		Билабиальные звуки	
По способу образования/ по месту образования	Открытые	Узкие	Открытые	Узкие
Твердые	А	Ы	О	У
Мягкие	Е	І	Ө	Ү

Два полугласных звука (**й, у**) старокыпчакского языка сочетаются и с твердыми гласными, и с мягкими гласными [Айдаров, Құрышжанов, Томанов 1971, 191]. Несмотря на то, что А. К. Құрышжанов эти два звука назвал «полугласными» и ввел их в разряд гласных звуков, специалист по фонетике современного казахского языка С. Мырзабеков считает, что гласные звуки **и, у** (например: бу, су, ки, қи и т.д.) не встречаются в составе исконно казахских слов. Интересно, что большинство исследователей и авторов учебников, заведомо зная, что каждый из этих двух букв является соединением двух звуков – **ұу, үу, ыу, іу, ый, ій**, относят их в разряд гласных.

А. Жунисбеков и другие ученые, исследующие гласные звуки казахского языка экспериментальным способом, эти две буквы (**й, у**), встречающиеся в составе исконно казахских слов, не считают их гласными звуками. Тем не менее, С. Мырзабеков вслед за фонетикой русского языка буквы **й** и **у** относит к разряду гласных звуков. Он пишет: «Они не имеют фонематических (различающее значение слов) свойств и поэтому считаются только одной фонемой» [Мырзабеков 1993, 18]. Несмотря на существующие точки зрения по уже сформулированным заключениям, в старокыпчакском языке 8 гласных звуков, а с учетом, что они встречаются в исконно казахских словах, то в современном казахском их – 9.

В старокыпчакском языке имеется 22 согласных звука [Құрышжанов, Көмеков, Дүйсен 2007, 145]. Академик В. В. Радлов считал, что сфера казахского консонантизма состоит из двадцати согласных и классифицировал согласные звуки на четыре акустико-артикуляционные группы: глухие (q, k, t, p, s, š), звонкие (γ, g, d, b, z, ž), сонорные (n, m, ŋ, r, l, I) и промежуточные или полугласные (w, j) [Vergleichende Grammatik...]. П.М. Мелиоранский рассматривая состав звуков, отмечает, что сфера казахского консонантизма состоит из двадцати одного согласного звука (б, п, м, т, д, н, ж, з, ш, с, р, л, к, г, ч, h, к, ғ, ң, й, й-дж) [Мелиоранский 1894]. Мы будем опираться на уже сформированное заключение по кыпчаковедению. Они представлены в таблице 2.

Таблица 2.

## Согласные звуки

Способ образования	По участию голоса		Губные согласные	Губно-зубные согласные	Зубные согласные	Переднеязычные согласные	Переднеязычные согласные	Среднеязычные согласные	Заднеязычные согласные	Гортанные согласные	
	глухие	звонкие									
Смычная	глухие	аффрикат	глухие								
			звонкие								
		аффрикат	глухие				ч				h
			звонкие								
Шипящие	сонорные	глухие		ф	с		ш		х		
		звонкие		в	з		ж		г		
		Плавные	у			л	р, й				
		Носовые	м			н			ң		

Описанные выше два полугласных (звуки **y** и **й**) рассматриваются в этом разряде [8, с. 145]. Звуки, обозначенные в таблице как **φ**, **x** используются только в составе вводных слов и принято считать их звуками, отсутствующими в старокыпчакском языке. Благодаря своим природным фонетическим особенностям, некоторые согласные звуки старокыпчакского языка обрели свое место в письменном памятнике, как «Сверкающая сокровищница». Это – **b** (ب), **p** (پ), **t** (ت), **c** (ج), **č** (چ), **h** (ح), **d** (د), **r** (ر), **z** (ز), **s** (ص), **š** (ش), **ğ** (غ), **q** (ق), **k** (ك), **g** (ك), **l** (ل), **m** (م), **n** (ن), **ŋ** (ن), **v** (و), **y** (ي), **u** (و).

Так выглядят примеры использования указанного ряда согласных звуков на языке письменного памятника, см.: *boynuzi* بۆيۈزۈيۈيۈز мүйіз, мүйізі (рог, рога) (5b,4), *budur* بۇدۇر бұдыр (шероховатый) (3b,14); *paugambar* پايغۇمبار Пайғамбар (Пророк) (2a,7), *pül* پۇل акша, пұл (деньги) (2b,16), *qarlan* قارلان кабылан (леопард) (7b,5), *qarqara* قارقارا қап-қара (черный-пречерный) (9a,13); *teñiz* تەڭىز теңіз (море) (4a,15), *Teñgiri* (Teñiri) تەڭرى Алла, Тәңірі (Аллах, Тенгри) (2a,7); *самға* قامچا шатыр қазығы (опора крыши) (9a,2), *Sanım* سەنәм Жаным – кісі есімі (имя человека) (18b,5); *čekirdek* كەڭرىكەك дән, дәнек (зерно, семечко) (5b,2), *čekirge* كەڭرىكەك шегіртке (кузнечик) (8a,2); *diz* دىز тізе (колени) (13b,6), *dost* دوست дос, жолдас (друг, спутник) (14a,8); *gast* گەست рас, тура (правда, честный) (3b,8), *quru incir* قۇرۇق ئىنجىر кептірілген інжір (сушеный инжир) (6b,13); *ziyan* زىيان зиян (вред) (4b,10), *zurna* زورنا зурнай, керней (зурна, карнай) (11b,4); *sinek* سەنەك шыбын (муха) (9b,8), *siñir* سىڭىر сиңір (сухожилие) (13b,10); *şıraq* شىراق шырақ, канделябр (лучина, подсвечник) (10a,16), *şıyam* شىيام шаян (рак) (9a,15), *şibin* شىبىن шыбын (муха) (9b,8); *qabırğa* قابىرغا қабырға (ребро, стена) (13a,16), *sağ* ساغ оңтүстік; оң жақ (юг, правая сторона) (4b,12); *qan* قان қан (кровь) (13b,4), *qanat* قانات қанат (крыло) (12a,12); *köp(i)* كۆپ мамықты ит (пушистая собака) (7b,11), *kögüs* كۆگۈس төс, төс сүйегі (грудь, грудина) (13a,16), *kök* كۆк аспан, көк жүзі (небо, небосвод) (2a,9); *Güharbay* گۇھارباي Гүһарбай – кісі есімі (имя человека) (19a,1), *gülaф* گۈل гүл (цветок) (5a,7); *lanqurt* (lanqurt) لانقۇرت шошқа құрты (свиной цепень) (9b,4), *laša* لاشا биықтың басы (плечо) (13b,12), *legün* لەگۈن леген, үлкен ыдыс, бочка (таз, большая чаша, бочка) (11a,4); *miuq* مۇيۇق миық (усы) (13a,10), *mindim* مېندىم міндім (я сел) (23a,14); *pöker* پۆكەر нөкер (нукер) (14a,8), *paqare* پاقران нығыра, нағара (литавры) (11b,4); *teñiz suu* تەڭىز سۇи теңіз суы (морская вода) (4a,13); *уанақ* واناق иектің төменгі жағы, тамақ (подбородок, горло) (13b,12), *уапсіқ* уапсіқ қалта-құлта, дорба (мешочек, торба) (12b,4); *tuzlu su* تۇزلى سۇ тұзды су (соленая вода) (4b,2), *tulquq* تۇلۇق тұлпып (тулуп) (8b,16), *tulum* تۇلۇم тұлым (коса, пучок волос) (13b,14) и т.п.

Исходя из транскрипции, звучания и смысловой особенности лексем, за исключением некоторого фонологического своеобразия, наблюдается большое сходство с казахским языком.

Тем не менее, в археографике письменного памятника, в орфограммах старокыпчакских слов встречаются различия, которые исключены в большинстве письменных памятников. Некоторые из них встречаются в графической подаче гласных звуков, а иные – в орфограммах согласных звуков. Также в письменном памятнике имеются некоторые различия в использовании диакритических символов, часто применяемых при написании старокыпчакских лексем и в разночтении их со стороны ученых-исследователей.

Результаты сопоставления археографических, орфограммических, фонологических, фонетических систем языкового материалов рукописи позволяют сделать ряд заключений.

Орфограмма звуков устной речи мамлюкских кыпчаков, отраженная в рукописи, максимально приближена к фонологической, фонетической системе тюркских языков, в частности, к старокыпчакскому и огузскому языкам. Исходя из фонологии рукописи, на момент написания, времени исполнения памятника, в устной речи мамлюкских кыпчаков имеются особенности фонологического говора, относящиеся к 1) мамлюкским кыпчакам; 2) мамлюкским тюркам (огузским тюркам); 3) представителям групп местных народов. Особенности фонологического говора устной речи мамлюкских кыпчаков, определяется использованием глухих согласных **t**, **š**, звонких согласных **c** (**ж**), звуков твердых гласных, как **й** (**ү**). Особенности фонологического говора, относящиеся к устной речи мамлюкских тюрков (огузских тюрков), определяется наличием звонкого согласного звука **d** и некоторых лексем, использованных на языке памятника, на современном турецком языке. Лексем, используемые на современном турецком языке, не используются ни на одном из кыпчакских языков.

Была обнаружена максимальная близость к современному казахскому языку фонологической системы лексем, лексического ряда, лексико-семантического применения старокыпчакского языка, которая встречается в языке памятника.

#### Литература

Айдаров Ғ., Құрышжанов Ә., Томанов М. Көне түркі жазба ескерткіштерінің тілі. – Алматы: Мектеп, 1971. – 272 б.

- Құрышжанов Ә., Көмеков Б.Е. С.Ж. Дүйсен. – Алматы: Қазақ Қыздар педагогикалық институты баспасы, 2007. – 352 б.
- Әбілқасымов Б. Әбілғазы ханның «Түркі шежіресі» және оның тілі. – Алматы, Арыс, 2001. – 246 б.
- Боранбаев С.Е. Эски кыпчак тили менен казак тилинин генетикалык катыштыгы. Автореф. дис. д-ра филол. наук. – Бишкек, 2008. – 19 с.
- Мелиоранский П.М. Краткая грамматика казак-киргизского языка. Часть I. Фонетика и этимология. – СПб., 1894.
- Мелиоранский П.М. Араб филолог о турецком языке. – Санкт-Петербург: тип. Имп. Акад. наук, 1900. – 297 с.
- Мырзабеков С. М. Қазақ тілінің фонетикасы: Оқу құралы. – Алматы: “Қазақ университеті”, 1993. – 136 б.
- Щербак А.М. Грамматика староузбекского языка. – Москва-Ленинград: Издательство АН СССР, 1962. – 274 с.
- Pritsak O. Das Kiptschakische // *Philologiae Turcicae Fundamenta*. – Wiesbaden, 1959. – Bd. 1. – Pp. 74–87.
- Vergleichende Grammatik der nördlichen Türkssprachen. Erster Theil. Phonetik der nördlichen Türkssprachen. – Leipzig, 1883. – 318 с.

## ФРАЗЕОЛОГИЧЕСКИЕ ЕДИНИЦЫ СО ЗНАЧЕНИЕМ «РАДОСТЬ» В РУССКОМ ЯЗЫКЕ

В статье на основе дефиниций существующих толковых словарей раскрывается понятие радость в русском языке, анализируются фразеологизмы со значением «радость», строится языковая картина мира. Проведенное исследование показывает малое их количество по сравнению с фразеологизмами других семантических групп. Во многих фразеологизмах видна связь с религией, причём не только с христианством, но и со старинными языческими верованиями славян.

**Ключевые слова:** радость, фразеология, русский язык, эмоции, чувства.

Радость, прежде всего, проявляется как психическая форма, которая сводит очень сильные эмоции и душевные состояния, фиксирует поведение, систему. Эмоции превращаются в лирику только тогда, когда попадают в психическую форму. От мучения рождается эстетический прорыв. То есть эмоциональная лирика не только вызывает у нас скорбь, но и также помогает преодолеть это настроение. Эстетический прорыв (эстетический прорыв для поэтов, вызывающих скорбь) – регулирует поведение человека, регулирует противоречия во внутренней вселенной, оздоравливает душу.

В условиях глобализации, определяемой как «активизация процесса усиления взаимосвязанности мира, всевозрастающего воздействия на социальную действительность отдельных стран различных факторов международного значения: экономических и политических связей, культурного и информационного обмена и т. п.» необходимо по-новому взглянуть на культурные и, в первую очередь, лингвистические ценности, имеющиеся у русского народа. Современная фразеология уделяет особенно много внимания данной проблеме. Например, Т.М. Грушевская, Н.В. Луговая пишут, что «во фразеологизме в большей мере, чем в других языковых единицах, проявляется субъективный человеческий фактор, отражающий лингвокреативный потенциал человеческого мышления» [Грушевская 2006, 29]. Л.Ю.Буянова, Е.Г. Коваленко считают, что специфику ФЕ обобщает «фразеологическая картина мира..., своеобразная сетка, накидываемая на наше восприятие, на его оценку... посредством специфических мыслительных процедур – категоризации, объективации, структуризации, языковой концептуализации» [Буянова 2012, 3]. Учитывая данные научные тенденции и наблюдения, рассмотреть в данном исследовании фразеологизмы со значением «радость» в русском языке было очень интересно и даже поучительно с точки зрения понимания русской языковой картины мира. Понятие языковой картины мира очень актуально в современной лингвистике. Под языковой картиной мира мы вслед за В. Плунонгом понимаем представление о мире, отражение мира в значениях и категориях языка, в его форме и содержании. В языке выражается мировосприятие и мировидение народа. Картина мира есть целостный глобальный образ мира, который является результатом всей духовной активности человека, а не какой-либо одной её стороны. В. Плунонган отмечает, что язык отражает общие представления говорящих на нём о том, как устроен мир. А эти представления будут лишь одной из возможных картин мира, и в разных языках они должны различаться – иногда очень сильно, иногда едва заметно, в зависимости от того, насколько совпадают культура, обычаи, традиции разных народов. Получается, что язык – своего рода зеркало, которое стоит между нами и миром [Плунонган 2002, 334].

Прежде чем анализировать фразеологизмы со значением «радость» и отражение ими русской языковой картины мира, раскроем понятие радости. Несомненно, это понятие является обозначением одного из важнейших в психике человека чувств. Наиболее полное его определение, нам кажется, дано в толковом словаре русского языка С.И. Ожегова, Н.Ю. Шведовой: «РАДОСТЬ. 1. Весёлое чувство, ощущение большого душевного удовлетворения. Испытывать радость. Доставить радость. Вне себя от радости (очень рад). С радостью помогу (очень охотно, с полной готовностью). 2. То, что (тот, кто) вызывает такое чувство. Радости жизни. Дети – радость матери. Радость ты моя! (часто в обращении). 3. Радостное, счастливое событие, обстоятельство. В семье радость: приехал сын. Случилась большая радость». Кроме того, в словарной статье приводится два фразеологизма, включающих слово радость: «На радостях (разг.) – по случаю радости, удачи. Погуляем на радостях. С какой радости? (разг. неодобр.) – чего ради, почему» [Ожегов, Шведова 1999, 640]. Все приведённые в словаре значения слова «радость» представляются нам очень важными, поэтому при работе с фразеологическим словарём мы выбирали фразеологизмы широкой семантики в соответствии с названными выше значениями.



Мы обнаружили 19 фразеологизмов, подходящих по семантике: бог миловал; бог (боже) <ты> мой!; на верху блаженства; воспрянуть духом; любо-дорого; как (будто, словно, точно) маслом по сердцу; на седьмом небе; попадать (попасть) на седьмое небо; будет и на моей [твоей, его, ее, нашей, вашей, их] улице праздник; разлюли малина; под весёлую руку; слава богу; слава те (тебе) господи; срывать цветы удовольствия; в своё удовольствие; с какой радости; собачья радость; одна радость в глазу, на радостях, не в радость (в радость).

Аналогичен по значению фразеологизм попадать (попасть) на седьмое небо в значении 'испытывать чувство восторга, блаженства, безграничного счастья и т. п. от чего-либо'. Приведённые в словарях примеры демонстрируют единство авторов в использовании данных фразеологизмов для описания психического состояния наивысшей радости, счастья героя: «Ах, как хорошо, что вы приехали! С вашим приездом я прямо на седьмое небо попала!» Ряд фразеологизмов отражает оптимизм русского народа, веру в то, что будущее будет счастливым, более радостным, чем настоящее, а настоящее, если оно грустное, всегда сменится светлым будущим. Например: будет и на моей [твоей, его, ее, нашей, вашей, их] улице праздник в значении 'выражение предвкушения радости, удачи, торжества в будущем; выражение надежды на лучшее будущее, на возможность восторжествовать над кем-либо или чем-либо'; воспрянуть духом в значении 'обрести хорошее настроение, бодрость, оживиться, преодолеть чувство подавленности, уныния'. Анализ стилистической окраски фразеологизмов со значением «радость» показывает, что большая их часть имеет разговорную окраску. При этом в словарях из всех анализируемых ФЕ помечена только *разлюли малина* (прост.) в значении 'очень хорошо, замечательно', например: «А ты, Пашка, оставайся, – сказал доктор, хлопая Пашку по плечу... У меня, брат, хорошо, разлюли малина!» (Чехов, Беглец). Нам кажется, что, несмотря на отсутствие помет, к разговорным фразеологизмам можно отнести *бог (боже) <ты> мой!; на верху блаженства; как (будто, словно, точно) маслом по сердцу; на седьмом небе; будет и на моей [твоей, его, ее, нашей, вашей, их] улице праздник* и т. п. Более высокую стилистическую окраску мы отмечаем только у ФЕ 'воспрянуть духом'. Преимущественно разговорная стилистическая окраска не выделяет фразеологизмы со значением «радость» из общей массы остальных фразеологизмов языка, большая часть из которых имеет сниженную стилистическую окраску: разговорную, просторечную. Нам показалось интересным рассмотреть отдельно фразеологизмы, имеющие в своём составе компонент «радость». В словаре под редакцией А.И. Молоткова их пять: *с какой радости* в значении 'зачем? почему? по какой причине?'; *не в радость* в значении 'не даёт удовлетворения, не приносит радости, счастья'; *одна радость в глазу* в значении 'единственное утешение'; *собачья радость* в значении 'колбаса самого низкого качества, сорта'; *на радостях* в значении 'по случаю какой-либо удачи, какого-либо радостного события; от радости'. Анализ их семантики показывает, что почти все они так или иначе связаны со значением «радость». Имели такую семантику раньше, а сейчас утратили в связи с изменением лексического значения, его метафоризацией и т. п. ФЕ *с какой радости, собачья радость*. Остальные сохранили значение «радость» и сейчас: *одна радость в глазу, на радостях, в радость*. А ФЕ *не в радость* имеет значение, противоположное значению «радость» благодаря использованию вместе с отрицательной частицей НЕ. Стилистическая окраска данных ФЕ аналогична окраске фразеологизмов, рассмотренных выше: в основном разговорная, а у ФЕ *с какой радости* и *собачья радость* – просторечная.

Если учесть, что «фразеологизмы можно трактовать как результат языковой материализации быта народа посредством включения в их образ наименования и особого понимания некоторой реалии» [Василенко 2014], то, заканчивая анализ фразеологизмов русского языка, имеющих семантику радости, можно отметить следующее. Их количество мало по сравнению с фразеологизмами других семантических групп. Возможно, это связано с историей народа, испытавшего немало трагических и печальных событий, или с суевериями, запрещавшими хвалить, хвалиться, говорить о радостных событиях, чтобы «не сглазить». Во многих фразеологизмах мы видим связь с религией, причём не только с христианством, но и со старинными языческими верованиями славян. Русские люди издавна связывали всё хорошее, происходящее в их жизни, с божественной помощью и воздавали за это хвалу высшим силам. Тем не менее, фразеологизмы со значением «радость» отражают оптимизм русского народа, его веру в счастливое будущее, которое в его сознании крепко связано с понятием радости.

### Литература

Буянова Л.Ю., Коваленко Е.Г. Русский фразеологизм как ментально-когнитивное средство языковой концептуализации сферы моральных качеств личности. – М.: ФЛИНТА; Наука, 2012. – 184 с.

Василенко А. П. Общее и частное в типологии компонентов-ихтионимов (на материале русской и французской фразеологии) // Научно-методический электронный журнал «Концепт». – 2014. – Т. 20. – С. 1921–1925.

Грушевская Т.М., Луговая Н.В. Идеографическое поле «Психоэмоциональное состояние человека» в русской и французской фразеологии: принципы построения и характеристика // Дискурсивное пространство: эволюция и интерпретации. – Краснодар: КубГУ, 2006. – С. 25–36.

Ожегов С.И., Шведова Н.Ю. Толковый словарь русского языка. – М.: Азбуковник, 1999. – 944 с.

Плунгян В. Язык и картина мира // Энциклопедия для детей. Т. 10. Русский язык. Языкознание. – М.: Аванта+, 2002. – С. 334–337.

## СРАВНИТЕЛЬНЫЙ АНАЛИЗ ЧАСТОТНЫХ СЛОВАРЕЙ КАЗАХСКОГО ЯЗЫКА

В статье анализируются имеющиеся в казахском языкознании частотные словари. Для этого проведено выделение частотных зон и исследование слов по лексико-грамматическим и семантическим разрядам. Выявлено, что вне зависимости от их объема, сверхчастотные зоны словарей одинаковы, идентичны и соотношения существительных, глаголов и прилагательных как в разрезе лексем, так и в разрезе словоупотреблений. Наблюдается некоторое различие в авторских частотных словарях, что связано с индивидуальным стилем писателей.

Частотные словари национального языка являются ценными учебными пособиями в изучении родного языка. Учитывая это, в 2016 г. был издан объединенный частотный словарь казахского языка для общеобразовательных целей объемом в 36265 лексем.

**Ключевые слова:** казахский язык, частотный словарь, лексема.

А. Жаңабекова, А. Салимова  
Алматы қ., Қазақстан

## ҚАЗАҚ ТІЛІНІҢ ЖИІЛІК СӨЗДІКТЕРІНЕ САЛЫСТЫРМАЛЫ ТАЛДАУ

Мақалада қазақ тіл білімінде баспада шыққан жиілік сөздіктері талданады. Қазақ тілінің сөздіктері жиілік зоналарға бөлінде және сөздер лексика-грамматикалық және семантикалық категориялар бойынша зерттелді. Олардың көлеміне қарамай, сөздіктердің үстірт жиіліктері бірдей, ал зат есімдер, етістіктер мен сын есімдердің арақатынасы лексемалар тұрғысынан да, сөз қолдану тұрғысынан да бірдей екендігі анықталды. Авторлардың жиілік сөздіктерінде өзіндік ерекшелік бар, ол жазушылардың жеке стилімен байланысты.

Жиілік сөздіктер ана тілін оқытуда құнды құрал болып табылады. Осыны ескере отырып, 2016 жылы жалпы білім беру мақсатында 36 265 лексемадан тұратын қазақ тілінің бірлескен сөздігі жарық көрді.

Қазақ тілінде 70-жылдары Қ.Бектаевтың басшылығымен және жетекшілігімен А.Қ.Жұбанов, С.Мырзабеков, А.Белботаев, Ә.Ахабаев т.б. зерттеушілер жиілік сөздіктер құрастыру ісімен айналысып, Абай тілі сөздігінің жиілік сөздігі, М.Әуезовтің 20 томдық шығармалар текстінің жиілік сөздіктері, сонымен қатар, ертегі тілінің, публицистика тілінің, математика терминдерінің жиілік сөздіктері жарық көрді.

Осы жиілік сөздіктерден кейін қазақ тілінде жиілік сөздіктер құрастыру ісі тоқтап қалды. 2016 жылы А.Байтұрсынұлы атындағы ТБИ Қолданбалы лингвистика бөлімі профессор Асқар Жұбановтың жетекшілігімен бұрын жарық көрген сөздіктердің басын біріктіру арқылы әртүрлі стильдерді қамтитын біртұтас жиілік сөздікті құрастырып, баспаға ұсынды [Жұбанов 2016]. Төменде сол жиілік сөздіктегі жиі кездесетін сөздердің бастапқы бөлігінен үзінді береміз. Қазақ тілінің жиілікті-әліпбилі сөздігінен» жиілігі 10 000-нан жоғары сөздер мыналар. **1-кесте.**

Көріп отырғанымыздай, *бол* және *де* тістіктерінің жиілігі өте жоғары. Жиілік бойынша жоғары қолданысты *есімдіктер*, *көмекші етістіктер*, *шылаулар* құрайды.

Бүгінгі таңда мемлекеттік тіліміздің қолданыс қызметін арттыру мақсатындағы заманауи қажеттіліктерге сәйкес қолданбалы бағыттағы ғылыми жобаларға ерекше мән беріліп отырғандығы белгілі. Бұл ретте 2016 жылы Білім және ғылым министрі Сағадиевтің тапсыруымен мемлекеттік тілді оқытудың тиімді әдістемесін жасау мақсатында лексика-грамматикалық минимумдар үшін қажетті қазақ тілінің жиілік сөздігін құрастыру ісі А.Байтұрсынұлы атындағы Тіл білімі институтына тапсырылған болатын. Осы тапсырма бойынша институт «Жалпы білім берудегі қазақ тілінің жиілік сөздігін» құрастырды [Жалпы білім берудегі... 2016]. «Жалпы білім берудегі қазақ тілінің жиілік сөздігін» құрастыруға көркем әдебиет, публицистикалық стиль, ғылыми, публицистикалық, сөйлеу стильдер мәтіндері қамтылған, яғни қазақ тілінің 5 функционалдық стилі сөздіктің базасын құрайды. 36 265 сөзтізбеден (сөз, лексикалық бірлік) тұратын бұл сөздік дәстүрлі сөздіктер түрлері болып табылатын үш түрлі құрылымда (Әліпбилі-жиілік сөздік; Кері әліпбилі-жиілік сөздік; Жиілікті-әліпбилі сөздік) жасалған. Сонымен қатар бұл сөздік әрбір стильдер бойынша да «Таралым сөздігінде»

(Распределительный словарь) жіктеліп берілген. Сонымен қатар неше түрлі сөз таптары мен стильдер бойынша түрлері электрондық нұсқада жасалған.

Кесте 1 – «Қазақ тілінің жиілікті-әліпбилі сөздігінен» үзінді

СӨЗ/сөз табы	Абсолютті жиілік	Бір топ сөздің мәтінді қамту пайызы
бол/ет.	47716	2,2695
де/ет.	40178	4,1805
бұл/ес.	26338	5,4332
е/ет.	26122	6,6756
өз/ес.	23635	7,7997
кел/ет.	22566	8,8730
ал/ет.	22522	9,9442
ол/ес.	22262	11,0031
да/шл.	22013	12,0501
де/шл.	20282	13,0147
сол/ес.	17723	13,8577
мен/ес.	17341	14,6825
айт/ет.	16410	15,4630
бір/са.	15276	16,1895
бер/ет.	14691	16,8883
отыр/ет.	12382	17,4772
осы/ес.	11818	18,0393
жоқ/мд.	11415	18,5822
сөз/зт.	11309	19,1201
түр/ет.	11022	19,6443
кет/ет.	10635	20,1501
жүр/ет.	10555	20,6522
бір/ес.	10344	21,1442

«Жалпы білім берудегі қазақ тілінің жиілік сөздігінің» жиілікті-әліпбилі сөздік түрінен ең жиі кездесетін жиілігі 20 000-нан жоғары сөздерден үзінді берейік. **2-кесте.**

Бұл сөздікте де *бол* көмекші етістігінің жиілігі басқаларына қарағанда өте жоғары. Бұл кестеден де көмекші *етістіктер, шылаулар* мен *есімдіктердің* жоғары қолданыстағы сөздер екендігін байқаймыз. Сонымен қатар жоғарыда берілген біріктірілген «Қазақ тілінің жиілік сөздігіндегі» көрсеткіштермен ұқсас.

Жиілік сөздік жасаудың әдіс-тәсілдері әртүрлі. Соның бірі – мәтінді сөзформалар сөздігіне түсіріп, сөзді түбірге келтіру. Екіншісі – мәтінді морфологиялық анализатор арқылы өткізіп, түбір мен оған қойылатын сөз табы белгісімен бірге автоматты жолмен алу. Біріншісі көп қол жұмысын керек етеді. Мұндайда қосымша жалғанғанда түбір сөздердің омонимдерін ажырату мүмкін болса, түбір формадағы сөздерге келгенде олардың сөз табына қатысын анықтау қиын болады.

Кесте 2 – «Жалпы білім берудегі қазақ тілінің жиілікті-әліпбилі сөздігінен» үзінді

СӨЗ/сөз табы	Барлығы
бол/ет	108133
ол/ес	62796
мен/шл	58406
де/ет	46200
ал/ет	43359
бер/ет	40295

және/шл	39765
кел/ет	34526
де/шл	33943
бұл/ес	32850
да/шл	32109
жыл/зт	30882
бір/ес	30288
өз/ес	29594
тұр/ет	28643
сөз/зт	26107
осы/ес	25995
адам/зт	25614
қазақ/зт	25069
айт/ет	23421
білім/зт	22915
үшін/шл	22598
кет/ет	22105
ел/зт	22098
отыр/ет	22038
қандай/ес	21378
екен/ет	21328
біз/ес	21267
сол/ес	20532

Екіншісі оңай, тез жүзеге асырылғанмен, жүз пайыздық шынайылықты бермейді. Бұл әдісте те омонимдер автоматты түрде ажыратылмайды. Автоматты жолмен омонимдері ажыратылған жиілік сөздіктерді тек омоним ажыратылған мәтіндерден ғана алуға болады.

Осы орайда Қолданбалы лингвистика бөлімі қызметкерлері проза жанры (М.Әуезов, Ә. Кекілбаев т.б. ) бойынша 500 000 сөзқолданыстан тұратын мәтіндерге «қолдап» морфологиялық белгіленім қойып, сол мәтіннен, яғни сөз таптарына қатысы 100 пайыз дұрыстықты көрсететін мәтіннен жиілік сөздік құрастырады [Қазақ сөзформа құрылымының жиілік сөздігі 2019]. Осы шағын жиілік сөздіктің үзіндісімен таныс болайық.

Кесте 3 – Қазақ тіліндегі сөздің жиілікті-әліпбилі сөздігінен үзінді

Рет саны №	Сөз/Сөз табы	Абсолютті жиілік	Бір топ сөздің мәтінді қамту пайызы	Рет саны №	Сөз/Сөз табы	Абсолютті жиілік	Бір топ сөздің мәтінді қамту пайызы
1	2	3	4				
1	<i>бол/ет.</i>	666	2,320	14	<i>жат/ет.</i>	218	16,701
2	<i>де/ет.</i>	541	4,204	15	<i>ол/ес.</i>	215	17,450
3	<i>еді/ет.</i>	479	5,873	16	<i>осы/ес.</i>	214	18,196
4	<i>кел/ет.</i>	407	7,290	17	<i>көр/ет.</i>	193	18,868
5	<i>бір/са.</i>	352	8,516	18	<i>қал/ет.</i>	179	19,491
6	<i>ал/ет.</i>	335	9,683	19	<i>үй/зт.</i>	178	20,111
7	<i>да/шл.</i>	334	10,846	20	<i>айт/ет.</i>	176	20,724
8	<i>бұл/ес.</i>	299	11,888	21	<i>бер/ет.</i>	171	21,320
9	<i>тұр/ет.</i>	240	12,724	22	<i>жоқ/мд.</i>	166	21,898
10	<i>өз/ес.</i>	238	13,553	23	<i>күн/зт.</i>	166	22,476
11	<i>де/шл.</i>	231	14,357	24	<i>шық/ет.</i>	163	23,044
12	<i>сол/ес.</i>	231	15,162	25	<i>кет/ет.</i>	161	23,605
13	<i>отыр/ет.</i>	224	15,942	26	<i>жүр/ет.</i>	144	24,107
				27	<i>көз/зт.</i>	144	24,608

Рет саны №	Сөз/Сөз табы	Абсолютті жиілік	Бір топ сөздің мәтінді қамту пайызы
28	<i>жер/зт.</i>	143	25,106
29	<i>қара/ет.</i>	136	25,580
30	<i>іш/зт.</i>	135	26,050
31	<i>бала/зт.</i>	132	26,510
32	<i>бас/зт.</i>	131	26,966
33	<i>біл/ет.</i>	119	27,381
34	<i>көп/сн.</i>	118	27,792
35	<i>сөз/зт.</i>	115	28,192
36	<i>бар/ес.</i>	114	28,589
37	<i>ауыл/зт.</i>	113	28,983
38	<i>бар/ет.</i>	113	29,377
39	<i>мен/шл.</i>	113	29,770
40	<i>түс/ет.</i>	109	30,150
41	<i>үлкен/сн.</i>	109	30,529
42	<i>соң/шл.</i>	105	30,895
43	<i>қыл/ет.</i>	99	31,240
44	<i>бар/мд.</i>	98	31,581
45	<i>жақ/зт.</i>	97	31,919
46	<i>ат/зт.</i>	94	32,247
47	<i>екі/са.</i>	91	32,564
48	<i>ет/ет.</i>	90	32,877
49	<i>сал/ет.</i>	89	33,187
50	<i>қой/ет.</i>	88	33,494
51	<i>не/ес.</i>	86	33,793
52	<i>алды/ке.</i>	84	34,086
53	<i>ел/зт.</i>	84	34,378
54	<i>баста/ет.</i>	81	34,660
55	<i>өт/ет.</i>	81	34,943
56	<i>кісі/зт.</i>	79	35,218
57	<i>бірақ/шл.</i>	75	35,479
58	<i>мен/ес.</i>	74	35,737
59	<i>кір/ет.</i>	73	35,991
60	<i>ақ/сн.</i>	72	36,242
61	<i>қол/зт.</i>	71	36,489
62	<i>енді/үс.</i>	67	36,722
63	<i>жол/зт.</i>	67	36,956
64	<i>адам/зт.</i>	64	37,179
65	<i>бет/зт.</i>	63	37,398
66	<i>күл/ет.</i>	63	37,618
67	<i>қара/сн.</i>	63	37,837
68	<i>жалғыз/сн.</i>	61	38,049
69	<i>жан/зт.</i>	61	38,262
70	<i>мал/зт.</i>	61	38,474
71	<i>ғой/шл.</i>	60	38,683
72	<i>қас/зт.</i>	60	38,892
73	<i>тап/ет.</i>	60	39,101
74	<i>аз/сн.</i>	59	39,307
75	<i>үсті/ке.</i>	59	39,512

Рет саны №	Сөз/Сөз табы	Абсолютті жиілік	Бір топ сөздің мәтінді қамту пайызы
76	<i>жігіт/зт.</i>	58	39,714
77	<i>қарай/шл.</i>	58	39,916
78	<i>ғана/шл.</i>	57	40,115
79	<i>бас/ет.</i>	56	40,310
80	<i>және/шл.</i>	56	40,505
81	<i>тарт/ет.</i>	56	40,700
82	<i>қыз/зт.</i>	55	40,892
83	<i>күйеу/зт.</i>	54	41,080
84	<i>жас/сн.</i>	53	41,264
85	<i>жет/ет.</i>	53	41,449
86	<i>кез/зт.</i>	53	41,634
87	<i>қатты/сн.</i>	51	41,811
88	<i>тағы/шл.</i>	51	41,989
89	<i>шеше/зт.</i>	51	42,166
90	<i>керек/мд.</i>	49	42,337
91	<i>бой/зт.</i>	48	42,504
92	<i>тіл/зт.</i>	48	42,672
93	<i>жыл/зт.</i>	47	42,835
94	<i>уақыт/зт.</i>	47	42,999
95	<i>басқа/сн.</i>	45	43,156
96	<i>біз/ес.</i>	44	43,309
97	<i>суық/сн.</i>	43	43,459
98	<i>түс/зт.</i>	43	43,608
99	<i>әлі/үс.</i>	42	43,755
100	<i>ойла/ет.</i>	42	43,901
101	<i>түн/зт.</i>	42	44,047
102	<i>ақ/шл.</i>	41	44,190
103	<i>арт/ет.</i>	41	44,333
104	<i>қалың/сн.</i>	41	44,476
105	<i>әңгіме/зт.</i>	40	44,615
106	<i>жібер/ет.</i>	40	44,754
107	<i>мына/ес.</i>	40	44,894
108	<i>әйел/зт.</i>	38	45,026
109	<i>ки/ет.</i>	38	45,158
110	<i>көңіл/зт.</i>	38	45,291
111	<i>көш/зт.</i>	38	45,423
112	<i>тау/зт.</i>	38	45,556
113	<i>әке/зт.</i>	37	45,684
114	<i>жай/зт.</i>	37	45,813
115	<i>қазақ/зт.</i>	37	45,942
116	<i>қала/зт.</i>	37	46,071
117	<i>қар/зт.</i>	37	46,200
118	<i>тай/зт.</i>	37	46,329
119	<i>ара/зт.</i>	36	46,454
120	<i>аш/ет.</i>	36	46,580
121	<i>барлық/ес.</i>	36	46,705
122	<i>қызыл/сн.</i>	36	46,830
123	<i>сен/ес.</i>	36	46,956

Рет саны №	Сөз/Сөз табы	Абсолютті жиілік	Бір топ сөздің мәтінді қамту пайызы
124	<i>су/зт.</i>	36	47,081
125	<i>жақсы/сн.</i>	35	47,203
126	<i>орта/зт.</i>	35	47,325
127	<i>соқ/ет.</i>	35	47,447
128	<i>есті/ет.</i>	34	47,565
129	<i>жұрт/зт.</i>	34	47,684
130	<i>қайт/ет.</i>	34	47,802
-----	-----	-----	-----

Рет саны №	Сөз/Сөз табы	Абсолютті жиілік	Бір топ сөздің мәтінді қамту пайызы
4515	<i>істік/зт.</i>	1	99,979
4516	<i>іші-бауыр/зт.</i>	1	99,983
4517	<i>ішіп-же/ет.</i>	1	99,986
4518	<i>экономикалық/сн.</i>	1	99,990
4519	<i>январь/зт.</i>	1	99,993
4520	<i>япырай/од.</i>	1	99,997
4521	<i>япырау/од.</i>	1	100
	<b>Барлығы:</b>		28710

Бұл шағын сөздіктегі сөздердің жиілікті-әліпбилі сөздігі бойынша, **3-кестеде** мәтіннің 50 пайызын қамтитын жиі қолданылатын сөздер тізімі (150-ші топ деген сөзге дейін) толықтай берілген, ал 60, 70, 80, 90, 100 пайызын қамтитын сөздер үзік түрінде қысқартылған. Жиілігінің кему тәртібіне қарай сөздер 10 топқа топтастырылып берілген. Ең алғашқы топқа жиілік сөздіктегі ең жиі кездесетін «бол» етістігі мәтін ішінде **666 рет** қолданып, жалғыз өзі-ақ мәтіннің **2,320** пайызын қамтыған. Болетістігінің ең жоғарғы жиілікті көрсетуі бұдан бұрын шыққан сөздіктерде де орын алған. Бол етістігінің жоғары жиілікті көрсетуі оның көмекші етістік болып, грамматикалық мағына үстеуші қызметімен тікелей байланысты деуге болады. Өйткені жиі қолданылу, грамматикалық абстракция жасайтын грамматикалық бірліктерге тән қасиет. Жиілік сөздікте **бол** етістігі сияқты жоғары жиілікте тұрған көп сөз осы көмекші етістіктер. Мысалы, *де, еді, кел, ал, тұр, отыр, жатыр, көр, айт, бер, шық, кет, жүр, бар, түс, қыл, сал, қой, баста* т.б. Сол сияқты жиілік сөздікте шылаулардан *да, де, мен, бірақ, гой, қарай, гана, және, тағы* шылаулары, сондай-ақ есімдіктер (*бұл, өз, сол, ол, осы, бар, мен*) жиі кездескен. Бұл жиілік сөздік тек көркем шығарма мәтіндерінен алынғандықтан, термин сөздер аз, тіптен кездеспейді десе де болады.

Енді С.Сәдуақасұлының жиілік сөздігіндегі жиі кездесетін сөздердің бастапқы бөлігінен үзінді беру арқылы жоғарыдағы сөздіктермен салыстырамыз.

Кесте 4 – С.Сәдуақасұлы шығармаларындағы сөздің жиілікті-әліпбилі сөздігінен үзінді

№	Сөз	Абсолютті жиілік	Қатынастық жиілік	Қосынды қатынастық жиілік	Бір топ сөздің мәтінді қамту пайызы
1	бол/ет	3633	0,02757	0,02757	2,75670
2	де/ет	3188	0,02419	0,05176	5,17574
3	қазақ/зт	1686	0,01279	0,06455	6,45506
4	ол/ес	1574	0,01194	0,07649	7,64941
5	ал/ет	1422	0,01079	0,08728	8,72841
6	да/шл	1338	0,01015	0,09744	9,74368
7	бұл/ес	1302	0,00988	0,10732	10,73163
8	кел/ет	1302	0,00988	0,11720	11,71958
9	бір/ес	1270	0,00964	0,12683	12,68325
10	біз/ес	1213	0,00920	0,13604	13,60367
11	керек/мд	1157	0,00878	0,14482	14,48159
12	айт/ет	1151	0,00873	0,15355	15,35496
13	де/шл	1087	0,00825	0,16180	16,17977
14	осы/ес	941	0,00714	0,16894	16,89380
15	жок/мд	938	0,00712	0,17606	17,60555

№	Сөз	Абсолютті жиілік	Қатынастық жиілік	Қосынды қатынастық жиілік	Бір топ сөздің мәтінді қамту пайызы
16	отыр/ет	909	0,00690	0,18295	18,29529
17	жер/зт	879	0,00667	0,18962	18,96227
18	еді/ет	864	0,00656	0,19618	19,61787
19	мен/шл	820	0,00622	0,20240	20,24008
20	сөз/зт	813	0,00617	0,20857	20,85698
21	өз/ес	810	0,00615	0,21472	21,47161
22	көп/сн	769	0,00584	0,22055	22,05512
23	бар/ес	747	0,00567	0,22622	22,62194
24	бер/ет	743	0,00564	0,23186	23,18572
25	бәрі/ес	685	0,00520	0,23705	23,70550
26	ауыл/зт	668	0,00507	0,24212	24,21237
27	жүр/ет	631	0,00479	0,24691	24,69117
28	екен/ет	608	0,00461	0,25153	25,15252
29	түр/ет	599	0,00455	0,25607	25,60704
30	қал/ет	591	0,00448	0,26055	26,05548
31	емес/ет	587	0,00445	0,26501	26,50090
32	шық/ет	577	0,00438	0,26939	26,93872
33	біл/ет	573	0,00435	0,27374	27,37351
34	кет/ет	571	0,00433	0,27807	27,80678
35	сен/ес	552	0,00419	0,28226	28,22564
36	жат/ет	548	0,00416	0,28641	28,64145
37	жұмыс/зт	547	0,00415	0,29057	29,05652
38	ел/зт	542	0,00411	0,29468	29,46778
39	күн/зт	525	0,00398	0,29866	29,86615
40	іш/зт	519	0,00394	0,30260	30,25996
41	мен/ес	507	0,00385	0,30645	30,64467
42	адам/зт	481	0,00365	0,31010	31,00965
43	қара/ет	474	0,00360	0,31369	31,36932
44	не/ес	474	0,00360	0,31729	31,72899
45	сол/ес	461	0,00350	0,32079	32,07879
46	орыс/зт	435	0,00330	0,32409	32,40887
47	көр/ет	423	0,00321	0,32730	32,72984
48	шаруа/зт	419	0,00318	0,33048	33,04777
49	халық/зт	415	0,00315	0,33363	33,36267
50	іс/зт	408	0,00310	0,33672	33,67226
51	мәселе/зт	402	0,00305	0,33977	33,97730
52	үй/зт	402	0,00305	0,34282	34,28233
53	жас/зт	398	0,00302	0,34584	34,58433
54	жұрт/зт	396	0,00300	0,34885	34,88482
55	бірақ/шл	384	0,00291	0,35176	35,17619
56	жақ/зт	380	0,00288	0,35465	35,46453
57	нәрсе/зт	380	0,00288	0,35753	35,75288
58	қол/зт	378	0,00287	0,36040	36,03970
59	үлкен/сн	378	0,00287	0,36327	36,32652
60	ұйым/зт	374	0,00284	0,36610	36,61031
61	екі/са	370	0,00281	0,36891	36,89107
62	жол/зт	366	0,00278	0,37169	37,16879
63	туралы/шл	359	0,00272	0,37441	37,44119



№	Сөз	Абсолютті жиілік	Қатынастық жиілік	Қосынды қатынастық жиілік	Бір топ сөздің мәтінді қамту пайызы
64	түс/ет	358	0,00272	0,37713	37,71284
65	сал/ет	357	0,00271	0,37984	37,98373
66	көз/зт	355	0,00269	0,38253	38,25310
67	кыл/ет	349	0,00265	0,38518	38,51792
68	бар/ет	341	0,00259	0,38777	38,77667
69	бала/зт	327	0,00248	0,39025	39,02480
70	біреу/ес	322	0,00244	0,39269	39,26913
71	басқа/сн	312	0,00237	0,39506	39,50587
72	үшін/шл	311	0,00236	0,39742	39,74186
73	мал/зт	310	0,00235	0,39977	39,97708
74	тұрмыс/зт	308	0,00234	0,40211	40,21079
75	алды/ке	303	0,00230	0,40441	40,44071
76	оқы/ет	301	0,00228	0,40669	40,66910
77	сондықтан/шл	300	0,00228	0,40897	40,89674

С.Сәдуақасұлы шығармалары мәтіндерінің жиілік сөздігі бойынша да ерекше көзге түсетін басқа жиілік сөздіктермен ұқсас жайт – *бол* көмекші етістігінің жиілігі бойынша тағы да бірінші орынға шығуы. 4 сөздіктегі жиілік мәліметтерді салыстырсақ, *де* етістігінің жиілігі бойынша екінші орында тұруы үш сөздікте бірдей де, тек «Жалпы білім берудегі қазақ тілінің жиілік сөздігінде» 4 орынға кеткен.

Басқа 3 сөздіктен С.Сәдуақасұлы шығармаларының жиілік сөздігіндегі үшінші орында тұрған сөз – *қазақ* сөзінің болуы жазушының осы сөзді жиі қолданғанын көрсетеді.

4 сөздікке де ортақ сипат – есімдіктердің (*ол, бұл, бір, біз, осы, өз, сол* т.б.) жоғары жиілікте кездесуі. Сонымен қатар 4 сөздікте де көмекші етістіктердің (*кел, кет, ал, отыр, тұр, жүр, бер, айт, көр, қал* т.б.) жиі қолданысы, яғни жиілікті сөздіктегі жоғарғы қатардан орын алуы. Сондай-ақ *да/де* шылауының жиі кездесетінде қатарында орын алуы да ортақ.

Жекелеген атауыш сөздерге келсек:

– 500 сөзқоланыстан алынған жиілік сөздікте (негізінен М.Әуезов, Ә.Кекілбаев шығармаларының мәтіні) – *үй, күн, көз, бала, бас, сөз, ауыл, ат* сөздері;

– «Жалпы білім берудегі қазақ тілінің жиілік сөздігінде» – *жыл, адам, қазақ, білім* сөздерінің жиі кездесуі;

– біріктірілген «Қазақ тілінің жиілік сөздігінде» – *сөз, күн, жер, үй, бас, адам, бала, халық* сөздерінің;

– С.Сәдуақасұлының шығармаларында – *қазақ, сөз, жер, ауыл, жұмыс, күн, адам, шаруа, халық, үй, іс, жұрт, мәселе, ұйым* сөздерінің жиі кездесуі байқалады.

Сонымен қатар сын есім сөздерден барлығына ортақ жиі кездесетін сөз – *көп* және *үлкен* сөздері.

Бұл сөздіктердегі атауыш сөздерге қатысты жиі кездесетін ортақ сөздер – қазақ халқына жақын, маңызды ұғымдар екенін көреміз. Олар: *қазақ, сөз, жер, ауыл, үй, халық, күн, бала, ат, адам* т.б.

М.Әуезовтің 20 томдық шығармалар жинағы бойынша әртүрлі сөздіктер алу жұмысын ҚР ҰҒА-ның А.Байтұрсынұлы атындағы Тіл білімі институтының Статистика тобы 1986 жылы қолға алған болатын. Сол жұмыстың нәтижесінде «М.О.Әуезовтің 20 томдық шығармалар текстерінің жиілік сөздіктері» (22 баспа табак) 1995 жылы жарық көрді [Бектаев, Жұбанов., Мырзабеков 1995].

Кесте 5 – М.Әуезовтің 20 томдық шығармаларының жиілік сөздігінен үзінді

СӨЗ/сөз табы	Барлығы
бол/ет	38944
де/ет	31901
бұл/ес	22893

СӨЗ/сөз табы	Барлығы
е/ет	21649
өз/ес	20471
да/шл	18254
ал/ет	17958
кел/ет	17797
де/шл	17293
ол/ес	17269
сол/ес	15599
мен/ес	14717
айт/ет	14273
бер/ет	11925
бір/ес	10293
бір/са	10244
отыр/ет	10082
тұр/ет	8742
сөз/зт	8575
кет/ет	8571
жүр/ет	8536
шық/ет	8221
қал/ет	7974
көр/ет	7628
бар/ет	7378
көп/сн	7302

Тізбедегі 29483 сөз (жалқы есімді қоспағанда) жиіліктерінің кему ретімен орналасқан.

Бұл жиілік сөздікте де бірінші орында «**бол**» етістігі тұр. Ол М.Әуезовтің шығармалар жинағының тексінде 38944 ретқайталанып кездеседі екен, бұл барлық текстің 2,28%-ін құрайды. Екінші орында «**де**» етістігі тексте 31901 рет кездеседі (1,87%). Осы екі етістік барлық текстің 4,15%-ін қамтитынын байқауға болады. Сөздіктегі осы сияқты ең жиі кездесетін әртүрлі сөздердің барлық сөзқолданыстың қанша процентін қамтитынын жиілік сөздіктен көруге болады. Тек қана тізбедегі алғашқы 10 сөз (**бол/ет., де/ет., бұл/ес., е/ет., өз/ес., да/шл., ал/ет., кел/ет., де/шл., ол/ес.**) барлық текстің 13%-ін қамтиды. Ал осы сияқты сөздіктің бас жағында тұрған 31 сөз текстің 1/4-ін яғни 25%-ін, 174 сөз -1/2-ін, яғни 50%-ін, 1292 сөз – 80%-ін қамтитынына көз жеткізуге болады.

Демек, М.Әуезовтің жиілік сөздігіндегі ең жиі қолданылатын сөздер жоғарыда салыстырылған 3 сөздік және С.Сәдуақасұлының шығармаларының жиілік сөздігімен ұқсас болып келеді.

Сонымен қатар жоғарыда аталған барлық сөздікте сөз таптарына қатысты мына жайтты аңғаруға болады. Ол – **зат есім, етістік, сын есімдердің** жиі кездесуі. С.Сәдуақасұлы шығармаларында **есімдіктер** сын есімдерге қарағанда мәтінде жиі кездескен. Сондықтан тұрақты жоғары жиілікті зат есімдер мен етістіктер алып тұр деп тұжырымдауға болады. Бұл деректерден қазақ тілінің сөз байлығы зат есім мен етістік сөздер екендігіне көз жеткіземіз.

#### Әдебиеттер

Бектаев Қ.Б., Жұбанов А.Қ., Мырзабеков С., Белботаев А.Б. М.О.Әуезовтің 20 томдық шығармалар текстерінің жиілік сөздіктері. – Алматы-Түркістан, 1995. – 346 б.

Жалпы білім берудегі қазақ тілінің жиілік сөздігі. – Алматы, 2016. – 1472 б.

Қазақ сөзформа құрылымының жиілік сөздігі. – Алматы, 2019.

Жұбанов А., Жаңабекова А., Карбозова Б., Қожахметова А. Қазақ тілінің жиілік сөздігі. – Алматы: «Қазақ тілі» баспасы, 2016. – 665 б.

## ЯЗЫКОВЫЕ ОСОБЕННОСТИ БАШКИР САЛАВАТСКОГО РАЙОНА РЕСПУБЛИКИ БАШКОРТОСТАН

В данной статье анализируются языковые особенности башкир Салаватского района РБ на фонетическом, лексическом и морфологическом уровнях.

Язык данного населения относится к айскому говору восточного диалекта башкирского языка и обнаруживает основные характеристики восточного диалекта.

Спецификой говора является *ж*-анлаут, который считается более древней формой и был представлен гораздо шире, чем в современный период. Особенностью говора является также употребление сокращенных форм: 1) глагола 2 л. настоящего времени: *бараң* ‘ты идешь’ – лит. *бараһың, барағыз* ‘вы идете’ – лит. *бараһығыз*; 2) побудительного наклонения: *барың* ‘иди, идите’, *китең* ‘уходи, уходите’ и т.д.

**Ключевые слова:** айский говор, Салаватский район, анлаутный «ж», языковые особенности.

Л. К. Ишкильдина  
Өфө к., Рәсәй

## БАШКОРТОСТАН РЕСПУБЛИКАҢЫ САЛАУАТ РАЙОНЫ БАШКОРТТАРЫНЫҢ ТЕЛ ҮЗЕНСӘЛЕКТӘРЕ

Салауат районы Башкортостандың төньяк-көнсығышында урынлашкан. Башкорт ауылдары исемлегенә түбәндәгеләр инә: Аркауыл, Ахун, Башкорт Илсекәйе, Бишәүзәр, 1-се Изелбай, 2-се Изелбай, Изрис, Илтәй, Ишембай, Йонос, Көсәпәй, Комсомол, Күсәләр, Калмаккол, Калмактар, Карағол, Кызырбәк, Лағыр, Малаяз, Миндеш, Мөсәт, Мөрсәлим, Мәсетле, Мәхмүт, Радио, Сыбаркул, Ташауыл, Тирмән, Урмансы, Шәрәк, Әй, Әлкә, Юлай, Язғыйорт, Янғанту, Яңы Картаулы, Яңы Сүрәкәй, Яубүләк, Яхъя. Бында шайтан көзәй, кыр көзәй, мырзалар, тырнаклы, түбәләс, табын ырыулары башкорттары йәшәй. Салауат районында башкорттар 67 % тәшкил итә.

Салауат районы башкорттарының теле көнсығыш диалекттың Әй һөйләшенә карай. Был һөйләш дөйөм көнсығыш диалектының һызаттарын үз эсенә ала. Улар иҫәбенә: 1) 4 төрлө күплек ялғауы: *-тар/ -тәр, -лар /-ләр, -зар /-зәр, -дар/ -дәр*; 2) *-как /-кәк* сифат ялғауы (*йокак, кыскак, йәбешкәк*); 3) *-тык/ -тек, -дык /-дек* исем яһаусы ялғаулар (*аштык, урындык, утындык*); 4) *-та/ -тә* кылым яһаусы ялғауы (*эштәү, йоктау, баштау*) һ.б. инә.

Тел үзенсәлектәренә ентекләберәк тукталып үтәйек.

### Фонетик үзенсәлектәр

Әй һөйләшендә башкорт әзәби теленә хас булғанса туғыз һузынқы фонема бар: [a], [ә], [o], [ө], [ы], [е], [у], [ү], [и].

«А» өнә – иренләшмәгән арткы рәт түбәнге күтәрелеш һузынқыһы: *ак, азак, асык, арна* һ.б. Был һөйләштә [a] фонемаһы әзәби телдәге һәм башка һөйләштәрзәге [ә] урынында килә: *асы – әз. әсе, асыткы – әз. әсетке, арсыу – әз. әрсеу, һалмак – әз. һәлмәк, сасрай – әз. сәсрәй*. Шул йәһәттән был райондың башкорттарын “каты һөйләшә” тизәр.

Шулай ук «а» һузынқыһы 3-сө зат алмаштарында әзәби телендәге «у» урынында кулланыла: *аны – әз. уны, анан – әз. унан, аңа – әз. уға, анда – әз. унда, аның – әз. уның* һ.б.

[И] фонемаһы башка һузынқылар урынында килә ала: *сискә – әз. сәскә, сирыш – әз. сырыш, киңәш – әз. кәңәш*.

Был һөйләштә өй дифтонгыһы ү монофтонгына әйләнә: *үрә – әз. өйрә, үрзәк – әз. өйрәк, үрәнсек – әз. өйрәнсек, үрәтеу – өйрәтеу* һ.б.

Әй һөйләшендә тартынқы өндәр системаһы төп башкорт һүззәрәндә 24 фонема менән билдәләнә: [б], [w], [г], [ғ], [д], [з], [й], [ж], [з], [к], [к], [л], [м], [н], [ң], [п], [р], [с], [с], [т], [ф], [х], [һ], [ш].

[Б] фонемаһы әзәби телдән айырмалы рәүештә һүз азағында килә ала: *мәктәб – әз. мәктәп, китаб – әз. китап, хараб – әз. харап* һ.б.

Һүз уртаһындағы шартлаулы ике ирен «п» тартынқыһы спирантлаша *n > ф*: *керпек > керфек, япрак > яфрак, тупрак > туфрак*.

Һүз башындағы [т] тартынкыһы был һөйләштә яңғыраулана – т > д: *дөрөү* – *эз. төрөү, дән* – *эз. тән, дәрән* – *эз. тәрән, дулкын* – *эз. тулкын* һ.б.

Фарингаль [һ] фонемаһының кулланылышы был һөйләштә эзәби телдән айырыла:

1) Ул һүз башындағы [с], [ш] урынында килә ала: *һүмес* – *эз. сүмес, һабага* (жребий) – *эз. шыбага, һыйык* – *эз. шыйык*.

2) Ижек башындағы «һ» урынына «с» әйтелә: *бауырсақ, тунса, үләкә, ақсақ, тансык*.

3) Үзләштерелгән һүзәрзә һүз башындағы «һ» өнө төшөп калырға мөмкин: *әйбәт, өнәр, алакәт* – *эз. һәйбәт, һөнәр, һәлакәт*.

Был һөйләшкә редукция күренеше хас – һүз эсендәге «л», «н», «р» сонор өндәрә төшөп кала: *буган* – *эз. булган, тугайны* – *эз. торгайны, һаган* – *эз. һалган, муса* – *эз. мунса, китерә* – *эз. килтерә* һ.б.

Салауат районының иң төп фонетик үзенсәлектәрәненә берене – һүз башындағы «ж» өнөнөн кулланылыуы. Арткы рәт һузынкылары менән килгәндә был тартынкы каты, ә алгы рәт һузынкылары менән – йомшак әйтелә. Ул эзәби телендәге «й» өнө урынында килә: *жай* – *эз. йай, жарыу* (һуғыу, тукмау) – *эз. ярыу, жайзау* – *эз. йайзау, жим* – *эз. ем, жиң* – *эз. ең, жите* – *эз. ете* һ.б. Ул бөтөн һүзәрзә лә «й»-ны алыштырып килмәй: *еләк, ер, еп, йөрәк, йылан* кеүек һүзәрзә «й» өнө һаклана. Әммә Н.Х. Мәксүтова әйтеп үтеүенсә: “Экспрессив әйтелештә был һүзәр *ж* менән әйтелә: *кара жылан! жир жоткыры! жаман кара!*” [Мәксүтова 1976, 40]. Экспедиция күзәтеүәрә буйынса, ололар телендә [ж] фонемаһы күберәк һүзәрзә кулланылыуы асыкланды (*жәйәү, жите, житәү, жиңеү, жиң, жыйырсык, жәй* һ.б.), ә йәшерәк быуында ул юғала бара. Тимәк, Әй һөйләшәндә боронго вақытта «ж» өнө кинерәк кулланылышта булған һәм, бәлки, тел уртаһы «й» тартынкыһы һүз башында бөтөнләй булмаған (йә һирәк кенә һүзәрзә кулланылған) тигән һығымта яһай алабыз.

Төрки телдәрәндә *й* – *ж* сиратлашыуы хәл ителмәгән проблемалар рәтенә инә. Бер төркөм ғалимдар, һүз башындағы *й* – өнөн боронғорак тип караһа, икенсә төркөм – *ж* йәки *дж* аффрикатаһы беренсел тип исәпләй. Әммә ләкин һуңғы тикшеренеүәрә буйынса Мәскәү ғалимдары бында бөтөнләй икенсә төрлө катмарлы фонема булған тип исәпләйзәр һәм уны \**з*–тамғаһы менән билдәләйзәр [кара СИГТЯ, 2006]. Ошо \**з*–фонеманан кыпсақ телдәрәндә *дж* аффрикатаһы барлыкка килгән. Тимәк, боронго кыпсақ осоро өсөн *дж* аффрикатаһы тергезелә һәм уның артабан таркалыуы аркаһында «ж» өнө барлыкка килгән. Был күренеш татар, казак, кыргыз, каракалпак телдәрәндә күзәтелә. Әммә шуны әйтергә кәрәк, татар эзәби телендә *ж* өнө бөтөн һүзәрзә лә кулланылмай: «ө», «о» һузынкылары алдынан «й» килә (*йөгән, йөзек, йодрык, йомшак*), «а» алдынан килгәндә «й» өнө *я* хәрәфе менән билдәләнә (*яңа, яфрак, ябык* һ.б.).

Тимәк, Әй һөйләшәндәге [ж] фонемаһы ул кыпсақ телдәрәнә хас булған боронго фонетик күренеш. Кызғансыка каршы был тартынкы күп һүзәрзә эзәби телдәге тел уртаһы сонор [й] фонемаһына алмаштырыла.

Тюркологияла, кыпсақ телдәрә өсөн, ә был төркөмгә башкорт теле лә инә, һүз башындағы «й» өнө – яңы күренеш тип карала [СИГТЯ 2006, 25]. Әммә башкорт теленә көнъяк диалектындағы «й» өнө боронго кыпсақ «ж» (< *дж*) тартынкыһынан үсешкәнме, әллә ул боронго төрки \**з*–фонемаһынан икенсә төрлө күсеш арткаһында барлыкка килгәнме, әйтеүә ауыр. Һәр хәлдә, Әй һөйләше өсөн «ж» өнө «й» өнөнә карағанда боронғорак.

### Лексик үзенсәлектәр

Әй һөйләше көнсығыш диалектының башка һөйләштәрәнән һәм эзәби телдән лексик яктан да айырыла. Уларзы түбәндәге төркөмдәргә бүлөп карарға мөмкин:

1. Туғанлык терминдары: *инәй* ‘эсәй’, *атай* /*әтей* ‘атай’, *кыззар* ‘балдыззар’, *карттай* ‘ола-тай’, *кәртнәй* / *кәтәй* ‘өләсәй’, *тәнәй* ‘кәйнеш’, *парут* ‘ырыу, ара’, *апсын* / *афын* ‘килендәш’, *иркә* ‘кесе туғандарға карата’, *эне* ‘кусты’, *хәрәндәш* ‘сестра’, *бабай* / *быуай* ‘ата-эсәйзән оло ир кеше’.

2. Үсөмлөк атамалары: *бәрәңге* ‘картуф’, *аштык* ‘ашлык’, *бозай* ‘бойзай’, *тилегән орзок* ‘тилебәрән’, *зәрзә* ‘еңле балтырған’, *кымыззык* ‘кымызлык’, *окһон* ‘оқсон’, *тарма* ‘киндер’, *жим* ‘ем’,

3. Азык-түлек: *асылык* ‘ойотко’, *бауырсақ* ‘бауырһақ’, *шәңкә* ‘бәрәмәс’, *үрә* ‘өйрә’, *корт* ‘корот’, *әпәй* ‘икмәк’, *күмәс* ‘мейестә бешкән сөсө икмәк’, *әйрән* ‘эрәмсек һыуы’, *писук* ‘шәкәр’, *жәймә* ‘йәймә’, *йыуаса* ‘майза бешкән камыр азык’.

4. Һауыт-һаба, өй-кәрәк ярағы, мейестә өлөштәрә: *балғалак*, *аштыһыу* ‘помой’, *бизрә* ‘бизрә, күнәк’, *жәйкә* ‘йәйпәк, сынаяк асты’, *уктау* ‘уклау’, *ижау* ‘аш сүмесе’, *һүмес* ‘сүмес’, *кыткыс* ‘кысқыс’, *тәгәс* ‘сынаяк’, *һозгак* ‘һоқко’, *люшкә* ‘йүшкә’, *өрзөктәмә* ‘өрлөк шаршауы’.

5. Кейем-һалым, бизәнәү әйберзәрә: *балдак* ‘йөзөк’, *балитәк* ‘катмар’, *бәшәй* ‘йөнойок’, *һырга*, *ефәк* ‘ебәк’, *күлмәк* ‘күлдәк’, *түшәлдәрек* ‘катын-кыззың кейем эсенән күкрәккә кейелә торған әйберә’, *яка* /*яга* ‘кейем яғаһы’, *нағыш* ‘бизәк’.

6. Йола, ғөрөф-ғәзәт, мифологияға бәйлә һүззәр, һүзбәйләнештәр: *бирнә* ‘кыззың артынан килгән мал-мөлкәт’, *буа* ‘сөннәт’, *буага тотоу* ‘сөннәтләү’, *димдәу* ‘димләү’, *биҫер талау* ‘бүҫерзе имләү йолаһы’, *нәстә кағылыу* ‘ен-зәхмәт кағылыу’, *морондок инәй, морондок атай* ‘килен кешегә егет яғынан билдәләнгән әсәй-атай; әзәби телдә: *киәмәтлек әсәй, киәмәтлек атай*’, *ел тейеу (кағылыу)* ‘ен тейеу’, *ығырмау* ‘үнербау’.

7. Кош-корт, йәнлек, бөжәк атамалары: *бәнкә* ‘тауык, каз, өйрәк себешә’, *дараж* ‘тараш’, *күгәүен/кегәүен, осйөзөк* ‘камка’, *үрзәк* ‘өйрәк’, *йәндәк* ‘кыр хайуандары’.

8. Һөйләшкә хас сифаттар: *асан* ‘ябык’, *дәу* ‘зур’, *тәмде* ‘тәмле’, *көстө* ‘көслө’.

9. Кылымдар: *кереу* ‘инеу’, *төпсәшеу* ‘төпсөнөу’, *йылау* ‘илау’, *ыҫтау* ‘ыҫлау’, *танһыныу* ‘һокланыу’, *бакырыу* ‘кыскырыу’, *ашактау* ‘ашаклау’, *үрәтеу* ‘өйрәтеу’, *жыйыштырыу* ‘1) ерләу, күмеу; 2) йыйыштырыу’.

### Морфологик үзенсәлектәр

1. Был Һөйләштә башкорт теленәң көнсығыш диалектына һәм әзәби телгә хас булғанса, күплек ялғаулары дүрт фонетик вариантта кулланыла (тартынкылар ассимиляцияһына ярашлы): *-тар/-тәр, -дар/-дәр, -зар/-зәр, -лар/-ләр*: *аттар, урмандар, кыззар, балалар* һ.б.

2. Әй Һөйләшендә исем яһаусы ялғаулар әзәби телдәгә формаларзан айырыла: а) *-тык/-тек, -дык/-дек, -зык/-зек*: *утындык, тоззөк, батырзык*; б) *-как/-кәк*: *һуғышкак, боскак* ‘малдың аяк тиреһе’, *корошкак, сәскәк* ‘каюу’; в) *-кыс/-кес*: *алдаткыс* ‘алдаксы’, *уаткыс* ‘ватыусы’; г) *-мак/-мәк*: *баҫмак* ‘баҫкыс’, *каҫмак* ‘йәш баланың башындағы кауак’, *кылтамак*; д) *-сак/-сәк*: *муйынсак, жыйынсак, бауырсак* һ.б.

3. Һөләштә сифат ялғаулары көнсығыш диалектындағы кеүек үк дүрт вариантта кулланыла *-ты/-те, -ды/-де, -зы/-зе, -лы/-ле*: *атты* – әз. *атлы, тәмде* – әз. *тәмле, кыззы* – әз. *кызлы, балды* – әз. *баллы* һ.б.

4. Сифат яһаусы ялғаулар был Һөйләштә күп төрлө: а) *-как/-кәк*: *кыҫкак, йокак, карышкак*; б) *-кас/-кәс*: *атакас* ‘атаһына окшаған’, *инәкәс, ярзыкәс* ‘ярлы’; в) *-кай/-кәй*: *кылтыкай* ‘кылтым’, *остокай, ямакай*; г) *-сан/-сән, -сак/-сәк/-сәк*: *каксан* ‘какса’, *кыйылмасан* ‘кыйыуһыз’, *күңелсәк* һ.б.

5. Кылым яһаусы ялғаулар көнсығыш диалекттағыса дүрт вариантта йөрөй: *-та/-тә, -да/-дә, -за/-зә, -ла/-лә*: *эштәу* – әз. *эшләу, һөйзәшеу* – әз. *һөйләшеу, һуңдау* – әз. *һуңлау, йоктау* – әз. *йоклау* һ.б.

6. Кылымдың 2-се зат хәзәрәгә заман формаһы кыска вариантта кулланыла: *алаң* – әз. *алаһың, алағыз* – әз. *алаһығыз, көтәң* – әз. *көтәһең, көтәгез* – әз. *көтәһегез*. Ә 1-се зат берлектәгә форма *-ым, -ырым* ялғауы аша белдерелә: *алырым* – әз. *алырым; барым, барырым* – әз. *барырым*.

7. Әй Һөйләшендә бойорок һөйкәләшенәң боронғо кыска варианты кулланыла *-ың/-ең*: *килең* – әз. *килегез, кайтың* – әз. *кайтығыз, барың* – әз. *барығыз*. Был форма көнсығыш диалект өсөн хас күренеш.

8. Теләк һөйкәләшә *-мак/-мәк, -мака/-мәкә, -максы/-мәксе* ялғаулары һәм “бул” ярзамсы кылымы менән яһала: *бармак бугайны, килмәк бугайным, күрмәкә бугайным* һ.б.

Шулай итеп, Салауат районы башкорттарының теле көнсығыш диалекттың Әй Һөйләшенә карай. Үрзә карап үтелгәнсә, был Һөйләштә дөйөм көнсығыш диалектының һызаттары күзәтелә, әммә унан айырмалы специфик үзенсәлектәрә лә бар. Шулар рәтенә һүз башындағы [ж] фонемаһы инә. Әй Һөйләшендә был кыпсак осоронан калған боронғо фонетик күренеш һәм сонор [й] тартынкыһына карағанда беренсел һанала.

### Әзәбиәт

Баишев Т.Г. Башкирские диалекты в их отношении к литературному языку. 2-е изд. – Уфа: Гилем, 2006. – 132 с.

Башкорт теленәң диалекттары һүзлегә / Төзөүселәрә: М.И. Дилмөхәмәтов, У.Ф. Нәзәрғолов, С.Ф. Сабирийәнова, Ф.Ф. Гәрәева. – Өфө: Китап, 2002. – 432 б.

Кейекбаев Ж.Ф. Башкорт диалекттары һәм уларзың тарихына кыскаса инеш // Филми язмалар. Башкорт филологияһы серияһы / Башкорт дәүләт педагогия университеты, III сығарылыш, № 2. – Өфө, 1958. – 27–81 б.

Максютова Н.Х. Восточный диалект башкирского языка. – М., 1976. – 290 с.

СИГТЯ = Сравнительно-историческая грамматика тюркских языков: Пратюркский язык-основа. Картина мира пратюркского этноса по данным языка / Отв. ред. Тенишев Э. Р., Дыбо А. В. – М., 2006. – 767 с.

## ЛЕКСИКО-СЕМАНТИЧЕСКИЕ КЛАССЫ СУЩЕСТВИТЕЛЬНЫХ И ГЛАГОЛОВ КАЗАХСКОГО ЯЗЫКА

Статья посвящена проблеме семантической классификации лексики существительных и глаголов казахского языка. Рассмотрены существующие в казахском языкознании денотативно-десигнативный, парадигматический и синтагматический принципы при решении поставленных задач. За основу предложенной авторами семантической классификации принят денотативно-десигнативный принцип, дополненный в отдельных случаях двумя другими перечисленными выше принципами. При рассмотрении существительных казахского языка на основе денотативно-десигнативного принципа выделены 11 семантических классов существительных и 10 классов глагольной лексики, которые разложены на лексико-семантические группы, связанные парадигматическими отношениями.

**Ключевые слова:** лексико-семантические группы, денотативно-десигнативный принцип, казахский язык, парадигматические отношения.

В соотношении «язык – реальная действительность» язык служит не только обозначением реальной действительности, но в то же время является носителем обобщенного отражения действительности, органически соединяясь с соответствующими понятиями или отдельными их признаками, лежащими в основе семантических ценностей языковых единиц.

Общеизвестно, что лексические средства, наряду со звуковыми и грамматическими, составляют основу любого языка, в частности, и казахского. Слово, называя предметы, явления реального мира в их обобщенном виде, выступает в языке основным средством номинации. Являясь структурной ячейкой строя языка, оно имеет определенную морфологическую оформленность и грамматическую значимость. Слово охватывает фактически все языковые функции: номинативную (обозначения), сигнификативную (обобщения), коммуникативную (общения) и прагматическую (экспрессивного выражения и эмоционального воздействия).

Словарные знаки слов таких классов, как существительные, глаголы и прилагательные, обладают двумя основными семиологическими функциями – наименования и описания. Совместимость их в пределах слова порождает назывные высказывания, предикативные и лексические синтагмы. Они являются тем основным звеном в словарном составе языка, той постоянной значимой основой, которая обеспечивает «переход» от языка к мысли, от системы к речи, от слов к синтагматическим единицам большим, чем слова.

При семантической классификации лексики казахского языка имеются свои трудности и своя специфика. Это одна из труднейших задач семантики, которая не находит единого решения и в других языках мира. Количество таких классов и подклассов, их состав зависят от целей исследования, лингвистической школы и целого ряда других факторов.

Работ, посвященных исследованию казахского языка в области семантики, относительно мало. Здесь стоит назвать имя М.Оразова, его монографию «Қазақ тілінің семантикасы» и его исследования по семантике казахского глагола [Оразов 1983; Оразов 1991], которые затрагивают многие вопросы семантики казахского языка. Имеется и ряд других работ по некоторым проблемным вопросам семантики казахского языка [Ахметжанова 1989]. Хотелось только отметить, что результаты этих работ во многом помогли нам при решении поставленных перед нами задач по вопросам формализации содержания казахского текста.

В самом общем виде можно выделить три принципа семантической классификации лексики казахского языка:

- 1) денотативно-десигнативный;
- 2) парадигматический;
- 3) синтагматический [Оразов 1991; Васильев 1981; Зубов 1985].

Денотативно-десигнативный принцип учитывает онтологическое расчленение действительности. При этом слова объединяются в группы в соответствии с тем, что они обозначают в реальной действительности (названия животных, названия растений, наименования определенных классов предметов и т.п.). При парадигматической классификации слова объединяются в группы на основе тождества определенных семантических компонентов, выделенных в значении слов (модальные и ка-

узативные глаголы и т.п.). По синтагматическому принципу слова объединяются в лексико-семантические группы с учетом количества и качества валентностей классифицируемых слов [Оразов 1991, 189-192; Васильев 1981, 39-42].

Исходя из целей исследования за основу нашей семантической классификации принят денотативно-десигнативный принцип, хотя в некоторых случаях он будет дополняться двумя другими перечисленными выше принципами.

**Семантическая классификация имен существительных (зат есім) казахского языка.** Как известно, основным критерием разграничения имен существительных (наименований явлений, предметов, процессов, понятий) казахского языка, как и в других языках, на отдельные типы является прежде всего анализ их прямого номинативного значения, детерминированного типом денотата и сигнификата. При этом «слово» можно рассматривать двояко, т.е. как единицу первичного означивания в системе номинативных средств языка и как единицу, семантически реализованную в речи. Нами различаются две модификации «денотатов»: денотат как объект обозначения единичного предмета или класса предметов, а также денотат как предмет референции знака в речи.

А.А.Уфимцева в своей книге «Типы словесных знаков» [Уфимцева 1974, 102] подразделяет сигнификат как компонент знакового значения слов на шесть понятий, которые, как нам кажется, справедливы и для казахского языка. Имена с денотативным и денотативно-сигнификативным значением относятся к разряду имен, называемых конкретной лексикой, а слова с сигнификативной и сигнификативно-денотативной основой значения составляют абстрактную лексику любого языка, в том числе и казахского. Необходимо заметить, что определяющим для семантики именных лексем является обязательное наличие в их знаковом значении денотата и сигнификата. Если словесный знак служит специализированным названием (термином) в какой-либо узкой предметной области, то доминирует денотат над сигнификатом. В случае, когда словесный знак служит названием абстрактного понятия, то, наоборот, или сигнификат доминирует над денотатом, или словесный знак имеет чисто сигнификативное значение.

Именами существительными в казахском языке, так же как и во многих тюркских языках, являются различные наименования предметов и понятий в широком смысле слова. Это различные имена: собственные, например имена людей, географические названия и т.д.; нарицательные – названия человека: *кісі (человек), шеше (мать), әйел (женщина)*; названия животных; названия растений; названия различных веществ и предметов быта; научные, технические и другие термины; наименования разных обобщенных отвлеченных понятий: *ел (народ), ру (род), тағдыр (судьба)* и т.д.; наименования действий, явлений, состояний: *жүріс (ходьба), хабар (известие), ұйқы (сон)* и т.д.; названия результатов действия: *жиналыс (собрание; ср. жинау – собираться), кеспе (лапша; ср. кесу – резать)*; названия орудий (инструментов) действия: *бұрғы (сверло; ср. бұрау – винтить, завинчивать), сыпырғыш (метла, веник), ара (пила)* и т.п.

Как пишет М.Оразов, среди слов казахских существительных самыми частыми по сложности их значений являются нарицательные имена. Эти имена существительные по отношению к значению имеют в языке изменчивый характер, т.е. иногда у них превалирует денотативное значение, иногда сигнификативное или же приобретает как бы суммарное значение – и денотативное, и сигнификативное. Примером может служить использование слов в метафорическом и метонимическом значениях. Безусловно, когда слова употребляются в переносном значении, то изменяется и их денотативное и сигнификативное значение. Для убедительности сказанного можно сравнить слово «нар» в двух следующих предложениях: *Қырық балам бар еді, қырқы да бірдей нар еді.* (Батырлар жыры). *Айырдан туган жампоз бар – Нарға жүгін салғызбас* (Махамбет) [Оразов 1991, 98]. В первом предложении слово «нар» применено в переносном значении как сильный, «боевой», а во втором – в прямом значении как однопольный верблюд. Их денотативные и сигнификативные значения несравнимы.

В количественном отношении тематические группы имен существительных в любом языке, в том числе и в казахском языке довольно велики. М.Оразов для примера приводит 8 разновидностей таких групп:

- 1) родственные термины (туыскандық, атаулар);
- 2) человек и названия частей его тела (дене мүше атаулары);
- 3) названия домашних животных (төрт түлік мал атаулары); 4) названия одежды, обуви человека (киім-кешек атаулары);
- 5) название птиц (құс атаулары);
- 6) название диких животных (жабайы аң атаулары);
- 7) название обиходных предметов-сосудов (посуды) (ыдыс-аяқ атаулары);
- 8) названия болезни человека (адамға тән ауру атаулары).

В нашем исследовании в основу семантической классификации казахских существительных положено деление реальной действительности на определенные зоны, сегменты [Зубов 1985 и др.]. С учетом целей нашего исследования, анализируемого материала, а также отдельных идей, изложенных в работах [Оразов 1991; Васильев 1981; Апресян 1967] нами выделены семантические классы существительных (см. *табл. 1*). Как видно из таблицы, денотативно-десигнативный принцип использован здесь при выделении наиболее крупных семантических классов существительных. Более мелкие лексико-семантические группы, например «части тела человека», «воинские звания», «дни недели» и др., связаны парадигматическими отношениями.

**Семантическая классификация глаголов. Знаковое значение глагольных лексем.** Имена существительные отражают и обозначают предметы или явления, а глаголы – отношения последних. Значение имен существительных – более автономно и абсолютно. Значение глагольных лексем – относительно и детерминировано многими факторами. В глагольных лексемах, выражающих понятие отношения, их взаимодействие как бы выносится за рамки глагольной лексемы и переносится в область синтагматических отношений действия к его объекту или субъекту (или одновременно). Поэтому соотносительность глагольных лексем с предметным рядом предполагает наличие смысловых связей глагола с предметными именами.

Таблица 1.

**Семантические классы существительных,  
использованных в модели порождения казахского текста**

Сегменты реальной действительности	Наименование семантического класса	Код класса
1. Наука	Общефилософские понятия	ЗТ01
	Общенаучные понятия	ЗТ02
	Общемедицинские понятия	ЗТ03
	Основные понятия психологии	ЗТ04
	Основные понятия нейропсихологии	ЗТ05
	Основные понятия логики	ЗТ06
	Основные понятия геометрии	ЗТ07
	Основные географические понятия	ЗТ08
	Основные понятия языкознания	ЗТ09
	Основные понятия литературоведения	ЗТ10
	Общемедицинские понятия	ЗТ11
	Основные понятия музыкального искусства	ЗТ12
	Основные юридические и уголовные понятия	ЗТ13
	Основные понятия логики	ЗТ14
	Основные понятия геометрии	ЗТ15
	Основные географические понятия	ЗТ16
	Основные понятия языкознания	ЗТ17
	Основные понятия литературоведения	ЗТ18
	Основные понятия изобразительного искусства	ЗТ19
	Основные понятия музыкального искусства	ЗТ20
	Основные юридические и уголовные понятия	ЗТ21
	Понятия, связанные с авиацией	ЗТ22
	Понятия, связан-е с устройством отдельн. Механ-го объекта	ЗТ23
	Информационные понятия	ЗТ24
2. Природа	Оценка	ЗТ31
	Предметы природы	ЗТ31
	Явления природы	ЗТ32
	Предметы природы – артефакты	ЗТ33
	Состояния природы	ЗТ34
	Понятия, описыв-ие структуру предм-в и части предметов	ЗТ35
	Вещества, металлы	ЗТ38
Действия предметов природы	ЗТ39	



3. Человек и его мир	Человек и его части тела	ЗТ40
	Части лица и головы человека	ЗТ41
	Состояния и ощущения, специфичные для человека	ЗТ42
	Действия, специфичные для человека	ЗТ43
	Понятия, используемые при характеристике человека	ЗТ44
	Предм-ты, одежды и обуви чел-ка и понятия, связан. с этим	ЗТ45
	Предметы-артефакты, используемые человеком в целях гигиены, для украшения и т.п.	ЗТ46
	Объекты, используемые для передвижения человека и понятия, с этим связанные	ЗТ47
	Предметы питания человека	ЗТ48
	Понятия, связанные с жилищем человека	ЗТ49
	Понятия, связанные со смертью и захоронением	ЗТ50
	Имена, отчества и фамилии людей	ЗТ51
	Национальность человека	ЗТ52
	Предметы-сосуды, которыми пользуется человек, и связанные с этим понятия	ЗТ53
	Предметы, используемые на отдыхе, охоте, рыбалке, и связанные с этим понятия	ЗТ54
	Наименование гражданских специальностей, должностей, степеней, званий	ЗТ55
	Предметы-артефакты, используемые при учении, труде и т.п.	ЗТ56
	Понятия, связанные с взаимоотношением людей	ЗТ57
	Понятия, связан. Родствен-ми отношениями между людьми	ЗТ58
4. Животный мир	Награды человека	ЗТ59
	Крупные домашние животные и части их тела	ЗТ60
	Насекомые и части их тела	ЗТ61
	Птицы и части их тела	ЗТ63
	Действия, специфичные для животных	ЗТ64
	Предметы питания животных, птиц и насекомых	ЗТ65
5. Растения	Растения и их части	ЗТ66
	Деревья и их части	ЗТ67
6. Война и понятия, с ней связанные	Предметы боевой техники, оружия и их части	ЗТ68
	Наименование военных специальностей	ЗТ69
	Воинские звания	ЗТ70
	Понятия, связанные с оказанием помощи во время войны	ЗТ71
	Названия воинских подразделений	ЗТ72
	Понятия, связанные с войной	ЗТ73
7. Время	Временные понятия	ЗТ74
	Дни недели	ЗТ75
8. Религия	Понятия, связанные с обучением и трудом	ЗТ76
9. Учеба и труд	Организация учебы, труда	ЗТ77
	Понятия, связанные с обучением и трудом	ЗТ78
10. Меры	Меры длины, веса, скорости и т.п.	ЗТ79
11. География	Имена населенных пунктов	ЗТ80
	Имена стран, материка, частей света, островов, гор и др.	ЗТ81
	Имена рек, морей, озер, проливов	ЗТ82

Знаковое значение глагольных лексем представляет собой сложный феномен, детерминируемый тремя факторами: 1) соотносительностью с реальной действительностью, которая представляет собой в данном случае не мир предметов, а мир их отношений, действий и состояний; 2) категориальной семантикой сочетающихся с глаголом предметных имен; 3) типом смысловых отношений между действием, его субъектом и объектом, а соответственно типом моделей «объект – действие», «действие – объект», «субъект – действие – объект» [Уфимцева 1974, 113].

По сравнению с именами существительными класс глагольных лексем представляет большие трудности при выделении типов по характеру их знакового значения.

Во-первых, собственно лексическое (индивидуальное) значение глагольных слов осложнено целым рядом семантических признаков, выражающих способ протекания действия, обстоятельства

его свершения, видовременные характеристики глагольного действия, модальные и экспрессивные наслонения.

Во-вторых, определение глагольной семантики затруднено уточнением, какой из признаков присущ только данной лексеме, какой – целой группировке, а какой обобщен до уровня категории глагольных лексем и т.п.

Традиционно глаголы по их семантике делятся на три класса:

- 1) глаголы действия;
- 2) глаголы состояния;
- 3) глаголы процессов.

У каждой лексико-семантической группы – свои парадигматические параметры, семантические признаки. Так, для глаголов «движения», довольно хорошо изученных во многих языках, выделяются следующие семантические (лексические) категории: направленные движения, скорость движения, манера передвижения, средство или среда передвижения и т.п.

Для семантики эмотивных глаголов существует иная шкала – глаголы воздействия, глаголы восприятия, глаголы оценки или субъективного отношения и т.п. [Уфимцева 1974, 148].

М. Оразов в своей книге «Қазақ тілінің семантикасы» пишет, что неконкретность лексического значения глагола в основном зависит от его денотативного и сигнификативного значения. Когда речь идет о денотативном значении глагола, мы говорим о движении предмета, физическом действии (амал-эрекет) и о состоянии (қалпы). Если учесть, что движение происходит в какое-то определенное время и в определенном месте, то имеется в виду присутствие субъекта, который осуществляет физическое действие, известна также цель передвижения. При этом лексическое значение глагола дополняют множество других элементов. В качестве примера он приводит следующее предложение: *Шаңғышылар селоза кіргенде көз байланып та қалған еді. (М.Сундетов). /Когда лыжники вошли в село, уже наступило сумерки/*. В этом предложении рассматривается структура значения глагола *кір* (входи):

- 1) наличие движения;
- 2) движется некоторый предмет или человек (субъект);
- 3) движение происходит на земле;
- 4) переход через некоторую границу (вошли на территорию села);
- 5) имеется элемент значения войти «внутри». Перечисленные значения определяются благодаря отношениям и связям слов в данном предложении.

Таким образом, действие *кіру* (входить) происходит на лыжне (на снегу) движением лыжников, *кіру* означает вхождение на территорию, где проживают люди (в село). Благодаря сочетанию слова *кір* с другими словами (*село, шаңғышылар*), мы узнаем направление движения, что оно происходит «извне в село».

Глагол *кір*, сочетаясь с другими словами, может принять абсолютно другое значение. Например, *Мен 1962 жылы партия ұйымына кірдім*. В этом предложении слово *кір* уже не означает движения, а дает понять, что он стал членом партийной организации. Последний пример позволяет говорить о том, что значение слова непосредственно не зависит от значения словосочетаний (от контекста, в широком смысле) в предложении. В первом примере глагол *кір* сочетается с объектом, означающим местность, поэтому его граница является основой понимания лексического значения из сем движения. Во втором примере глагол *кір* сочетается с понятием, т.е. названием организации, здесь сема глагола *кір* единственная – стать членом данной организации. Изменение значения некоторых глаголов требует очень тонкого подхода при анализе их семантики. Например, глаголы движения в одних сочетаниях входят в структуру глагола общения, а в других – в структуру глагола мышления, Сравните: *үйге келу* (прийти домой), *қарсы келу* (возразить), *ойға келу* (вспомнить), *тілге келу* (заговорить) и др. В структуре предложения глаголы образуют отношения с определенными словами. Такие отношения происходят между субъектом или объектом и движением, между физическим действием (*амал-эрекет*) и объектом или субъектом.

**Тематические и лексико-семантические классы казахских глаголов.** Основанием для отнесения слов в различные тематические и лексико-семантические классы являются обобщающие (жалпылаушы) семы. Необходимо отметить, что в индо-европейских языках имеется определенный опыт в разделении на лексико-семантические классы отдельных частей речи, но в отношении тюркских языков, в частности казахского, подобные исследования начали появляться только в последнее 10-летие (большинство из них имеют сопоставительный с русским языком характер). В настоящее время описаны следующие лексико-семантические группы глаголов казахского языка: глаголов физиологического воздействия [Қасабекова 1985], глаголов речи [Курамысова 1986], глаголов ориентированного положения в пространстве [Дүйсекова 1996], глаголов движения [Шайкенова], эмотивные глаголы [Ибадильдина 1997], глаголы физического восприятия [Уразымбетова 1998], глаголы созидания [Токтарова] и др.

В своей семантической классификации знаменательных классов слов казахского языка мы опирались на результаты исследований М.Оразова и работы вышеупомянутых авторов.

В центре лингвистической проблематики находится изучение фундаментальных семантических категорий со сложной формально-содержательной структурой. К числу таких категорий, отражающих одну из форм всеобщей связи явлений и фактов объективной действительности, следует отнести категорию каузальности (причинности).

В исследованиях казахского языка категория каузальности не была объектом специального изучения. Каузальные отношения в языке рассматривались в пределах отдельно взятых лингвистических задач. Например, в работах А.К.Калыбаевой-Хасеновой, С.В.Шкурдина и др. на морфемном уровне рассматривались глаголы в форме понудительного залога, образующие морфологическую каузативность. На синтаксическом уровне изучались сложноподчиненные и бессоюзные сложные предложения с причинно-следственным характером отношений между составляющими их частями (М.Б.Балакаев, Т.Кордабаев, Н.Х.Демесинова, К.Есенов, Т.Ж.Тажибаева, У.Салиева и др.).

В последнее 10-летие научная школа, возглавляемая доктором филологических наук профессором З.К.Ахметжановой, намного продвинула вперед функционально-семантическое направление исследования в сопоставительном плане изучения казахского и русского языков. При семантической классификации каузативных глаголов казахского языка, мы в своей работе в основном придерживались положений работы Г.М.Алиной «Каузальные отношения в системе русского и казахского языков» (рукопись канд. дис., руков. З.К.Ахметжанова).

Необходимо отметить, что бытийные глаголы казахского языка еще не освоены до конца. В существующих классификациях (И.Маманова, А.Ыскакова, Н.Оралбаевой, М.Оразова, Н.Душимовой и др.) лексико-семантических групп казахского языка бытийные глаголы не выделяются как отдельная группа. Глагол «быть» является основным средством выражения понятия бытия в русском языке. В казахском языке, глаголы с чисто бытийным значением, такие, как *еді (екен, емес), бол*, характеризуются в основном как вспомогательные, которые самостоятельно не употребляются, но образуют описательные формы времен и наклонений основных глаголов.

А. Ыскаков, в свою очередь, определяет глаголы, образованные от корня *-е (еді, екен, емес)*, как неполные слова, которые утратили свое лексическое значение либо полностью, либо наполовину и играют в предложении вспомогательную роль, придавая грамматическое значение отдельным словам (*бала еді, қырық-отыздар шамасында екен, қара-торы ала емес*).

Учитывая все высказывания исследователей о семантических классах казахских глаголов, мы для наглядности свели их в одну таблицу (*табл. 2*). По нашей классификации семантический класс казахских глаголов делится на 10 разновидностей: I. Глаголы существования; II. Глаголы физических действий; III. Глаголы движения; IV. Глаголы, означающие отношения между людьми, животными и неживыми объектами; V. Глаголы психической деятельности; VI. Глаголы физического состояния и нормы; VII. Глаголы психического состояния; VIII. Фазисные глаголы; IX. Модальные глаголы; X. Каузативные глаголы.

Каждый класс глагола может быть разделен еще и на семантические подклассы. В данной статье мы решили показать лишь фрагмент, охватив некоторые семантические классы и подклассы казахских глаголов. В таблице показана следующая информация: 1. В отдельной строке под римской нумерацией дается семантический класс глаголов; 2. В первом столбце таблицы указаны семантические подклассы каждого класса; 3. Во втором столбце – соответствующие примеры к каждому подклассу казахских глаголов; 4. В третьем столбце даны соответствующие коды каждого класса и подкласса, которые будут использованы при записи формул предложений, абзацев и текста.

Таблица 2.

**Семантические классы казахских глаголов,  
используемых в модели порождения текста (фрагмент)**

Семантический подкласс глагола	Примеры	Код
1	2	3
<b>Семантический класс: I. Глаголы существования</b>		
I. Бытийные глаголы:	<i>Менде жақсы оқуға мүмкіншілік бар,</i>	ET101
<i>Бар</i>	<i>бірақ оған тұқ ынтам жоқ.</i>	
<i>Жоқ</i>	<i>Оның жасы әлі кәмелетке толған жоқ.</i>	ET102

Бол (болу, болмау)	<i>Сөзіңді айтып бол. Кеше жиналыс болды</i>	ЕТ103
Е (еді, екен, емес)	<i>Алматы Қазақстанның астанасы еді. Өзі бір қызық адам екен. Оның аты қара-торы емес.</i>	ЕТ104
<b>Семантический класс: II. Глаголы физических действий</b>		
1. Глаголы физического изменения: созидания и производства, соединения и деления и т.д.	<i>Жасау, соғу, тігу, тұрғызу, салу, бөлшектеу, екішеу, ыдырату и т.д.</i>	ЕТ201
2. Глаголы активных физических действий: разрушения и порчи	<i>Қирату, бүлдіру, бұзу, сындыру, құлату и т.д.</i>	ЕТ202
3. Глаголы эмоциональных физических действий	<i>Аузы ашылу, мәз болу, күлу, ыржиу, шапалақтау, секіру (қуанғаннан)</i>	ЕТ203
4. Глаголы употребляющиеся при приеме пищи, свойственных для человека и животного мира	<i>Ішу, жеу, ұрттау, жұту, асау, дәм тату, сорптау и т.д.</i>	ЕТ204
5. Глаголы прочих физических действий, свойственных человеку и животному миру	<i>Жүгіру, секіру, ырғу, құсу, таптау, жылжу, қыймылдау ....</i>	ЕТ205
6. Глаголы физических действий, описывающих неживую природу	<i>Күркіреу, сарқырау, арсылдау, сілкіну и т.д.</i>	ЕТ206
7. Глаголы прочих физических действий, свойственных неживой природе	<i>Ағу, бату (күн), құлау, қайнау, дірілдеу, ысу, суу, гүрсілдеу и т.д.</i>	ЕТ207
8. Глаголы, связанные со звуками окружающей среды, глаголы звуковыражения	<i>Тарсылдау, сатырлау, сарқырау, ызыңдау, гүрсілдеу, қоңқылдау и т.д.</i>	ЕТ208
9. Глаголы физических действий, свойственных только животному миру	<i>Ұлу, ұру, ырылдау, мөңку, мөңіреу, жорту, шабу и т.д.</i>	ЕТ209
<b>Семантический класс: III. Глаголы движения</b>		
1. Глаголы, означающие движения, происходящие на земле	<i>Жүру, жүгіру, еңбектеу, домалау, сырғанау и т.д.</i>	ЕТ301
2. Глаголы движения, происходящие на воде (в воде)	<i>Жүзу, сүңгу, малту, қалқу и т.д.</i>	ЕТ302
3. Глаголы движения, происходящие в воздухе	<i>Ұшу, самғау, қалықтау, шүйлігу, зымырау и т.д.</i>	ЕТ303
4. Глаголы движения, означающие направленное движение	<i>Өрлеу, ылдилау, келу, кету, жақындау, ұзау</i>	ЕТ304
5. Глаголы движения, означающие хаотичность движения	<i>Аралау, шарлау, кезу, қыдыру, атқу и т.д.</i>	ЕТ305
<b>Семантический класс: IV. Глаголы, означающие отношения между людьми, животными и неживыми</b>		
1. Глаголы выравнивания	<i>Алу, қосу, теңестіру, салыстыру, бөлу и т.д.</i>	ЕТ401
2. Глаголы принадлежности	<i>Иемдену, меншіктеу, иелену, баурау и т.д.</i>	ЕТ402
3. Глаголы приветствия и прощания	<i>Амандасу, сәлемдесу, көрісу, қоштасу и т.д.</i>	ЕТ403
4. Глаголы почтения	<i>Құрметтеу, қадірлеу, ізеттеу, сыйлау и т.д.</i>	ЕТ404
5. Глаголы властвования	<i>Басқару, әмір ету, боп-салау, өкім ету и т.д.</i>	ЕТ405
6. Глаголы подчинения	<i>Көну, бас иіу, ырқына көну и т.д.</i>	ЕТ406
<b>Семантический класс: V. Глаголы психической деятельности</b>		
1. Глаголы ощущения: а) глаголы ощущения в общем смысле	<i>Сезу, сезіну, байқау, абайлау, аңғару, пайымдау, түйсіну ...</i>	ЕТ501
б) глаголы видения	<i>Көру, көз салу, көз татау, үңілу, бажырайу и т.д.</i>	ЕТ502
в) глаголы вкусовые	<i>Кермектену, ашқылтымдану, тату</i>	ЕТ503

г) глаголы осязания (связанные с запахом)	<i>Бықсу, мұңку, аңқу, сасу, борсу, қаңсу</i> и т.д.	ET504
д) глаголы осязания, связанные с порами тела человека и животных	<i>Сыздау, ашу, қышу, тызылдау, сырқырау</i> и т.д.	ET505
е) глаголы, связанные со слухом	<i>Тыңдау, есіту, құлақ салу, дауыстау, естілу</i> и т.д.	ET506
<b>2. Глаголы мышления:</b>		
а) глаголы мышления, связанные с процессом мышления вообще	<i>Ойлау, пайымдау, ой жүзгірту, ақылға салу, болжау, шамалау</i> и т.д.	ET507
б) глаголы предположения (приблизительности)	<i>Шамалау, болжау, тұспалдау, межелеу</i> и т.д.	ET508
в) глаголы, связанные с выводами, заключениями, обобщениями	<i>Қорытындылау, шешілу, кесіп айту</i> и т.д.	ET509
г) глаголы доверия	<i>Сену, достасу, дос көру, жақындасу, сенім арту.</i>	ET510
д) глаголы запоминания	<i>Ескер, ұмытпа, есте сақта, зердеде қалдыру ...</i>	ET511
<b>3. Глаголы психического воздействия</b>		
4. Глаголы внимания и желания	<i>Қуанту, күлдіру, қорқыту, ойландыру</i> и т.д. ( <i>жақсы, ақылды, сыпайы, ұстамды</i> ) болу и т.д.	ET512
5 Глаголы активной психической деятельности	<i>Айқайлау, ұрысу, зар жылау, қарқылдау</i> и т.д.	ET514
<b>6. Глаголы эмоциональные:</b>		
а) глаголы восхищения	<i>Таңдану, қуану, таңырқау</i> и т.д.	ET515
б) глаголы радости	<i>Шаттану, қуану, мәз болу ...</i>	ET516
в) глаголы удовлетворения	<i>Рахаттану, қанағаттану, шаттану</i> и т.д.	ET517
г) глаголы эмоционального переживания (скорби)	<i>Қапалану, қайғыру, қабырғасы қайысу</i> и т.д.	ET518
<b>Семантический класс: VI. Глаголы физического состояния и нормы</b>		
1. Специфичных для элементов неживой природы (изменение цвета)	<i>Жылтырау, жарқырау, ағару, қарайу, сарғайу, сәулелену</i> и т.д.	ET601
2. Глаголы физического сохранения	<i>Сақталу, бұзылмау, ашымау...</i>	ET602
а) означающие внутренние психологические состояния человека	<i>Қасарысу, ашулану, қайғыру, толқу, қуану, ойлану, армандау ...</i>	ET603
б) означающие субъективные черты человека (наружные)	<i>Кісімсу, адамсу, баласыну, сасқалақтау</i> и т.д.	ET604
3. Глаголы, передающие изменение объема предмета, вещества	<i>Шөгү, ұзару, солу, толысу, зораю, кеңею</i> и т.д.	ET605
4. Глаголы, передающие изменение предмета в количественном отношении	<i>Азаю, көбею, молаю, ортаю, қосылу</i> и т.д.	ET606
5. Глаголы, означающие изменение внутренних свойств человека или предмета	<i>Іру, шіру, үгілу, тозу, босау, өзгеру</i> и т.д.	ET607
6. Глаголы, употребляющиеся в связи с привычками людей	<i>Әдеттену, үйрену, дағдылану, жалқаулану ...</i>	ET608
7. Глаголы, связанные с процессом рождения и роста, свойственных человеку или животным	<i>Туу, өсу, өну, ержету, жігіт болу</i> и т.д.	ET609
<b>Семантический класс: VII. Глаголы психического состояния</b>		
1. Глаголы, свойственные только человеку	<i>Қайғыру, күңірену, ренжу, күлу, жылау, қуану</i>	ET701
2. Глаголы, связанные с природными явлениями	<i>Азынау, гуілдеу, борандату, желдету</i> и т.д.	ET702
3. Глаголы, свойственные человеку и особям животного мира	<i>Қайғыру, қуану, ашулану, ызалану</i> и т.д.	ET703
4. Глаголы, приписываемые элементам неживой природы	<i>Қайғыру, мұңаю, күңірену, қуану, жадырау...</i>	ET704

5. Глаголы, свойственные человеку, особям животного мира и риписываемые элементам неживой природы	<i>Қайғыру, мұңаю, жадырау, күңірену, қуану, күлімсіреу и т.д.</i>	ЕТ705
<b>Семантический класс: VIII. Фазисные и модальные глаголы</b>		
1. Начало действия	<i>Бастау, кірісу, жасағалы-, кіріскелі-</i>	ЕТ801
2. Продолжения действия	<i>Жалғастыру, созу, атқару, -ып, -іп (отыр)...</i>	ЕТ802
3. Конец действия	<i>Аяқтау, қою, бітіру, тоқтату ...</i>	ЕТ803
и.т.д	и.т.д	и.т.д

В статье мы опускаем подробный табличный вид каузативных глаголов, которые нами заимствованы из рукописи кандидатской диссертации Г.М.Алиной «Каузальные отношения в системе русского и казахского языков» и делятся на следующие классы и подклассы:

I. Собственно-каузативные глаголы:

- 1) понуждение; 2) вызывание, причинение;
- 3) побуждение; 4) обусловливание; 5) воздействие;
- 6) допущение.

II. Частно-каузативные глаголы:

- 1) созидания объектов;
- 2) уничтожения объекта;
- 3) со значением рождения;
- 4) со значением лишения жизни кого-либо;
- 5) со значением изменения объектов;
- 6) со значением каузации контакта предметов;
- 7) со значением каузации функциональных отношений;
- 8) со значением каузации движения;
- 9) со значением каузации посевивных отношений;

### Литература

- Апресян Ю. Д. Экспериментальное исследование семантики русского глагола. – М., 1967. – 252 с.
- Ахметжанова З. К. Функционально-семантические поля русского и казахского языков. – Алма-Ата, 1989. – 106 с.
- Васильев Л. М. Семантика русского глагола. М., 1981. – 184 с.
- Дүйсекова К. К. Семантика глаголов ориентированного положения в пространстве в казахском языке. Автореф дисс. Канд. филол. наук. – Алматы, 1996. – 26 с.
- Зубов А. В. Вероятностно-алгоритмическая модель порождения текста. Автореф дис. докт. филол. наук. – Минск, 1985. – 399с.
- Ибадильдина З. К. Эмотивная лексика казахского и русского языков: Автореф. дис канд филол. наук. – Алматы, 1997. – 26 с.
- Курамысова К. Н. Семантико-грамматические особенности глаголов речи // Актуальные проблемы общественных наук. Агма-Ата. 1986. – Ч. 2. – 206 с.
- Қасабекова Қ. Қ. Қимыл-қозғалыс етістіктердің жасалу ерекшеліктері // Қазақстан мектебі, 1985. – № 4. – 27 б.
- Оразов М. Қазақ тілінің семантикасы. Алматы, 1991. – 213 б.
- Оразов М. Семантика казахского глагола: дис. докт. филол. наук. – Алма-Ата, 1983.
- Токтарова Т. Ж. Семантика глаголов созидания казахского и русского языков: Автореф. дис. канд. филол. наук. – Алматы, 1998. – 25 с.
- Уразымбетова А. Б. Глаголы физического восприятия русского и казахского языков и некоторые аспекты их перевода: Автореф. дис. канд. филол. наук. – Алматы, 1998. С. – 25.
- Уфимцева А. Типы словесных знаков. – М., 1974.
- Шайкенова Ж. А. Семантика глаголов движения в казахском и русском языках: Автореф дисс. канд. филол. наук. – Алматы, 1981. – 23 с.

## О ЗВУКАХ ЗАИМСТВОВАННЫХ БУКВ КАЗАХСКОГО АЛФАВИТА

В статье рассматриваются звуки заимствованных букв из русского языка в казахском алфавите. В частности, указывается звуковой состав букв **й, у, я, ю, щ**, определяется способ расщепления их на составные части. Доказывается, что эти символы являются буквами, относящимися к звуковой системе казахского языка.

**Ключевые слова:** буква, слог, артикуляционная база, перцепция, звук, сингармонизм, перенос, морфем.

А. Қаршығаева  
Алматы қ., Қазақстан

## ҚАЗАҚ ТІЛІНДЕГІ КІРМЕ ӘРІПТЕРДІҢ ДЫБЫС МӘНІ

Мақалада қазақ әліпбиіндегі кірме таңбалардың дыбыс мәні туралы айтылады. Атап айтқанда, и, у, я, ю, щ дыбыстарының дыбыс құрамы, олардың құрамды бөліктерінің жігін ажыратудың тәсілдері көрсетіледі. Бұл таңбалардың қазақ тілінің дыбыс жүйесіне жат әріптер екені дәлелденеді.

**Тірек сөздер:** әріп, буын, жасалым қоры, естілім, дыбыс, үндесім әуез, тасымал, морфем.

Қазіргі қазақ әліпбиінің тізбегінде төл таңбалармен (әріптермен) қатар кірме таңбалар араласып жүр. Орыс тілінен және орыс тілі арқылы басқа тілдерден ауысқан сөздердің айтылуы /орфоэпия/ біркелкі емес, ондай ауысқан сөздердің бірқыдыруы қазақ тілінің дыбысталу жүйесіне қарай айтылса, бірқыдыруы сол ауысқан сөздердің өзіне тән ерекшелігіне қарай айтылады. Сөздердің ауызша айтылуы, негізінде еркін болатын құбылыс.

Төркіні бөлек кірме таңбалардың жасалым-естілімі (артикуляция-перцепциясы) бөгде болғандықтан, олар қазақ сөзінің дыбыс/әріп санын, үндесім әуенін, морфем құрамын, буын тұрқын, тасымал ретін, сондай-ақ, сөйлеу ырғағын түгел бұзылып айтылып жүр.

«Жазу реформасын қолға алғандағы мақсатымыз орыс тілінен енген кірме таңбаларды қазақша таратып жазу болатын. «Я» дегеніңіз қазақтың екі дыбысының тіркесі. Ал енді «ю» дегеніңіз қазақтың үш дыбысының тіркесі. Сонда қазақ тілінің буын құрамы бұзылмайды. Үндесім әуезі, морфем құрамы бұзылмайды. Қазіргі кирилл негізіндегі көп сөздеріміз тілдің үндесім әуезін бұзып отыр.

Профессор Ә.Жүнісбектің айтуынша, тілімізге еніп кеткен орыс тілінің емле-ережесінен арылу қажет.

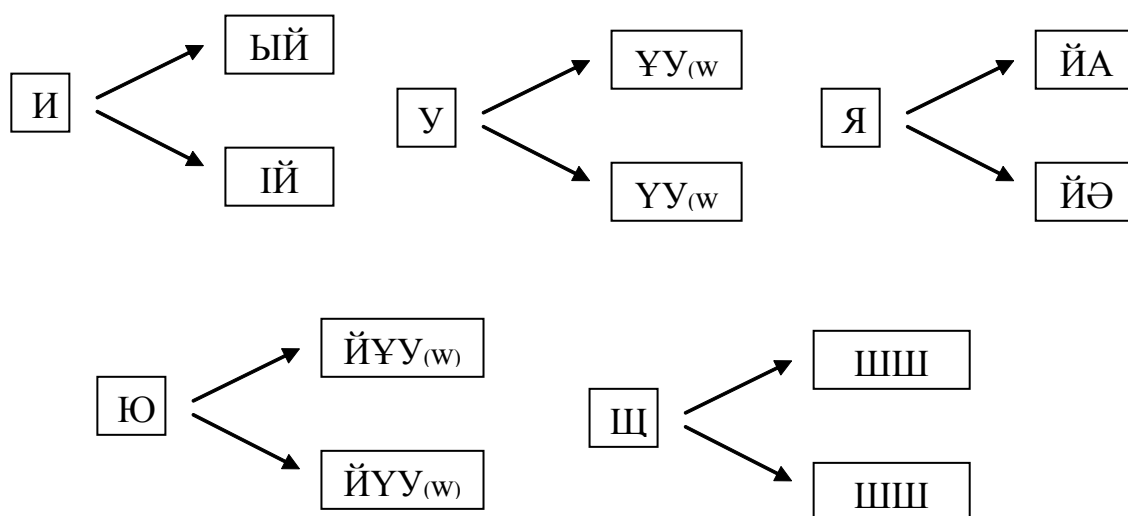
Қазақ тілінің төл сөздерінің жазылымына бірнеше кірме дыбыстар еніп кетті. Оның өзі қазақ тілтануымының теориялық фонетикасына да, практикалық фонетикасына нұқсан келтіріп жатыр. Теориялық тұрғыдан – қазақ тілінде жоқ «фонемалардың» санын көбейтіп жіберді, практикалық тұрғыдан – оқу-әдістеме шараларын қиындатып жіберді. Сонымен кірме әріптердің тіл бұзар әсері күшейіп барады:

- бірінші, қазақ сөзінің үндесім әуезін бұзып барады;
- екінші, қазақ тілінің буын жігін бұзып болды;
- үшінші, қазақ тілінің морфем желісін бұзып болды;
- төртінші, қазақ сөзінің тасымал ретін шатыстырып бітті.

Ол кірме әріптер: И, У, Я, Ю, Щ. Қазақ тілінің төл сөздерінің құрамында жазылып жүрген кірме әріптердің дыбыс мәні төмендегідей (мый, тій, сұу, сұурөт, жая, әлія, айуу, үйуу, ашшы, кешше, тұшшы) [Жүнісбек 2008]:

Мұндағы **у** таңбасының дыбыс мәнін латын әліпбиінің құрамындағы **w** әрпімен белгіленіп отыр. Өйткені қазақ тілінде **ерін-еріндік, жуысыңқы, үнді** дауыссыз бар. Ол дауыссыз халықаралық практикада осы таңбамен беріледі (ондай дыбыс ағылшын тілінде де бар, ағылшын тілінде сауатты өсіп келе жатқан жас ұрпақ оны тез аңғарып, жеңіл қабылдайды). Ал орыс тілінде ондай дауыссыз дыбыс жоқ. Олардың **Кокшетау** дегенді **Кокчетав** деп өз тіліндегі **тіс-еріндік, жуысыңқы, ұяңв** дауыссызымен алмастыратындығы сондықтан. Қазақ тіліндегі дауыссыз **у** мен орыс тіліндегі дауысты **у** дыбыстарын естілім ұқсастығына қарап өзара шатастырылып жүр. Тіптен **й** мен **у** әріптері

(дыбыстары емес), жоғарыда айтқандай, өз алдына «фонема» делінеді. Оны қазақ тілінің дамуы деп түсіндіруге тырысушылық та жоқ емес.



Өкінішке орай, орыс тілінің емле-ережесі арқылы қазақ тілінің төл сөздерінің жазылымына енген **й, у, я, ю** таңбаларының (дыбыстарының емес) әрқайсысын өз алдына дербес дауысты дыбыстар деп қабылдау бар. Өйткені оқу бағдарламалары мен оқулықтар түгелдей осы ережені құптап отыр. Қазақ тілі дыбыстарының құрамына енгізіліп отырған «дауысты дыбыстардың» жөнін табу үшін олар «дауыстыдан кейін дауыссыз болады, дауыссыздан кейін дауысты болады» деген ереже қабылданды.

Кез келген жазу тілдің ішкі өзіне тән құрылымы мен табиғи айтылымына зиян тигізбеу керек. Керісінше, тілдің әдеби айтылымының қорғанышы болу керек. Осы тұста Әлімхан Жүнісбек өз еңбегінде былай талдап көрсетеді.

Ол үшін:

1. Жазу сөздің дыбыс құрамын дәл бейнелегенде ғана дұрыс жазу болады. Ендеше қазақ жазуы қазақ сөзінің үндесім әуезін дұрыс көрсетіп отыру керек, басқаша айтқанда, жуан айтылатын сөздің таңбасы мен жіңішке айтылатын сөздің таңбасының арасында көрнекі айырма болсын.

2. Жазу қазақ сөзінің буын құрамын бұзбау керек. Мысалы, **ми-ы, су-ы**[**мый-ы**], [**сұу-ы**] болмау керек, **мы-йы, сұ-уы** болу керек. Қазақ тілінде сөз ішінде буын дауыстыдан басталмайды. Оқулықтардағы осындай кереғарлықтардың қиындықтарын айтып, балаларға түсіндіре алмай, олардың алдында «бірде олай болады, бірде бұлай болады» деп екі сөйлеп, әлек болып жатқан ұстаз мұғалімдердің пікірі ескерусіз қалып жатыр.

3. Жазу қазақ сөзінің тасымал ережесіне сай болу керек. Тағы да **ми-ы, су-ы** болып тасымалданбау керек, **мы-йы, сұ-уы** болып тасымалдану керек.

4. Жазу қазақ тілінің морфем (түбір мен қосымшалар) желісін үзбеу керек. Мысалы, қазақ тілінің «егер түбір дауыссызға бітсе, қосымша дауыстыдан басталады» деген іргелі заңдылығы бар. Сонда **мың = мы-ңы, мыс = мы-сы, мый = мы-йы** болып, тіл жүйесі бұзылмайды.

Қазақ тіліндегі дифтонгтардың сингармологиялық мәртебесін (статусын) анықтау бүгінгі қазақ фонетика ғылымының өзекті мәселесіне айналып отыр [Жүнісбек 2018]. Мәселенің теориялық шешімі қазақ тілі дыбыстарының құрамын дәл анықтап, «қазақ тілінде қанша дауысты дыбыс бар?» деген сұраққа нақты жауап тауып беретін болады. Ал оның практикалық шешімі қазіргі қазақ жазуындағы қайшылықтардан арылуға мүмкіндік береді.

Жазуда **й, у** әріптерімен таңбаланып жүрген дыбыстарға, бір жағынан, «дауыстыдан кейін дауыссыз, дауыссыздан кейін дауысты болады» деген мектеп және жоғары оқу орындарының оқулықтарында жаңсақ анықтама берілсе, екінші жағынан, олар «дауысты мен дауыссыздардың тіркесін белгілейтін таңба» деген академиялық грамматикадағы ақиқат анықтама беріледі.

Өзара артикуляциялық үйлесімі күшті, естілімі ұқсас болып келетін қысаң дауыстылар мен үнді дауыстылар тіркесінің құрамды бөліктерін ажыратып алу оңай болмайды (оның үстіне орыс тілінің **ы, и** дауыстыларының естілімі мен қазақ тілінің **ый, йй** дыбыс тіркестерінің естілімі өзара ұқсас болуы да өз әсерін тигізді. Кезінде қазақтың төл сөздерінің құрамында **и** таңбасын жазу туралы ереже қабылдағанда, сөз жоқ, осы жай себеп болған болу керек).



Олардың құрамды бөліктерінің жігін ажыратудың бірнеше тәсілі бар.

Солардың бірі – морфемалық талдау: егер өзге морфемалық жігі анық тіркестермен ұқсас келсе, яғни арасынан морфема жігі өтетін болса, онда олар дербес-дербес дыбыстардың тіркесі болып саналады.

Ендеше, бірінші, морфемалық жігі анық тіркестермен салыстырып шығу керек. Төмендегі түбір морфемалар дыбыс құрамының әртүрлілігіне қарамай, бірыңғай тілдік бірліктер болып табылады (көпшілікке түсінікті болу үшін тіркес транскрипциялары да кирилше беріп отыр):

**ай [ай], әй [әй], ей [ей], ой [ой], өй [өй], ұй [ұй], үй [үй], и [ый], и [ий]**

Алғашқы жеті сөздің (тіркестің) дыбыс құрамына ешкім күманданбайды, ал соңғы **и** арқылы жазылып отырған сөзге (тіркеске) келгенде күмән басталады. Шындығында, барлығы ды әртүрлі дауысты дыбыстар мен үнді, жуысыңқы, тіл ортасы (тілшік) **й** дауыссызының тіркесі болып келеді. Егер **ай** немесе **үй** тіркестерінің дыбыс құрамы күман туғызбайтын болса, онда солармен теңдес **й, и** тіркестерінің де дыбыс құрамы күмән туғызбау керек.

Тіптен, басқаны былай қойғанда, **и** таңбасы бір дауысты дыбыс болса, онда **ми** деген сөздің құрамында неге жуан айтылып, **ми** деген сөздің құрамында неге жіңішке айтылып, екі түрлі дыбысталады деп ойлану керек. Өйткені **и** таңбасы бірде құрамында жуан **ы** дауыстысы бар **ий** тіркесін, бірде құрамында жіңішке **і** дауыстысы бар **ий** тіркесін белгілейді.

Енді **у** таңбасымен келетін бірыңғай сөздер тобы алынды (**у** таңбасының мәні түсінікті болу үшін ол арнайы **w** таңбасымен белгіленіп отыр):

**ау [aw], әу [әw], оу [ow], өу [өw], у [ұw], у [үw]**

Алғашқы төрт сөздің (қазақ тілінде **ы, і, е** дауыстылары **у[w]** дауыссызымен тіркеспейді) дыбыс құрамына ешкім күманданбайды, ал соңғы **у** арқылы жазылып отырған екі сөзге (тіркеске) келгенде тағы да күмән басталады. Шындығында, барлығы ды әртүрлі дауысты дыбыстар мен үнді, жуысыңқы, еріндік **у** дауыссызының тіркесі болып келеді. Егер **ау** немесе **әу** сөздерінің дыбыс құрамына күман тумайтын болса, онда солармен теңдес **ұу, үу** сөздерінің (тіркестерінің) де дыбыс құрамына күмән тумас керек.

Тағы да, басқаны былай қойғанда, **у** таңбасы бір дауысты дыбыс болса, онда **алу** деген сөздің құрамында неге жуан айтылып, **елу** деген сөздің құрамында неге жіңішке айтылып, екі түрлі дыбысталады деп ойлану керек. Өйткені **у** таңбасы бірде құрамында жуан **ұ** дауыстысы бар **ұу** тіркесін, бірде құрамында жіңішке **ү** дауыстысы бар **үу** тіркесін белгілейді.

Екінші, түбір морфема мен қосымша морфеманың жігі:

**май [май], ма-йы [ма-йы], ми [мый], ми-ы [мы-йы], тау [тау], та-уы [та-уы], ту [тұу], ту-ы [тұ-уы]**

Қазақ тіліндегі қосымшалар талғампаз келеді. Соның бірі түбір морфеманың дауысты не дауыссызға бітуіне байланысты. Егер **май, тау** сөздерінің дауыссыз дыбысқа бітуіне байланысты қосымша морфема дауысты дыбыстан басталып тұрса, онда «дауысты дыбысқа бітіп тұр» деген **ми, ту** сөздеріне неге дауысты морфема жалғанып тұр? Ендеше бұл сөздер де дауыссыз дыбысқа бітіп тұр деген сөз. Олай болса, транскрипциядан көрініп тұрғандай, дыбыс тіркесінің екінші сыңары **й, у** дауыссыз дыбыстары болғаны. Бұған қоса, мысалға алынып отырған сөздердің бәрінің буын және морфем құрамы біркелкі екенін де ескеру керек.

Ендеше **й, у** таңбаларының дыбыс құрамын **ий, ий, ұу, үу** дыбыстарының тіркесі деп қарау керек. Өйткені қазақ жазуының құрамына енген кірме әріптердің дыбыстық мәнін анықтау аса күрделі фонетикалық мәселеге айналып отыр. Себебі қазақ тілінің оқулықтары мен әдістемелік құралдарында кірме әріптер бірде дауысты, бірде дауыссыз, ал енді бірде дыбыс тіркесі деп түсіндіріліп келеді. Соның салдарынан, өкінішке орай, олар қазақ тілінің жаңа әліпби құрамынан шығарылмай жатыр. Жаңа фонема болу үшін ол қазақ сөзінің **үндесім әуезін** бұзбайтын (қазақтың жуан үндесім әуезді **мый** сөзін жіңішке әузбен **ми** деп айтқызбайтын), **буын құрамын** өзгертпейтін (**ми-ы** деп тасымалдатпайтын), **морфем жігін** үзбейтін (**мый+ы** деп түбір мен қосымша тіркесіміне нұқсан келтірмейтін) болу керек.

Ал, қазақтың төл сөздерінің жазылымына еніп кетіп, дауысты дыбыс болып жүрген **я, ю** таңбаларының тек дыбыс құрамы жоғарыда көрсетілді. Дауысты болып жүрген **я** әрпі **йа, йә** тіркестерінің таңбасы болса, дауысты болып жүрген **ю** әрпі **йүү, йүү** тіркестерінің таңбасы болып табылады. Мысалы, **жая [жайа], сая [сайа], әлия [әлийә], дүрия [дүрүйә], аю [айүү], ою [ойүү], үю [үйүү], түйю [түйүү].**

Бұлардың да дыбыс құрамына жоғарыдағы әдістермен көз жеткізуге болады.

Қазақ сөзінің жаңсақ және ақиқат жазылымының салыстырма мысалдарының арасындағы буын, морфем, тасымал және әуез қайшылықтарды көрнекі байқауға болады. Жаңсақ жазылыммен

оның әдеби (төл) айтылымының арасында қазақ тілінің буын, морфем, тасымал және әуез заңдылықтарына кереғар қайшылықтар көзге ұрып тұр. Ал ақиқат жазылымымен оның әдеби (төл) айтылымының арасында ешқандай қайшылық жоқ. Тіптен *а-я-у*, *о-я-у*, *жа-я-удеп* буынға бөлудің ерсілігіне қарамай, қазақ тілінің болашақ иелері – жас ұрпақтың – көзін үйрете береміз бе?

Қазақ жазуының тарихында күрмегі шешілмей келе жатқан бұл мәселе әлі де ғалымдардың арасында талқылау көп. Өзге кірме таңбалардың жайы олардың бәрін қамту мүмкін емес. Алайда осындай талдаулардың арқасында келешек жас ұрпағымызды сауаттандырсақ екен деп ойлаймыз.

#### **Әдебиеттер:**

Байтұрсынов А. Тіл тағылымы (Қазақ тілі мен оқу-ағартуға қатысты еңбектері). – Алматы: Ана тілі, 1992. – 448 б.

Жүнісбек Ә.Әліпби ауыстыруды жазу реформасына айналдыру керек. – Орал: Ағартушы, 2007. – 32 б.

Жунисбек А. Введение в сингармоническую фонетику. – Алматы: Арыс, 2008. – 108 б.

Жүнісбек Ә. Қазақ тіл білімінің мәселелері. – Алматы: Абзал-ай, 2018.

## «СИБИРСКИЙ» И «НОВГОРОДСКИЙ» ЯЗЫКИ В СОВРЕМЕННОЙ РОССИИ КАК ЯЗЫКИ, СКРИПТЫ ИЛИ АРТЛАНГИ

Автор анализирует проблемы «сибирского» и «новгородского» языков в современной России. Предполагается, что идиомы, известные как «новгородский» и «сибирский» языки являются скриптами. Автор, с одной стороны, полагает, что современные скрипты – это письменные формы и языковые нормы, которые претендуют на статус нормы в виртуальном пространстве. С другой стороны, автор указывает, что «новгородский» и «сибирский» языки не могут быть определены как языки, поскольку не имеют все признаки языка. Анализируются основные особенности (отсутствие нормы, стандарта и письменной традиции, отсутствие постоянных пользователей языка, ситуативный и виртуальный характер) «сибирского» и «новгородского» языков. Поэтому, предполагается, что векторы развития «сибирского» и «новгородского» языков как «скрипт» неопределенны, что актуализирует их дальнейший междисциплинарный анализ.

**Ключевые слова:** языковое воображение, идентичность, «новгородский язык», «сибирский язык», скрипт.

Утверждение, что современная Российская Федерация является многонациональной страной, успело стать общим местом в историографии, а число исследований, сфокусированных на проблемах сохранения и развития национальных идентичностей, языков, литератур и культур насчитывает несколько тысяч библиографических позиций. Проблемы национальных идентичностей и языков тех наций, которые имеют свои государственности или их отдельные элементы в Российской Федерации, достаточно подробно описаны в академической российской и зарубежной историографии. Актуальная современная история национальных языков России, с одной стороны, не ограничивается только теми языками, которые имеют официально признанный статус как российским государством, так и академическим сообществом. Развитие интернета, тенденции кризиса и упадка традиционной офф-лайн культуры и литературы, с другой стороны, актуализируют потенции и возможности виртуального пространства для создания, развития и продвижения или восстановления и популяризации мертвых языков, что ведет в развитии нового явления, единое определение которого в научной литературе отсутствует. Речь идет о попытках интернет-активистов возродить и фактически создать заново те языки, которые не выдержали конкуренции с другими более успешными историческими соперниками и к первой четверти 21 века фактически принадлежат к числу мертвых языков.

Новейшая история российского интернета 2000–2010-х годов обеспечивает лингвистов двумя показательными случаями «изобретения» [Gorski Severo 2016] языков, которыми стали «сибирский» и «новгородский» языки.

Анализируя эти явления автор полагает, что относительно т.н. «сибирский язык» и «новгородский язык» использование понятия «язык» является некорректным. Автор в отличие от своих предшественников, которые писали об этой проблеме, не воспринимает «сибирский» и «новгородский языки» как сознательные антирусские проекты, но полагает, что другие факторы не позволяют воспринимать их как языки в значении «сложной знаковой системы, естественно или искусственно созданной и соотносящей понятийное содержание и типовое звучание (написание)». «Сибирский» и «новгородский языки» не могут быть определены как языки потому, что не имеют литературной нормы, не имеют развитой литературной традиции, не пережили процессы унификации, а вопрос о наличии их носителей, которые постоянно используют «сибирский» и «новгородский языки» для коммуникации является дискуссионным. Поэтому, анализируя «сибирский» и «новгородский языки», по мнению автора, более логичным является использование дефиниции «скрипта».

Понятие «скрипта» активно используется в романской филологии, где означает «письменную форму диалекта» или / и «региональную письменную традицию, совокупность приемов написания» [Маркова 2013]. Эти дефиниции вполне применимы для описания ситуации до начала 21 века, когда культура и идентичность были преимущественно офф-лайн явлениями. Корректируя определение «скрипта» в условиях динамичной виртуализации культуры, автор полагает, что скрипта – это письменная форма и совокупность приемов написания региональной языковой нормы или совокупности языковых явлений, которые претендуют на статус нормы в пространстве Интернета с использованием блогов, сайтов и других ресурсов.

В этой статье автор анализирует попытки лингвистических активистов создания и продвижения в интернете двух «языков» – сибирского и новгородского. «Сибирский язык» и «новгородский язык» как скрипты существуют исключительно как виртуальные он-лайн лингвистические конструкции, не имеют нормы и стандарта, а наличие их активных пользователей в ежедневной коммуникации является спорным.

Проанализируем основные особенности и направления развития «сибирского» и «новгородского» языков как скрипт в условиях культурной смены локализации языка, его миграции из офф-лайн пространства в интернет как сферу доминирования он-лайновых культурных практик и лингвистических явлений.

Фактически «сибирский» и «новгородский языки» не являются языками, но существуют и функционируют как скрипты, которые пытаются усилиями их создателей и условных носителей обрести статус языка. «Сибирский» и «новгородский языки» на современном этапе существуют почти исключительно и только в интернете, но в виртуальной реальности их роль и место маргинальны в сравнении с другими живыми и даже искусственными языками. «Сибирский» и «новгородский языки» следует относить к конструктам, хотя их сторонники и отрицают искусственный статус. Например, Ярослав Золотарев, один из создателей «сибирского языка», полагал, что «не думаю, что будет сложно сконструировать сибирский язык, залив в диалектную грамматику нужное количество татарских корней» [Фаворов 2007]. Однако попытки «конструирования» языка не привели к серьезным результатам.

Проект Сибирской Википедии (Сибирска Википеддя) оказался крайне неудачным: раздел, созданный в 2006 году, был удален уже в 2007 году. Попытки продвижения «сибирского» языка в интернете практически с самого начала воспринимались как политические и идеологические, а примеры «сибирского словаря» с фразами были больше похожи на литературную мистификацию: например, «Краткий русско-сибирский разговорник» содержал следующие примеры «сибирского языка» в переводе с русского: «Здесь кто-нибудь говорит по-русски? / Здесь кто-нибудь баёт на блошистом говоре-том?», «Я осуждаю колониальную политику / Я бессужу заимошну правильную», «Не стреляйте, я беженец! / Не бухайте, я утеканшык!» или «Я уже сжег паспорт. Позвольте мне перейти границу (реку) / Я уже пожег личну басму. Дозвольте мне перелазить передел (умку)» [Казинник, Ткачев 2011]. Попытки развития «сибирского языка» не привели к конкретным результатам, перейдя в политические дискуссии [Золотарев, 2016], а восприятие самого «языка» как литературной мистификации и радикального лингвистического творчества содействовало его постепенной маргинализации.

Сферы использования и применения «сибирского» и «новгородского» языков неясны и являются крайне условными: регулярные СМИ на этих «языках» отсутствуют, они не используются в сферах образования и государственного управления, но и «сибирский», и «новгородский» языки активно используются в проектах, направленных на воображение, изобретение и продвижение альтернативных форм региональных идентичностей с перспективой их трансформации в национальные идентичности новых наций, которые в условиях глобализации воспринимаются как «воображаемые сообщества» и становятся «изобретенными традициями».

В этом контексте появление «сибирского» и «новгородского» языков – частный случай лингвистического воображения, языкового строительства и маргинальная версия изобретения традиций и воображения сообществ. Дефиниция «маргинальность» в этом контексте используется, чтобы показать только степень невидимости «сибирского» и «новгородского языков» как культурных и интеллектуальных проектов. Поэтому связи проекта «сибирского языка» с политическими и идеологическими противоречиями, проблемами, попытками актуализировать региональную, сибирскую альтернативу московской политики в целом и современной российской версии идентичности в частности содействовали политизации и идеологизации дискуссий и дебатов относительно его развития.

«Новгородский» язык в этой ситуации значительно отличается от «сибирского языка»: попытки лингвистических активистов культивировать или «возродить» этот язык остаются практически невидимыми и маргинальными в контекстах других культурных активностей, которые вызвали более широкий резонанс как в обществе в целом, так и в академическом сообществе в частности. Если эксперименты по созданию «сибирского» языка, как утверждают лингвистические интернет-активисты, были остановлены под политическим давлением, что создало дополнительную рекламу проекту, то «новгородский» язык известен меньше. Если «сибирские» лингвистические активисты использовали в академических исследованиях живых региональных говоров и диалектов Сибири, то сторонники ревитализации «новгородского» языка ориентировались на академические штудии советских и российских авторов [Зализняк 2004], хотя тексты, написанные ими на восстановленном «новгородском» языке, являются противоречивыми и спорными, так как язык, предложенный ими, актуализирует

только одну из возможных и вероятных траекторий развития новгородского языка. Сторонники возрождения «новгородского» языка не стремятся превратить его в живой язык, но используют для культивирования идентичности в контекстах альтернативной истории, что позволяет локализовать «новгородский» язык среди артлангов [Пиперски 2017] или искусственных языков. «Новгородский язык» формально имеет грамматические и лексические нормы, будучи в значительной степени стандартизированным современными энтузиастами [Posthumus 2010; Posthumus 2019], но, как «сибирский язык», сферы его применения ограничены. Узус использования «новгородского» языка ограничивается несколькими интернет-ресурсами.

Подводя итоги, во внимание следует принимать несколько факторов.

Во-первых, «сибирский язык» и «новгородский язык» являются новейшими лингвистическими конструктами и результатами деятельности лингвистических интернет-активистов и активных пользователей различных сервисов интернета. Во-вторых, «сибирский язык» и «новгородский язык» могут быть локализованы среди скрипт, потому что не обладают всеми признаками языка. В-третьих, «сибирский язык» и «новгородский язык» являются почти исключительно виртуальными он-лайн проектами. В-четвертых, «сибирский язык» и «новгородский язык» стали последствиями кризиса традиционных моделей нации, нации-государства и идентичности, которые имели преимущественно офф-лайн основания книжной культуры. В-пятых, появление таких форм лингвистического конструктивизма как «сибирский язык» и «новгородский язык» стало следствием, с одной стороны, кризиса модели государства-нации, а, с другой стороны, незавершенностью процессов развития современной нации в России, где этот концепт исторически и политически не смог заменить концепт государства-империи, которая предусматривает отсутствие политической нации как сообщества граждан.

Таким образом, автор полагает, что основные векторы и траектории дальнейшего развития таких явлений как «сибирский язык» и «новгородский язык», которые фактически могут быть определены как «скрипты» или даже артланги (искусственные языки), но не «языки» в понимании лингвистики, неопределенны, что придает значительную актуальность их дальнейшему академическому изучению в условиях междисциплинарного подхода в контекстах лингвистики (грамматика, синтаксис, фонетика, лексикология), социологии (сферы использования и применения), культурологии (культурные роли языка), политологии (политические функции языка) и исследований идентичности (роль языков в трансформации идентичностей их носителей и активных пользователей).

### Литература

- Зализняк А.А. Древненовгородский диалект. – М.: «Языки славянской культуры», 2004. – 872 с.
- Золотарев Я. Сибирский язык: как запрещали лингвистический эксперимент (2016): [Электронный ресурс]. URL: <https://afterempire.info/2016/12/14/siberian/> (дата обращения: 10.09.2019).
- Казинник М., Ткачев Дм. Пятиминутный путеводитель по сибирскому сепаратизму // Esquire. 2011. № 71: [Электронный ресурс]. URL: <https://esquire.ru/relaxation/2172-5-min-siberia/> (дата обращения: 10.09.2019)
- Маркова Е. О скриптологическом исследовании средневековых французских рукописей // Научные ведомости БелГУ. Серия «Гуманитарные науки». 2013. № 27 (170). – С. 52–56.
- Пиперски А. Конструирование языков. От эсперанто до дотракийского. – М.: Альпина нон-фикшн, 2017. – 223 с.
- Фаворов П. Маршал говоров // Большой город. 2007. 14 марта: [Электронный ресурс]. URL: [http://bg.ru/society/marshal\\_govorov-6539/](http://bg.ru/society/marshal_govorov-6539/) (дата обращения: 10.09.2019).
- Gorski Severo Cr. The colonial invention of languages in America // Alfa: Revista de Linguística. 2016. Vol. 60. № 1. – P. 11–28.
- Posthumus M. A Reference Grammar of the Novegradian Language. Довѣднике по граматикѣ новеградескаево лизика (2010): [Электронный ресурс]. URL: [file:///C:/Users/User/Downloads/novegradian\\_grammar\\_web\\_20\(2\).pdf](file:///C:/Users/User/Downloads/novegradian_grammar_web_20(2).pdf) (дата обращения: 10.09.2019).
- Posthumus M. Novegradian Lexicon (2019): [Электронный ресурс]. URL: <https://www.veche.net/lexicon/novegradian> (дата обращения: 10.09.2019).

## О ПОНЯТИИ ВРЕМЕНИ И ПРОСТРАНСТВА В СКАЗОЧНОМ ДИСКУРСЕ (на примере казахских и русских народных сказок)

В статье рассматривается понятие времени и пространства в сказочном дискурсе. Проведенное исследование показывает, что пространственно-временные характеристики в сказках русского и казахского народов носят совокупный характер. Для данного жанра демонстрация главного героя, его действия в дальнейшем более важны чем констатация действия во времени и пространстве. В этом плане сказка действует вне времени и пространства.

Автором выделены доминантные и характерные лексемы в русском и казахском языках, характеризующие пространственно-временные понятия.

**Ключевые слова:** хронотоп, время, пространство, лексемы, доминант, сказочный дискурс.

Н. К. Култанбаева  
Ақтобе қ., Қазақстан

## ЕРТЕГІ ДИСКУРСЫНДАҒЫ КЕҢІСТІК ПЕН УАҚЫТ ҰҒЫМЫ (қазақ және орыс халық ертегілері бойынша)

Мақалада ертегі дискурсындағы кеңістік пен уақыт ұғымы туралы сөз болады. Қазақ және орыс халық ертегілеріндегі уақыт пен кеңістіктің көріну сипаты, ерекшеліктері мен ұқсастықтары, кеңістікке қатысты доминант лексемалар салыстырыла сипатталады.

**Түйін сөздер:** хронотоп, уақыт, кеңістік, доминант лексемалар, ертегі дискурсы.

Адам баласының ақпараттық өрісінің маңызды элементінің бірі – уақыт. Ол – адам санасы кеңістігінің ең маңызды қаруы. Адам санасының әрбір микроқұрылымы сол құрылымға ғана тән уақыт пен кеңістікке ие болады [Аитова 2010, 29].

Ал көркем уақыт – бұл көркем шығармадағы әрекеттің бірізділігі мен тәртібі. Кеңістік – көркем шығармадағы кейіпкер өмір сүретін барлық әрекеттер орнының жиынтығы. Уақыт пен кеңістік логикалық жақтан біріге отырып хронотоп құрады. Әрбір жазушы мен ақынның өзінің сүйікті хронотопы болады. Онда кейіпкер де, заттар да, сондай-ақ сөйлеу әрекеті де осы уақытқа бағынады. Алайда шығарма үнемі бірінші орынға басты әрекет етуші тұлға шығады. Жазушы немесе ақын өзіндік стиліне орай, әрқайсысы өзінің белгілі бір көркем тәсілімен уақыт пен кеңістікті қызықты етіп суреттейді.

М.М. Бахтин: «Уақыттың белгілері кеңістікте ашылады, ал кеңістік уақытпен өлшенеді және ұғынылады», – дейді. Сондай-ақ ол үнемі қайталанып отыратын идеяға оралатын циклдық құрылымды білдіретін халықтық– мифологиялық уақытты бөліп қарастырады. Сонымен бірге М.М.Бахтин хронотоп категориясы туралы былай деп жазады: «О событии, можно сообщать, осведомить, можно при этом дать точные указания о месте и времени его свершения. Но событие не становится образом. Хронотоп же дает существенную почву для показа — изображения событий» [Бахтин 1975, 235]. Бахтиннің айтуы бойынша образды жасау үшін оның бейнелік қызметі пайда болады. Хронотоптың жалпы мағынасы көркем шығармадағы оқиғаның белгілі бір кеңістік пен уақыттың өзара тәуелділігін де сипатталуымен қорытылады. Хронотоп ұғымы дәуірлердің эстетикалық тенденциясымен, автор стилімен, шығарманың жанрымен байланысты болып келеді.

Әрбір дәуір өзінің әлемін, өзінің стилін, ойлауын, өзінің уақыты мен кеңістігінің тұжырымдамасын жасайды. Ойлаудың ежелгі формасы уақытты антропоморфтық түсінумен белгіленген, оған моральдық ойлау бөлінген. Архитиптік уақыт антитарихи, ол цикльді дәл өлшемді қажет етпейді. Өмір кездейсоқтықтан босап, мәңгілік құрылымға айналады. Болашақ мифопоэтикалық дәстүр тағдырмен анықталады, әрекет етудің уақытша ұзақтығы кеңістікті жеңумен өлшенеді.

Бұл келтірілген тұжырымдардың барлығы жалпы көркем шығармадағы кеңістік пен уақыт теориясына байланысты болса, одан әрі біз мақаламызда ертегі дискурсындағы, нақтырақ айтсақ қазақ және орыс халықтары ертегісіндегі кеңістік пен уақыт тұжырымына тоқталатын боламыз.

Жалпы ертегі дискурсындағы хронотоп – баяндау ағымында болып жатқан уақыт, яғни ертегінің басталуынан бастап, негізгі кейіпкердің соңғы жеңісіне дейінгі, сондай -ақ оқиғаның өту

орны, ертегі кейіпкерінің басынан өткен сынақтарына дейінгі мерзім. Қиял ғажайып ертегідегі хронотоп шынайы географиялық нысан мен шынайы уақыттан қашықтатылған. Ал оның аталмыш екі халықтың ертегілерінде көріну сипаты әр түрлі болып келеді. Айталық, хронотоп орыс ертегілерінде нақты көрсетілмейді: әрекет ету уақытының көрінуі өте сирек кездеседі, ал әрекет ету орны жинақтық, әрі жалпы сипатта көрінеді. Қиял ғажайып ертегісінің топонимикасы үнемі әр түрлі болып келеді, алайда нақты емес, қандай да бір іс әрекеттің нақты орны айқын емес. Қиялға жайып ертегінің кеңістігі үнемі шартты түрде белгіленген болып келеді. Бұған мысал ретінде орыс ертегілерінің мәтінінде жиі кездесетін мына тіркестерді атауға болады: «за тридевять земель», волшебный и дремучий лес, избушка на курьих ножках және т.б. Мысалы, орыс халық ертегілеріндегі әрекет ету орны «*тридесятое государство*», «*некоторое царство, некоторое государство*», т.б. болып келсе, қазақтың қиял ғажайып ертегілерінде «*Ертеде, ерте-ертеде, ешкі құйрығы келтеде, Қаратаудың ойында, Қарасудың бойында Қазанқап деген бір кедей болыпты*», «*Ерте, ерте, ертеде Құбахан деген хан екен. Малының саны жоқ екен. Хандығының көлемі жеті айшылық жол екен. Құбахан жалғыз ұлкөреді, ұлының атын Тотан қояды*» деп Қаратау, Қарасу, Қазанқап, Құбахан деген секілді нақты жалқы есімдер көрінетін жерлері де кездесіп отырады. Осымен бірге «*Ертеде бір бай болған екен, төрт түлігі сай болған екен, бір перзентке зар болған екен*» немесе «*Ертегім ерте екен, ешкі жүні бөрте екен, ерте заманда бір аңшы болған екен*» деген секілді уақыты мен оқиғаның өту орны нақты көрсетілмеген жағдайлар да жиі кездеседі.

Халық ертегілерінің уақыты ол да нақты емес: «аз жүрді ме, көп жүрді ме» – «долго ли, коротко ли», «қанша өзен тасыды, қаншама су ақты» – «много воды утекло», «много ли, мало ли» – «аз ба, көп пе», «жил да был» – «ерте, ерте, ертеде» және т.б. Сондай – ақ қиял ғажайып ертегінің уақыты бір текті емес дамып отырады. Мысалы, не өте баяу («лежит тридцать три года») немесе өте жылдам («не по дням, а по часам», алты күнде жүріпті, алпыс күнде сөйлепті).

Ертегіде кейіпкер әрекеті уақыт пен кеңістіктен тыс болып жатқандай көрінеді. Айталық ертегідегі хронотоп ертегінің басталуынан-ақ көрінеді. Барлық ертегі сюжеті басты кейіпкердің үйінен кетуінен және оған түрлі жағдайлардың әсер етуінен басталады. Кейіпкер үйінен шыққаннан бастап түрлі оқиғаларға, сынақтарға тап болады. Одан әртүрлі кейіпкерлер мен оқиғаларға қатысты ертегі желісінің өрбуі, шиеленісуі, шарықтау шегі мен аяқталуы баяндалады.

Ертегінің көркемдік қасиеті жоғары, әрі халықтың сөз өнеріне айналған фольклордың эстетикалық формасының бір түрі болып саналады. Сондықтан ол қоғам дамуының әр кезеңіндегі халық идеясын бейнелейді, оның эстетикалық, этикалық түсініктерін танытады. Ал осылардың айқын көрінуінде ертегілік шығарманың композициясы мен тілі өзіндік рөл атқарады.

Қазақ ертегілерінің құрылымы өзге халықтар ертегілерінің композициясымен ұқсас: кейіпкерді таныстыратын **бастама**, оқиғаны эпикалық пафоста **баяндау** және бақытты **аяқталу**, міне бұлар көптеген халықтың классикалық ертегісіне тән элементтер. Мұндай ұқсастық жалпы ертегі жанрының типологиялық табиғатынан және халық идеясын мадақтау мақсатынан туған. Әрине, көп ел ертегісінің композициялық ұқсастығы мүлде бірдей деуге болмайды. Ұқсастық жалпы құрылымда, оның бөлшектерінде. Ал, бөлшектің ішкі құрылымы мен мазмұны әр халық ертегісінде әр басқа [Қасқабасов 2009, 193].

Ал ертегі сюжетіндегі оқиғалардың бірізділігі мен баяндалу тәртібі бірден бір кеңістік пен уақытқа тәуелді. Айталық, кейіпкер үйден шығысымен-ақ ертегінің оқиғасы күрделеніп, қоюлана түседі: оның атқаратын істері мен өтетін бөгеттері қиындай береді, бірақ ертегіге қосымша кейіпкерлер аз да болса енгізіліп отырады, ал енгізілген күннің өзінде (ғажайып көмекшілер, немесе қаскүнем) олар қаһарманның ерекше күшін, тағы басқа қасиетін аша түсуге қызмет етеді. Ертегінің эпикалық, яғни оқиғалық бөлшегі түгелімен дерлік кейіпкердің әртүрлі мақұлықтар, құбыжықтармен кездесіп, олармен шайқасуынан тұрады. Неше түрлі жауыздармен қақтығысып, оларды жеңуі – ертегінің нағыз шарықтау шегі. Ал, барлық жауын жеңген қаһарманның адал жарымен аман-есен еліне оралуы – көркем ертегінің шешімі (развязка).

Бұл – қарастырып отырған қазақ және орыс халықтары ертегісіне тән құрылым. Ал, енді олардың өзіндік ерекшеліктері қандай, осы мәселені зерттеп көрелік.

Орыс ертегі дискурсындағы айтарлықтай ерекшелік – ол ертегілік әрекеттегі уақыт пен кеңістіктің қалыс қалуы. Онда ертегінің басталуының екі түрі көрінеді:

– ертегі сюжетінің басталуы анық болмайды да бірден әрекет етуші тұлға аталады («*Жили в одном селе старик да старуха*», «*Жил-был царь с царицею*», «*Жила в одной деревне крестьянка, вдова*»);

– ертегі кейіпкерінің орны нақты көрінбейді («*В некотором царстве, в некотором государстве*», «*За тридевять земель, в тридесятом государстве*»).

Орыс ертегісінде әрекет етудің уақыты мен орны өте сирек кездеседі. Ол орыс ертегілерінде жинақтық мәнге және жалпы сипатқа ие екендігіне тағы бір мысал келтірсек, «*Прошло срочное время*», «*отслужил солдат службу*», «*Ехал как-то*», «*В старые годы*», «*В стародавние времена*» – осылайша әрекет ету хронологиясы аз көрсетіледі. Ертегіде әрекет уақыт пен кеңістіктен тыс өтіп жатады. Қиялдың, оқиғаның өрбуі ертегінің басталуы мен соңында нақты көрінеді. Ертегінің басталуында уақыт пен әрекет етуші кеңістік анық көрсетілмейді, ал соңында ертегінің көркем әдебиет екендігін нұсқайды. Орыс ертегілерінде уақыт пен кеңістіктен гөрі әрекет етуші тұлғаны, яғни басты кейіпкерді, оның сынақтарын көрсету сюжеттің одан әрі дамуы үшін маңыздырақ болып келеді.

В.Я. Пропп, кейін Е.М. Мелетинский және басқа да ертегіні зерттеушілері көрсеткендей қиял ғажайып ертегілердің сюжетінде оқиғаның өту орындары нақты емес доминант лексемалармен де берілетінін көрсеткен:

– «царство»: *Тридесятое царство, Неведомое царство, Подводное царство, Подземное царство, Змеиное царство*;

– «королевство»: *Тридесятое королевство, Иноземное королевство*;

– «государство»: *Тридевятое государство, Подсолнечное государство*; [Брейгер 2013, 119]

Ал қазақ ертегілеріне келер болсақ, онда оқиғаның өту орны ішінара болса да кездесіп отырады. Оған жоғарыда келтірілген мысалдарымыз дәлел бола алады. Қайталап айтар болсақ «*Ертеде, ерте-ертеде, ешкі құйрығы келтеде, Қаратаудың ойында, Қарасудың бойында Қазанқап деген бір кедей болыпты*»

«*Ерте, ерте, ертеде, ешкі жүні бөртеде, байлар киіп қырмызы, жарлының тоны келтеде, жердің үстінде, аспанның астында, әлемді таң еткен Келтен деген бір сараң бай болыпты*» немесе «*Ер Төстік*» ертегісінде кездесетін алғашқы сынақтардың бірі өтетін жер атауы – *Сорқұдық*, жер астындағы ел атауы – жылан *Бапы хан елі*, жер асты немесе жер үстіндегі сынақтар және т.б. атаулар аз да болса, оқиғаның болу орнынан хабардар етеді.

Сонымен бірге ертегі дискурсындағы кеңістік пен уақыттың тағы ерекшелігі ертегі кейіпкерлерінің қаншама уақыт өтсе де, уақыттың ағымына қарсы тұрып, жастық шағын сақтап қалуы. Бұл ерекшелік екі халықтың да ертегісінде кездеседі. Онда ертегінің бақытты болып аяқталуы мен ертегі қаһармандарының сол баяғы жас кейіптеріне оралуы суреттеліп отырады. Бұны кейіпкердің уақыт ағымынан мүлдем алшақ болуымен түсіндіруге болады. Яғни, оқиғаға қатысушылар үшін (бас қаһарман, етегіні баяндаушы, басты қаһарманды қоршаған түрлі кейіпкерлер) көркем уақыт түрлі ағымда өтіп жатады. Қиял ғажайып ертегіде уақыт аралығы ұзақ болып келеді. Алайда, осыған қарамастан ең бастапқыдай кейіпкерлер жас күйінде қала береді. Мысалы, «*Иван-царевич и белый полянин*» атты орыс ертегісінде «... Долго ли, коротко ли, низко ли, высоко ли – скоро сказка сказывается, да не скоро дело делается – приехал он в царство змеиное, убил царя-змея, освободил из неволи прекрасную королевну и женился на ней; после того воротился домой и стал с молодой женою жить-поживать да добра наживать» деп аяқталса, қазақ халқының ертегісінде Ер Төстік пен Кенжекей сұлудың араға бірнеше жылдар салып кездесуіндегі «...*бәрі де баяғыдағы айырылған күндегі қалыптарына келіп, қайтадан жасарады. Ер Төстік жас жігіт, Кенжекей жаңа түскен келіншек, Шалқұйрық бесті ат қалпына келеді. Одан кейін ұзақ жылдар өмір сүріп, мақсаттарына жетеді*» деген жолдар айғақ.

Қорыта айтқанда, мақаламызда сөз болған қазақ және орыс халықтары ертегі дискурсындағы кеңістік пен уақыттың көрінуі басқа көркем шығармаларға қарағанда айтарлықтай ерекшелікте екендігін байқадық. Оған себеп ертегі дискурсында көркем уақыт пен кеңістік шынайы географиялық нысан мен шынайы уақыттан қашықтатылып көрсетілуі; Ертегідегі әрекет ету уақыты өте сирек кездеседі, ал әрекет ету орны жинақтық, әрі жалпы сипатта көрінуі; Қиял ғажайып ертегінің топонимикасы үнемі нақты емес, қандай да бір іс әрекеттің нақты орны айқын еместігі; Қиял ғажайып ертегінің кеңістігі үнемі шартты түрде белгіленген болып келуінайтұға болады.

### Әдебиеттер

Аитова Н.Н. Поэтикалық ақпарат және оның уақыт пен кеңістікте таралуы // ҚазҰУ хабаршысы № 4. 2010 ж. – 29 б.

Бахтин М.М. Вопросы литературы и эстетики. – М., 1975. – С. 235.

Брейгер Ю.М. Вербализация испытания в сказочном дискурсе в русском и французском языках. Дисс. ... канд. филол. наук. – Волгоград, 2013. – 119 с.

Қасқабасов С.Қ. Фольклор туралы. – Алматы, 2009. – 193 б.

Қазақ ертегілері. II том. – Алматы: Жазушы. 2000. 273-б.



## СОСТАВ И ВАРИАНТЫ ЗВУКОВ В КАЗАХСКОМ ЯЗЫКЕ

В статье рассматривается современный состав звуков казахского литературного языка, проводится анализ их вариантов. Автором освещаются некоторые вопросы ассимиляции и диссимиляции звуков в речевом потоке.

**Ключевые слова:** казахский язык, звуки, состав и модификация.

А. Мухалиев  
Бишкек к., Кыргызстан

## КАЗАК ТИЛИНДЕГИ ТЫБЫШТАРДЫН КУРАМЫ ЖАНА ТҮРЛӨРҮ

Кыскача мазмуну: макалада казак адабий тил, алардын Чыгарылган үндөрдү заманбап бөлүгү менен мамиле.

**Ачык сөздөр:** Казак тили, үн, курамы жана өзгөртүү.

Казак тилиндеги тыбыштардын түрлөрүн кароо үчүн алды менен алардын курамын аныктоо керек. Анткени азыркы кезге чейин казак тилинин тыбыштык курамына деген көзкараш бир нукка түшкөн эмес. Окуулуктар менен окуу программаларында болсун, илимий монографиялар менен макалаларда болсун тыбыштардын саны ар түрдүүчө аталып келет. Ал эми тыбыштык түрлөрдү накта сыпаттап чыгуу үчүн казак тилинин тыбыштык курамы дурус аныкталган болуу керек. Казак тилинин тыбыштык курамы дурус аныкталбаса, анда тыбыш түрлөр жүйөсү бузулат.

Казак фонетикасынын XIX к. башталыш кезинде турган Ильминский, Мелиоранский, Радлов ж.б. болсо, XX к. башында А. Байтурсун уулу, Х. Досмухамед уулу, К. Кемеңгер уулу, К. Жубанов ж.б. казак фонетикасынын изилдөө деңгелин жаңы деңгелге көтөрдү, казак сөзүнүн төл айтым үлгүсү ушул изилдөөлөргө негиз болду. XX кылымдын экинчи жарымына карай И. Кеңесбаев, Б. Калиев, Ж. Аралбаев, А. Нурмаханова, С. Мырзабеков, А. Айгабылов ж.б. орыс жана батыш фонетистеринин фундаменталдуу фонема теориясына таянып талдап чыкты. Аталган аалымдардын изилдөөлөрү окуулуктар менен окуу программаларына толук кирип, ушул күнгө чейин пайдаланып келет. С. Мырзабеков өзүнүн көптөгөн эмгектеринде А. Байтурсун уулу, Х. Досмухамед уулунун эмгектерине таянып, казак сөзүнүн айтым үлгүлөрүнө байланыштуу мол мурас калтырды.

Казак тилинин тыбыш түрлөрүнө келгенде изилдөөчүлөрдүн тек жоон-ичкеликке гана маани берип келгенин көрөбүз. Жоон-ичкеликте алгач андагандардын бири Н. И. Ильминский болгон: «Подобно гласным согласные бывает толстые и тонкие» [Ильминский 1860, 8].

А. Байтурсун уулу казак тилинин тыбыштык курамын аныктаганда да тыбыштардын үндөшүм варианттарына баса көңүл бөлүп отурган. Аалым казак тилинин негизги арип үлгүлөрүн түгөлдөй келип, «өзгө 19 дыбыштардын һәр кайсысы бирде жуан, бирде жіңішке айтылат» [Байтурсунулы 1992, 324] деп белгилейт да, казак тилинде баш-аягы 43 үндүү-үнсүз бар деп корутунду жасайт. А. Байтурсун уулу жоон-ичкелик үндүү менен үнсүзгө бирдей таандык деп карап, казак тилинде 5 үндүү бар деген корутундуга келет.

Бул жерде казак тилинин тыбыштык курамы үндөшүм тургусунан жана да тилдин көлдөлөң абалына байланыштуу бирде жоон, бирде ичке айтылган тыбыштык белгилери эскерилүүдө. А. Байтурсун уулу пикиринин эң башкы баалуулугу деп үндөшүм күүсүнүн казак тилиндеги бардык тыбыштарга бирдей таандык деп отургандыгын айтабыз. Бул өзүнө чейинки изилдөөчүлөр маани бербеген тилдик (фонетикалык) кубулуштун казак тили үчүн маңызын ачып берген натыйжа болду.

Казак сөзүнүн үндөшүүсүнө ошол кездеги казак изилдөөчүлөрүнүн баары көңүл бурган. Х. Досмухамед уулу «Бір сөздін жуан, я жіңішке болып екі түрлі айтылуы казак-кыргыз тилиндегі әр дыбыстын биресе жуан, биресе жіңішке болып қос айтылуынан» [Досмухамедұлы 1993, 83] деген. Х. Досмухамед уулу жоон жана ичке айтылыштын артикуляциялык себебин да түшүндүрүп өткөн: «Дыбыс жуан айтылганда тілдің түбімен айтылып, тамақтан (көмейден) шыққандай болып сезіледі; ошол дыбыс жіңішке айтылганда тілдің ұшымен сөйлегендей болып, ауыздың алдыңғы жағынан шыққандай болып сезіледі» [Досмухамедұлы 1993, 83]. Арийне, ағезде азыркыдай калыпташкан артикуляциялык илимий аппарат болгон эмес, ошондуктан аныктама карапайым тил менен түшүндүрүлгөн.

Аалым «ошол тыбыш» деп бир тыбыштын эки үндөшүм варианты тууралуу айтып турганын баса көрсөтөт. Аны азыркы фонетикалык тил менен айтсак, тыбыштын «үндөшүм аллофону же варианты» болот.

К.Кемеңгер уулунун казак тилинин тыбыштык курамын талдаганында «фонема» негизги фонетикалык бирдик катары пайдаланганын көрөбүз. Ошону менен катар зарыл деген жеринде кээ бир тыбыштардын артикуляциясынын сыпаттамасын берген.

Казак тили тыбыштарынын түрлөр (вариант) курамы негизинен *к-к*, *з-з* үнсүздөрүнүн артикуляциясына байланыштуу болду. И.Кеңесбаев «К мен к бір фонеманың эки түрлү айтылуусу (немесе жуанды, жіңішкелі вариарты) деп карау дұрыс эмес» [Кеңесбаев, Мұсабаев 1975, 251] деп, алардын арасындагы артикуляциялык айырмачылыкты сөз маанисин ажыратуунун ыгы катары карайт. Ал эми калган «дооыссыздардың эрқайсысының жуан-жіңішкелі болып келуінің фонематикалык мағынаға әсері жок» [Кеңесбаев, Мұсабаев 1975, 251] деген корутундуга келет. Агездеги фонетикалык теория ушул жобону бек тутунган. Илимий эмгектер менен окулуктарда алар жеке фонема катары берилген.

Сингармонизм теориясынын өнүгүшүнө байланыштуу казак тилинин тыбыштык курамы үндөшүм тургусунан кайта карала баштат. Сингармонизм теориясы казак тили тыбыштарынын курамын тек «әліпби дыбыстармен» («әліпби дыбыстар» деп А.Жүнісбек азыркы алфавит тизимине кирген тыбыштарды атайт) чектөөгө болбосун көрсөттү. А.Байтурсун уулунун илимий мурасын ар тараптуу изилдей келип, М.Жүсіп уулу казак тили тыбыштарынын айтым курамы татаал экенин көрсөттү. Казак жазуусунун негизинде сингармонизм жатканын жана алдагы убакытта ымла-эрежеге сингармонизмдин тикелей катышы болуу керектигин далилдеп, сунуштар жасаган. Сингармонизмдин атайы изилдениши анын басым, тон тариздүү сөз жасоо кызма турганын табиятын ачып берди. Индо-европа тилдеринде басымдын, муунчул тилдерде тондун тилдеги тыбыштардын башын курап, сөз этип бириктирген кызматы түрк (анын ичинде казак тили) тилдеринде сингармонизмге таандык экен. Буга жараша казак тилинин тыбыштык курамынын жиктелимин сунат. Сингармонизм тургусунан казак тили тыбыштарынын үндөшүм түрлөрүн Ж. Назбиев «... метод сингармонической сегментации казахской речи... позволяет выявить и описать все произносительные единицы в казахском языке и провести их сингармоническую интерпретацию» деп талдап чыкты.

Казак тилиндеги тыбыштык түрлөр салттуу фонетикада атайы изилдөө объекти болбогону менен, бирди-жарым талданган учурларын кездештирүүгө болот. Бирок ал талдоолордун негизги максаты мындай белгилердин курамы менен жүйөсүн аныктоо эмес, ошол тыбыштар «фонемабы, фонема эмеспи?» деген суроонун жообун издөө болуп калат. Мисалы, *к-к*, *з-з* үнсүз тыбыштарынын сыңарларынын фонологиялык мартабасы «фонема деңгелиндеби» же, керисинче, «вариант деңгелиндеби» деген суроонун жообунда бирдей пикир жок. Бир топ аалымдар:... «**к, к, ғ, г** дауыссыздарын эзирше қалыптасқан түсінік-түйсік тұрғысынан танып баяндау керек болат» десе, башка бир топ аалымдар тилдеги фонологиялык кызма турган эскерип, варианттар деп карайт. Соңку пикирге орыс жана кыргыз фонетистери де кошулат.

Экинчи топтогу эмгектер казак тилинин төл тыбыштарын бөлүп карайт да, кирме ариптер менен кирме тыбыштарды чыгарып таштайт. Анын бир көрнөк үлгүсү катары соңку чыккан «Казак грамматикасын» алууга болот. Бул багыттагы эмгектерде төл «алипби» тыбыштардын варианттарына атайы көңүл бөлүнөт. Мунун натыйжасында казак тилинин тыбыштык курамы сан жагынан «алипби» тыбыштардын курамынан артык болуп чыгат. Анткени бир тыбыштын бир нече айтым түрлөрү камтылып, тыбыш санынын өлчөмү салттуу изилдөөлөр натыйжасынан артык болуп чыгат. Мисалы, А.Байтурсун уулу казак тили тыбыштарынын курамы 5 үндүү жана 17 үнсүз деп алып, алардын жооничке үндөшүм варианттарын эсепке алып, казак тилинде 43 тыбыш бар деп корутундулайт. Бул жерде А.Байтурсун уулу айткан 43 тыбыш менен азыркы 42 арипти сан жактан жакындыгына карап өз ара чатыштырбоо керек. Булар бир-бирине жакындабаган, башка-башка маселелер. Бул казак тилинин жооничке үндөшүү күүсүн эсепке алгандагы сандык курам.

Эгер А.Жүнісбек сунган жол менен казак тилинин тыбыштык курамын аныктай турган болсок, анда казак тилиндеги негизги үндүүлөр 3 (кең үндүү, кууш үндүү жана куранды үндүү) тыбыш, ал эми үнсүздөр 17 тыбыш, бардыгы 20 тыбыш болуп чыгат. Аалым казак тилиндеги төрт үндөшүм күүсүн эсепке алат, анда 9 үндүү тыбыш, 66 үнсүз тыбыш, бардыгы 75 тыбыш болуп чыгат. Ошентип казак тилиндеги үндөшүм тыбыштардын курамдык саны чыгат. Эгер казак тилиндеги тыбыштардын калган түрлөрүн териштирип чыга турган болсок, анда тыбыш варианттары дагы да көбөйүп кете берет.

Казак тилинин салттуу фонетикасында тыбыш өзгөрүү атайын изилдөө объекти болбогондуктан, тыбыштардын артикуляциялык сыпаттамасы тыбыштардын негизги белгилерин аныктоо менен аякталып олтурат. Мисалы, үндүүлөр үчүн алардын жасалуу орду, жаактын катышы менен эриндин

катышы жеткиликтүү болсо, үнсүздөр үчүн алардын жасалуу орду, жасалуу ыгы менен үндүн катышы жеткиликтүү болот. Мында бул белгилер тыбыштарды жалпылап сыпаттоого жеткени менен, алардын түр-варианттарын сыпаттоого жеткиликсиз болот. Ошондуктан азыркы казак фонетикасында, жалпы лингвистикадагыдай, негизги жана кошумча артикуляциялык белгилер курамы киргизилет. Анын азыркы үлгүсүн «Казак фонетикасынын» сингармонизмге арналган бөлүмүнөн көрүүгө болот. Анда казак тыбыштарынын үндөшүм күүсүнө байланыштуу, мисалы, жоон-ичке күүнү берүү үчүн илгериңки-кедериңки сыяктуу аталыш киргизилди.

Ошентип, салттуу фонетикадагы башы ачылбай келген, тыбыштык аныктамасында аралашып жүргөн жасалуу, айтылуу жана угулуу түшүнүктөрү жиктелип бериле турган болду. Мисалы, артикуляциялык «тил арты» менен угулуш «жоон» аталышы бир тыбыштын аныктамасында катар кездеше берет. Кээде артикуляциялык аталыштын ордуна угулуш аталышы колдонула берет. Тыбыш сыпаттамасынын жүйөгө түшкөн түрү артикуляциялык аныктама (жалаң артикуляциялык атоолор пайдаланылат), акустикалык аныктама (жалаң акустикалык атоолор пайдаланылат), угулуш аныктама (жалаң угулуш атоолору пайдаланылат) болуп чыгат.

Буга өзгөчө токтолуп жаткан себебибиз бар. Мындан ары тыбыштык талдоону А. Жүнисбек сунган жаңы ыкманы жетекчиликке алуу менен ошол тартипте жүргүзүү зарыл. Мунун натыйжасында тыбыштык варианттардын аныктамасы татаалданып кетет. Анткени тыбыштын негизги артикуляциялык белгилеринин курамына кошумча артикуляциялык, акустикалык жана перцепциялык белгилер кошулат.

Казак тыбыштарынын ассимиляциялык жана диссимиляциялык түрлөрү өтө татаал кубулуш болот. Анын үстүнө алардын ассимиляциялык жана диссимиляциялык түрлөрүнө арналган изилдөөлөр өтө аз, бирди-жарым гана.

Азыркы учурда белгилүү болгон изилдөөлөрдө үнсүздөрдүн артикуляциялык белгилерине карай бөлүнүшү иштелип чыккан. Ал эми үнсүздөрдүн ассимиляциялык жана диссимиляциялык түрлөрүнүн үн түйүндөрүнүн термелишине карай жиктелиши эми гана колго алынууда. Салттуу фонетикада үн түйүндөрүнүн термелишин «үндүн катышы» деп белгилейт. Бирок «үн катышы» деген атоо акустикалык фонетиканын термини болуп табылат. Ал эми биз үчүн фонетиканын изилдөө объекти деп тыбыштардын жасалышын ишке ашырган алардын артикуляциялык аспекти өтө олуттуу.

#### Адабияттар

Байтүрсүнүлы А. Тіл тағылымы. – Алматы: Ана тілі, 1992. – 446 б.

Досмұхамедұлы Х. Аламан. – Алматы: Ана тілі, 1991. – 81–99 б.

Ильминский Н.И. Материалы к изучению киргизского наречия. – Казань, 1860. – 123 с.

Кеңесбаев І., Мұсабаев Ғ. Қазіргі қазақ тілі. – Алматы: Ғылым, 1975. – 304 б.

Қазақ грамматикасы. – Астана, 2002. – 815 б.

## **ПСИХОЛИНГВИСТИЧЕСКИЕ КОНЦЕПЦИИ ОБУЧЕНИЯ И ИСПОЛЬЗОВАНИЯ ИНОСТРАННОГО ЯЗЫКА**

Психологическая структура овладения русским языком как иностранным во многом может быть интерпретирована как смысловая модель процесса освоения иноязычного лингвистического опыта даже в условиях русскоязычного обучения. Несмотря на отсутствие продуктивного языкового окружения, обучение русскоязычных учащихся иностранному языку должно учитывать перспективную коммуникативную ориентацию, направленную на возможно более полную интеграцию общающегося на иностранном языке студента или школьника в иноязычный социум.

**Ключевые слова:** психоллингвистика, иностранный язык, обучение, лингвистический аспект, технология обучения.

*J. K. Saginova  
Aktobe, Kazakhstan*

## **PSYCHOLINGUISTIC CONCEPTS OF LEARNING AND MASTERING A FOREIGN LANGUAGE**

The psychological structure of mastering the Russian language as a foreign language can be interpreted as a semantic model of the process of mastering foreign language linguistic experience even in the conditions of Russian-language learning. Despite the absence of a productive language environment, teaching Russian-speaking students a foreign language should take into account the perspective of communicative orientation, aimed at the fullest possible integration of communicating in a foreign language student or student in a foreign language society.

**Keywords:** psycholinguistics, foreign language, teaching, linguistic aspect, learning technology.

Since the 70s of the twentieth century, a number of psycholinguistic works have appeared in the Russian scientific literature, which have made a significant contribution to the development of psychological ideas of teaching a foreign language at different levels of educational practices. The first publication was the work of A. A. Leontiev "Some problems of teaching Russian as a foreign language (psycholinguistic essays)" (1970). It is one of the founders of the national psychology of foreign language teaching and psycholinguistics considers the broad context of Russian language teaching foreigners the psychological characteristics of learning a foreign language is important not only for the practice of teaching the Russian language, but has obviously significant importance for foreign language educational practices in terms of foreign-Russian bilingualism. Actually, foreign language educational practice should be considered any educational practice in which the goals of mastering any language as a foreign language are pursued. Of course, for the situation of teaching Russian language to foreign students need to determine not only the specifics of learning technologies, but the features of the external environment, when the Russian language, being foreign students, is not such for the majority of people around them. This, in fact, is the first psychological feature of the process of teaching the Russian language, which is considered in detail by the author. At the same time, A. A. Leontiev notes, such features of the Russian language as a foreign language act as invariant in relation to the conditions of teaching a foreign language in General, despite their obvious specificity. In other words, the psychological structure of mastering the Russian language as a foreign language, in many ways can be interpreted as a semantic model of the process of mastering a foreign language linguistic experience even in the conditions of Russian-language learning. Despite the absence of a productive language environment, the teaching of Russian-speaking students to a foreign language should, according to the author, take into account the perspective communicative orientation aimed at the fullest possible integration of the student or student communicating in a foreign language into a foreign-language society.

In his later works A. A. Leontiev examines in detail the psycholinguistic essence of the process of mastering a second foreign language. Defining this area of psycholinguistics as applied [Leontiev 1999, 218], the author defines the essence of the concept of "mastering a foreign language". We are interested in the field is characterized as the acquisition of a second language (commonly used in the community in which the developing child), the language of the national linguistic minority, the language of interethnic communication, and in the absence of native speakers in the language environment, mastering a foreign

language. Comprehensively analyzing the content and operational composition of the mechanisms of mastering foreign language linguistic experience, A. A. Leontiev defines teaching a non-native language as teaching speech activity [Leontiev 1999, 218]. In his opinion, teaching a foreign language is nothing but teaching speech activity with the help of a foreign language, specifying, in a later version – it is language acquisition as a teaching of speech communication with the help of this language. In addition, the author points to the presence of at least two more significant and necessary “constituent elements” of the development of foreign language linguistic experience-cognitive and personal aspects of language acquisition. Emphasizing the role of cultural component in mastering a foreign language, A. A. Leontiev speaks about the importance of learning the implementation of orientation in the set of linguistic knowledge and knowledge about the country of the studied language as it is carried out by a native speaker. This is the main task of language acquisition in the cognitive aspect [Leontiev 1999, 218].

The analysis of domestic psycholinguistic sources shows that a number of aspects of teaching a foreign language and, more broadly, the development of foreign language linguistic experience were studied in detail and considered by A. A. Zalevskaya (1973, 1981, 1983, 1996, 2000). Without dwelling on the linguistic aspects of mastering the second and foreign languages, developed in detail in the research of A. A. Zalevskaya, we point to the essential components of the psychological and didactic-psychological structure of its theoretical constructs. First of all, it is necessary to mark the distinction between the concepts of “second language” and “foreign language” verified by the author. In its interpretation, “the second language may not be a foreign language, but it can be studied in educational conditions, and a foreign language is sometimes the third, fourth, etc. language for an individual. Thus, the relation of these concepts is more complex than it seems” [Zalevskaya 2000, 292]. Here, the author defines the specifics of mastering and learning a foreign language – mastering a foreign language occurs in artificial learning situations with a limited grid of hours, but when learning under the guidance of a professional. Differences in the concepts of natural (everyday) and artificial (educational) bilingualism allow A. A. Zalevskaya to speak about the peculiarities of teaching a foreign language – a foreign language is “learned” through volitional efforts and using special methods and techniques. This is the general context of the problem discussed by the author, which explains the existing semantic bias towards linguistic orientation. Referring to the results of a number of studies, the author argues that there are more similarities than differences between the processes of mastering native and foreign languages, the mechanisms of speech activity in native and foreign languages are the same, students undergo similar stages of speech development [Zalevskaya 1996, 5]. In the same context, A. A. Zalevskaya defines the degree of differentiation between the processes of mastering a second and a foreign language, consisting in stipulating specific conditions that must be taken into account when teaching the language, despite the fact that the concepts of second and foreign language are interchangeable. Definition of the General approaches to the psycholinguistic analysis of processes of possession and mastering of the second (foreign) language. Zalewski to an extremely important psychological idea-in the conditions of changes in the goals and content of foreign language teaching, the teacher needs a personal application of a certain concept of learning [Zalevskaya 2000, 18].

The content of teaching foreign languages, according to A. A. Zalevskaya, includes the content of the subject, the content of the subject’s activities on mastering and using a foreign language and the language (speech), socio-cultural and activity components arising from these contents. At the same time, the latter assumes the presence of skills, abilities, strategies and supporting elements of mastering and teaching a foreign language [Zalevskaya 1996, 76]. Special attention the author receives the fact of the overall relationship of all these elements of the content of foreign language teaching, despite the fact that speech and socio-cultural components act simultaneously as the products of scientific analysis appearing in the content of the subject, and how the required property of the individual student through its activities on mastering and using a foreign language.

It is obvious that the expanding trend in the field of psychology of learning and mastering a foreign language involves the use of these marginal disciplines – from psycholinguistics, as the closest to the psychology of the sphere, to methods and pedagogy, representing not so distant from her field of scientific thought. In General, the approach formulated by A. A. Zalevsky’s approach to learning and mastering a foreign language, with reference to the problems of language and metalanguage knowledge, “language consciousness” and “sense of language” encourages, according to the author, to go beyond the methodology of language teaching in the field of psychology and psycholinguistics [Zalevskaya 1996, 85]. In addition, a special mention requires, in our opinion, another generalization of AA. Zalevskaya: most scientific publications in the field of psychology of teaching foreign languages by the broad masses of foreign language teachers are not sufficiently understood and evaluated as a guide to action in both theoretical and practical terms (one of the reasons for this is the lack of a mandatory course of psychology of teaching a foreign

language at the relevant faculties of universities) [Zalevskaya 1996, 112]. On the one hand, it is difficult to disagree with this conclusion, on the other-it should be interpreted not so much as the statement of a psycholinguist, but as a domestic teacher of a foreign language with experience of theoretical and experimental work in the field of psycholinguistics.

In addition to the actual research approaches, in the psycholinguistic scientific and educational literature is very widely discussed psycho-psycholinguistic problems of learning and mastering the second (foreign) language. However, due to the specificity of the genre of these educational and scientific publications (a qualified review of development trends and discussion-staged interpretation of the main theoretical provisions of this field of knowledge), they do not contain a new understanding of the essence of psychological processes of teaching a foreign language, and therefore, are of interest to us only as a general scientific fact.

#### **References**

Zalevskaya A. A. Introduction to psycholinguistics. – M.: Russian. state of gum-bat. UNT, 2000. – 382 p.

Zalevskaya A. A. Questions of the theory of mastering the second language in the psycholinguistic aspect. – Tver: Tver State University, 1996. – 196 p.

Leontiev A. A. Psycholinguistics in language acquisition. Fundamentals of psycholinguistics. – M.: Sense, 1999. – P. 218–228.

Leontiev A. A. Psychological bases of visibility in the textbook of a foreign language. Language and speech activity in General and pedagogical psychology: Selected psychological works. – Moscow: Moscow psychological and social Institute; Voronezh: NGO «MODEK», 2001. – P. 283–292.

## ОБ ИСТОРИИ ТЮРКСКОЙ ФОНОЛОГИИ И МОРФОЛОГИИ

Морфонология определяется как область, промежуточная между морфологией и фонологией, главное функциональное назначение которой состоит в организации экспонента морфемы. Существует множество взглядов относительно тюркской фонологии и морфонологии. В статье обсуждается история формирования тюркской фонологии и морфонологии.

**Ключевые слова:** тюркские языки, фонология, морфонология, фонология.

В 20-30-е годы XX века когда интенсивно разрабатывались вопросы звукового и грамматического строя в связи с созданием графики и орфографии тюркских языков, тюркская фонеморфология, опираясь на собственную почву, смогла не только обнаружить и описать исконные черты строя родственных языков, но и выработать основополагающие идеи, которые впоследствии так или иначе послужили толчком для развития некоторых фонологических направлений, в частности, Лондонской фонологической школы просодического анализа [Лайонз 1977]. Это связано с именами целого поколения исследователей и прежде всего ученых-просветителей А. Байтурсунова, И. Арабаева, К. Жубанова и К. Тыныстанова, ставших жертвами репрессий 1937 года. Этим ученым не суждено было представить целостное изложение своих взглядов на проблемы фонологии и морфонологии тюркских языков. Тем не менее их научное наследие и практические решения вопросов графики и орфографии пронизаны идеями сингармонизма [Арабайуулу, Сарсекеуулу; Байтурсуноулу 1992; Тыныстанов 1934 (1); Тыныстанов 1928].

У А. Байтурсунова, исследовавшего сингармонизм со стороны слова, превалировала идея о неразрывном единстве слова и сингармонизма и, следовательно, невозможности существования тюркского (казахского) слова вне сингармонизма. Он глубоко понимал, что сфера действия сингармонизма носит глобальный характер и в равной мере распространяется как на составляющие слова, так и на слово в целом. Практическое воплощение этой идеи осуществлено в «талантливом проекте казахской орфографии», где каждое слово характеризовалось наличием или отсутствием особой сингармонической пометы [Реформатский 1966, 119]. Эта идея, как уже отмечалось, успешно развивается в Лондонской фонологической школе просодического анализа.

Аналогичная идея прослеживается и в трудах К. Тыныстанова [Тыныстанов 1934 (1); Тыныстанов 1934 (2). Тыныстанов 1934 (3); Тыныстанов 1928]. В отличие от А. Байтурсунова, он уделяет внимание на раскрытие механизма реализации сингармонизма. Это стало возможным благодаря созданию логических основ исчисления словоформ путем комбинации слоговых типов как минимальных носителей сингармонизма. Он обосновывал положение, что сингармоническое воздействие первого слога на третий осуществляется через посредство второго. По-видимому, именно это дает основание Е. Д. Поливанову считать К. Тыныстанова одним из основателей теории морфонологии: «И тот факт, что т. Тыныстанов не просто прислушался к последнему слову западной науки, а наоборот – самостоятельно пошел по параллельному пути, создав собственную морфонологию киргизского языка (хотя и не называемую им этим именем), уже сам себе достаточно говорит о значительности достижений киргизского лингвиста. Этим я вовсе не хочу сказать, однако, что все вопросы морфонологического характера уже разрешены для киргизского языка Тыныстановым. Но во всяком случае сделано им уже достаточно для того, чтобы практические вопросы орфографии могли решаться на твердой базе теоретического материала» [Поливанов, Тыныстанов]. Более того, К. Тыныстанов по праву мог бы называться и основателем тюркской генеративной лингвистики. Разумеется, раскрытие подлинного значения научного наследия К. Тыныстанова, равно как и А. Байтурсунова, – это дело будущего [Садыков 1986; Садыков 1992; Садыков 2009; Садыков 2011].

Источником многих недоразумений в тюркской фонологии является трактовка звукового состава корня и аффикса. Эта антиномия одной из первых была отмечена И. А. Богородицким: «Вокализм корня представляет одну существенную особенность сравнительно с вокализмом придаточных морфем: в нем гласные выступают каждый в своей индивидуальности без чередования с другими гласными, тогда как в придаточных морфемах гласные ограничены двумя основными типами – либо гласный широкий, либо узкий, видоизменяющиеся сообразно с вокализмом корня в отношении палатальности или непалатальности, а также губности или ее отсутствия» [Богородицкий 1954: 117]. Эта антиномия была отмечена также Н. Трубецким [Трубецкой 2000, 113]. Сам Трубецкой наличие этой антиномии склонен был видеть только в вокализме, а именно: в вокализме корня различие всех

восьми гласных, а в вокализме аффикса – нейтрализацию таких гласных по признакам тембра и огубленности: «В тех тюркских языках, где последовательно осуществляется так называемая гармония гласных, приведенная выше система наличествует в полном объеме, то есть фонологически значимом виде, лишь в первом слоге слова; во всех прочих слогах тембровые противоположения нейтрализуются, и тембровые признаки гласных непервого слога в своей реализации равняются на гласные предшествующего слога» [Трубецкой 2000, 113]. На этом основании для вокализма аффикса он постулировал наличие двух архифонем, различающихся по признаку подъема и не различающихся по остальным двум признакам. Это решение, однако, входит в противоречие с определением архифонемы, сформулированным самим же Трубецким: об архифонеме можно говорить только применительно к одномерным оппозициям, в пределах которых две фонемы, различающиеся одним фонологически существенным признаком, перестают различаться [Трубецкой 2000, 84].

Существующие решения этого вопроса на самом деле являются разновидностями трактовки Н. Трубецкого. В этих решениях, как и у Трубецкого, речь идет в основном только о вокализме [Лайонез 1987, 10–11]. Что касается аспекта проблемы, который связан с наличием мены согласных на стыке морфем, совершенно выпал из поля зрения исследователей.

Когда обнаруживаются факты, противоречащие теории, по-видимому, существуют два пути выхода из создавшейся ситуации. Первый путь – отбросить старую теорию, создав взамен ее новую. В ракурсе обсуждаемых проблем применительно к казахскому языку, это было сделано А. Жунисбеком, изложившим основные положения теории сингармонической фонологии или сингармонологии в двух своих монографиях [Жунисбек 2009]. Второй путь – предложить новую версию существующих теорий, приводя ее в соответствие с фактами языка, что и было сделано в наших исследованиях.

Суть основных положений нашей концепции сводится к следующим позициям:

1. Тюркские языки обладают автономными системами как фонем, так и морфем. При этом фонемы и морфемы выступают как изоморфные по фреймструктуре единицы, главное функциональное назначение которых заключается в том, чтобы быть компонентом исходной формы морфемы.

2. Базой для формирования фонемы служат корневые морфемы, а для образования, морфемы – аффиксальные. Звуковая сторона корневых морфем задает программу для функционирования аффиксальных морфем и тем самым регламентирует сферу действия источников алломорфии аффикса, прежде всего – явлений сингармонизма, ассимиляции, диссимиляции и др.

3. Звуковая сторона аффиксальных морфем является внутрисистемным отражением, представляющим модель звуковой стороны, следовательно, систему фонем корневых морфем. Корневые фонемы обычно бывают позиционно закрепленными и контекстуально необусловленными, а аффиксальные фонемы, будучи распространителями определенных субстанциональных свойств корневых фонем, – позиционно закрепленными, но контекстуально обусловленными. Свойства корневых фонем делают их относительно устойчивыми, свойства аффиксальных – относительно изменчивыми, но только в строгих рамках сингармонизма и чередующихся в определенной позиции фонем. Отбор и распределение участников чередования происходит не случайно, а целенаправленно, исходя из принципа наличия субстанциональной общности в их признаковом содержании. В частности, согласные отбираются чаще всего по гоморганности (ср. чередования гоморганных согласных типа  $n\sim d\sim t$ ,  $l\sim d\sim t$  в кыргызском,  $n\sim d\sim t\sim z$ ,  $l\sim d\sim t\sim z$  в башкирском языках), реже – по гомогенности (ср. исторические соответствия гомогенных согласных типа  $l\sim ш$ ,  $p\sim з$ ), гласные – по раствору (ср. распределение аффиксальных гласных по раствору в кыргызском языке:  $y\sim u\sim y\sim \ddot{y}$ ,  $a\sim e\sim o\sim \ddot{o}$ ), реже – по огубленности (ср. распределение аффиксальных долгих гласных по огубленности в кыргызском языке:  $yy\sim \ddot{y}\ddot{y}$ ,  $oo\sim \ddot{o}\ddot{o}$ ).

4. Наличие альтернантов фонем, участвующих в чередовании в аффиксе, делает актуальной задачу установления тождества и различия таких фонем. Это в высокой степени способствует структуризации звукового пространства как самих аффиксов, так и корней и, следовательно, идентификации дифференциальных признаков системы фонем в целом. При этом признаки, общие для чередующихся фонем, выступают в качестве основания для их отождествления (например, для гласных  $y\sim u\sim y\sim \ddot{y}$  отождествляющим является верхний подъем, для гласных  $a\sim e\sim o\sim \ddot{o}$  идентифицирующим – нижний подъем), а признаки, индивидуальные для альтернантов, – для их различения (например, в пределах гласных  $y\sim u\sim y\sim \ddot{y}$ ,  $a\sim e\sim o\sim \ddot{o}$  различительными являются признаки огубленности и тональности).

5. Адаптированность свойств аффиксальных фонем к свойствам корневых фонем обеспечивает системе в целом выигрыш в том, что и те, и другие становятся субстанционально однородными по линии их признакового содержания. Однако такая однородность компенсируется неоднородностью распределения функций фонем и их ДП в корнях и аффиксах.

6. Корневые фонемы и их ДП имеют все возможности для полной реализации своих функций, в том числе и функции смыслоразличения (ср. Корни типа *сыз* ‘сырость’; ‘чертить’, *сыз* ‘вы’, *суз* ‘хму-



рый; черпать’, *суз* ‘бодать, цедить, плавать’, в составе которых полностью реализованы ДП всех фонем, как согласных, так и гласных), а у аффиксальных эти возможности ограничены (ср. четыре сингармонические варианты аффикса лишения типа *-сыз/-сиз/-суз/-сүз*, в контексте которых смысловозначительную функцию у гласных имеет только признак подъема).

Часть признаков у корневых фонем выполняют еще и дополнительную, кванторную функцию, указывающую на принадлежность корня к определенному виду сингармонизма. Таковы для корневых фонем сингармонически активные признаки тональности и огубленности:

	<i>ат</i>	<i>ит</i>	<i>уй</i>	<i>үй</i>
<b>Ы</b>	—	Х	—	Х
<b>И</b>	Х	—	Х	—
<b>У</b>	Х	Х	—	—
<b>А</b>	—	—	Х	Х

где *ы*, *и*, *у*, *а* – соответственно кванторы веллярного (твердого тона), палатального (мягкого тона), огубленного и не огубленного видов сингармонизма. Сплошная линия под слогом показывает на сферу действия вида, а крестики – отсутствие действия данного вида. При этом тональные кванторы являются основными видами сингармонизма, тогда как резонансные по огубленности кванторы – дополнительными видами, которые сопровождают одного из основных видов: либо веллярный вид, либо палатальный вид.

Как видим, корневые основы *ат* ‘лошадь’, *уй* ‘корова’ произносятся одинаково твердым тоном, но у последней при сопровождении огубленного резонанса, а у первой – неогубленного резонанса. Корневые основы *ит* ‘собака’, *үй* ‘дом’ отличаются от них только мягкотональным произношением при совпадении резонансных видов. Следовательно, у корневых фонем признаки тональности и огубленности одновременно выполняют как функцию смысловозначения, так и кванторную функцию сингармонического оформления корня.

В аффиксах признаки, выполняющие кванторносингармонические функции, отделены от признаков, выполняющих смысловозначительную функцию. Кванторносингармоническая функция в аффиксальных фонемах наличествует только в потенции и актуализируется в контексте словоформы в результате наложения кванторов сингармонизма корня на аффикс. Ср. схему сингармонического воздействия рассмотренных корней на показатели *-сыз/-сиз/-суз/-сүз* аффикса лишения в контексте сингармонически цельнооформленных производных *ат-сыз* ‘безлошадный’, *иш-сиз* ‘безработный’, *уй-суз* ‘тот, кто не имеет корову’, *үй-сүз* ‘бездомный’:

	<i>ат-сыз</i>	<i>иш-сиз</i>	<i>уй-суз</i>	<i>үй-сүз</i>
<b>Ы</b>	—	Х Х	—	Х Х
<b>И</b>	Х Х	—	Х Х	—
<b>У</b>	Х Х	Х Х	—	—
<b>А</b>	—	—	Х Х	Х Х

Отмежевание смысловозначительной функции от других функций происходит для аффиксальных фонем и в позиции чередующихся согласных. Последние функции вместе с их носителями (признаками) также наличествует в потенции и могут актуализироваться в контексте словоформы в результате действия явлений ассимиляции и диссимиляции на стыке корня и аффикса.

7. Функциональная неравнозначность корневых и аффиксальных фонем, с одной стороны, и спецификация признаков фонем по линии разграничения их смысловозначительной и несмысловозначительной функций, с другой, в конечном итоге ставит перед тюркской системой задачу формирования дополнительной единицы для аффиксов-морфемы, которая могла бы выступать, как и фонема, в качестве компонента морфемы, но уже не корневой, а аффиксальной, притом только в позиции чередующихся фонем. Наличие автономной системы морфем позволяет тюркским языкам придать максимальную гибкость в использовании автономной системы фонем как в корнях, так и в аффиксах.

Различие между фонемой и морфемой состоит в следующем:

1) в пределах исходной формы морфемы фонема характеризуется только заполненными валентностями, т.е. полным набором ее ДП, а морфема – как заполненными, так и пустыми валентностями.

Заполненные валентности у морфемы формируются набором ДП, общих для чередующихся фонем, а незаполненные, пустые валентности – набором ДП, индивидуальных для функционально сильной фонемы в ряду чередующихся альтернантов;

2) в пределах алломорфа морфемы путем реализации пустых валентностей происходит выравнивание признакового содержания морфемы и фонемы. Этим самым осуществляется переход морфемы в функционально сильную фонему, представляющую позицию чередующихся фонем. На этом основании разрешается фундаментальная для тюркских языков антиномия фонологии корня и фонологии аффикса, которая была отмечена Н.Трубецким;

3) в пределах алломорфа морфемы фонема обретает еще дополнительные валентности, при заполнении которых она характеризуется максимальным набором, включающим как ДП, так ИП. Заполнение пустых валентностей фонемы в этом случае происходит в результате воздействия контекстов как оси одновременности, так и оси последовательности. Для морфемы как таковой оно осуществляется вне зависимости от условий контекстного окружения, но по управляющему сигналу, поступающему от вышестоящих уровней системы, т.е. от уровня морфем, слов и др.

В-третьих, результаты наших исследований можно рассматривать и в морфонологическом ключе. В этом случае во весь рост встает круг проблем, связанных с определением объекта морфонологических исследований и выявлением природы и функции морфонологических единиц и явлений.

Исследование показало, что из трех выделенных Н.Трубецким разделов общей морфонологической теории [Трубецкой 1987, 116-117] приемлемыми для тюркских языков являются только первые два, но при этом надо сделать следующие оговорки:

1. *Теория фонологической структуры морфемы.* Как было показано в [Садыков 1992], существующие фонологические концепции, в том числе и концепция Н.Трубецкого, не отвечают в полной мере своеобразию строя тюркских языков. Следовательно, здесь, во-первых, речь должна идти не о фонологической, а морфонологической структуре морфем, во-вторых, под морфонологической следует понимать структуру исходной формы морфемы, компонентами которой могут быть не только фонемы, но и морфемы (см. выше), в-третьих, исходную форму морфемы следует отличать от ее основного варианта: первая выступает в роли двойного означающего, как значения морфемы, так и ее основного варианта, тогда как основной вариант – только реально зафиксированных вариантов. При этом исходная форма, выступающая в качестве инварианта всех реально зафиксированных вариантов, определяется путем синхронической реконструкции архетипа морфемы.

С другой стороны, компонентами основного варианта могут быть только фонемы и притом функционально сильные для данной позиции, в которой фонемы представлены полным набором как ДП, так и ИП. А компонентами исходной формы могут быть как фонемы, так и морфемы, при том, что фонемы здесь представлены только минимальным набором, содержащим чистые ДП. При этом существует только один путь для перехода от исходной формы к основному варианту морфемы и наоборот, т.е. переход от первой к конкретному варианту опосредован переходом от нее к основному варианту.

2. *Теория комбинаторных звуковых изменений,* которым подвергаются отдельные морфемы в морфемных сочетаниях. Это положение Н.Трубецкого приемлемо для тюркских языков в следующей редакции: *теория порождения алломорфов морфемы.* Эта формулировка предполагает включение всех источников алломорфии, а именно: явлений сингармонизма, ассимиляции, диссимиляции, элизии (опущение фонем), аугментации (добавления фонем), метатезы (перестановки фонем) и др. Таким образом, все источники алломорфии тюркских морфем целиком относятся к морфонологическим явлениям, которые оперируют только фонемами, т.е. единицами автономной системы фонем того или иного тюркского языка.

3. *Теория звуковых чередований,* выполняющих морфонологическую функцию. Если эту формулировку Н.Трубецкого понимать как внутреннюю флексию, характерную для флективных языков, то очевидно нет особой необходимости для ее распространения на тюркскую почву. Однако не исключена возможность, что такая потребность может возникнуть при разработке проблем, например, диахронической морфонологии. Если же рассматривать ее как чередование, выполняющее функцию вхождения морфем в словоформу по виду сингармонизма, то она автоматически попадает в компетенцию предыдущего пункта.

Итак, в нашей концепции основной объект морфонологии представлен в двуединстве статики и динамики. Это вытекает из следующих позиций:

1) главная цель морфонологии заключается в том, чтобы установить инвариант вариантов морфемы и определить набор правил, обеспечивающих порождение по инварианту всех ее вариантов. Таким инвариантом является исходная форма морфемы, компонентами которой могут быть не только

фонемы, но и морфемы. С синхронической точки зрения исходная форма морфемы представляет собой статический объект системы и динамический – процесса порождения речи (текста);

2) с диахронической точки зрения как исходная форма (архетип) морфемы, так и процесс ее трансформации представлены в динамике. Следовательно, синхронический архетип морфемы может совпадать или не совпадать с ее диахроническим архетипом. Механический перенос данных диахронии в синхронию здесь исключается, хотя ясно, что синхрония в скрытом виде воспроизводит диахронические моменты развития языков;

3) совпадение данных синхронии и диахронии может иметь на уровне алломорфов морфемы, т.е. тогда, когда речь идет о тождестве одного из вариантов, чаще всего – основного варианта, с ее проформой.

Таким образом, критерии определения объекта морфонологии и установления ее границ исходят из самого устройства тюркского слова и составляющих морфем. Эти критерии базируются на том, как и на базе чего происходит порождение тюркской словоформы, и гласят: основным объектом морфонологии являются исходная форма морфемы и так или иначе соотнесенные с ней ее варианты, а также правила, по которым от исходной формы происходит порождение всех вариантов.

Правила порождения вариантов морфемы – это суть действия источников алломорфии, т.е. таких, как явление сингармонизма, ассимиляции, диссимиляции, элизии, аугментации, метатезы, суперпозиции и др. Следовательно, и они подпадают в рубрику объекта морфонологии.

Правила, определяющие сферу действия источников алломорфии, изначально являются морфонологическими в силу закреплённости их за определенной позицией морфемы. Однако они могут обуславливаться разными контекстуальными условиями. Большая часть из них реализуется лишь в условиях фонологического контекста. Таковы правила чередования фонем по законам сингармонизма и стыка морфем.

Другая часть правил характеризуется тем, что для их реализации, кроме фонологического контекста, требуется наличие и определенных морфологических или лексических условий, т.е. сфера их действия распространяется только на часть определенных морфем или лексем. Таковы, в частности, правила выбора одного из синонимичных показателей каузатива, правила опущения, добавления, перестановки фонем, правила, определяющие тип склонения именных основ и тип спряжения глагольных лексем.

Различие морфонологических правил может выражаться и по линии их соотнесенности с морфемой. Как уже отмечалось, наличием такой соотнесенности характеризуется только первая из них, т.е. те, которые действуют в позиции чередования. Поскольку морфема может выделяться только в позиции, где происходит чередование либо коррелятивных фонем, либо других фонем рядом с коррелятивными, соотнесенность этих правил с морфемой не прямая, а косвенная. Это объясняется с тем, что для перехода морфемы в функционально сильную фонему необходимо использовать другое правило, которое имеет контекстуально независимый характер своего применения. Следовательно, этим и отличаются морфемные правила от всех прочих правил.

Тюркские языки могут характеризоваться наличием как сходств, так и различий в сфере использования морфонологических правил. Сходства и различия этих правил могут проявляться как по линии результатов их действия, так и в контекстуальных условиях их реализации. Примером для демонстрации этих особенностей могут служить, в частности, правила, определяющие типы склонения именных основ в тюркских языках [Абдувалиев, Садыков 1997, 71-77, 132, 136-137, 144-145, 151-154, 156-157; 13 Садыков 1986].

Тюркские языки знают только два типа склонения: тривиальный и нетривиальный типы:

Падеж	Тип склонения	
	Тривиальный	Нетривиальный
Основной	Ø	Ø
Родительный	-нын	-ын
Винительный	-ны	-ы (>Ø)
Дательный	-га	-а
Местный	-да	-да
Исходный	-дан	-дан (>-ан)

Единственным критерием, определяющим тип склонения именных основ, служит наличие или отсутствие начального согласного у показателей род., вин. и дат. падежей. Привлечение других критериев, основанных на учете возможных изменений в основе и дополнительных изменений в аффиксе,

представляется в этом случае теоретически несостоятельным в силу того, что это приводит к нарушению единого логического основания, используемого при выделении данных типов. Однако такая информация полезна для установления подтипов внутри каждого типа.

Правило тривиального типа сохраняет анлаут у падежного аффикса, а правило нетривиального типа его лишает. Эти морфонологические правила являются общими почти для всех тюркских языков. Исключение могут составлять лишь чувашский и, возможно, якутский языки. Различие тюркских языков здесь проявляется только в условиях реализации этих правил, т.е. в том, какие именно основы попадают под действие этих правил.

Унификацию условий реализации типов склонения обнаруживают огузские языки и в особенности турецкий язык. В последнем правило тривиального склонения реализуется только после основ на гласный, а правило нетривиального – после основ на согласный. Так, например, склонение основ *куту*~ *kutu* ‘коробка’, *кыз*~ *kiz* ‘девочка’, *сен*~*sen* ‘ты’ в кыргызском и турецком языках имеет следующую парадигму:

О.	куту-Ø	kutu-Ø	кыз-Ø	kiz-Ø	сен-Ø	sen-Ø
Р.	куту-нун	kutu- nun	кыз-дын	kiz-in	сен-ин	sen-in
В.	куту-ну	kutu-yu	кыз-ды	kiz-i	сен-и	sen-i
Д.	куту-га	kutu-ya	кыз-га	kiz-a	саг-а	san-a
М.	куту-да	kutu-da	кыз-да	kiz-da	сен-де	sen-de
И.	куту-дан	kutu-dan	кыз-дан	kiz-dan	сен-ден	sen-den

Эти данные как будто свидетельствуют о том, что кыргызский и турецкий языки сильно расходятся в рассматриваемой сфере. Однако это совсем не так. Эти расхождения имеют поверхностный характер и вызваны они особенностями распределения правил, определяющих тип склонения. Глубинные сходства, обусловленные генетическими связями двух языков, проявляются с большей рельефностью в местоименных парадигмах, в частности, в парадигме склонения местоимения *сен*~*sen*. Если сравнить парадигмы склонения данного местоимения, то невозможно не заметить, что и в том, и в другом языке падежные показатели полностью совпадают. А именно: род., вин. и дат. падежи имеют показатели, лишенные анлуата. Ср. кырг. род. -*ин*, вин. -*и*, дат. -*а*; тур. род. -*ин*, вин. -*и*, дат. -*а*. Наличие этих показателей, тождественных для род., вин. и дат. падежей, регламентировано одновременным действием правила нетривиального склонения в обоих языках.

Расхождение наблюдается лишь в одном случае: в кыргызском при дат.падеже местоименная основа *сен* представлена в варианте *саг-а*, тогда как турецкая основа *sen* – в варианте *сан-а*. Ясно, что преобразование *сен*-в *саг*-обусловлено наличием ауслатного перехода *н >ң>* в местоименных корнях. Ср.: *сен-га >саң-га > саг-га > саг-а*. Правомерность такой сегментации верифицируется материалом хакасского языка, в котором у личных местоимений при дат. падеже конечный согласный основы переходит, как и в кыргызском, в *г*-, а у падежного показателя, в отличие от кыргызского, путем стяжения анлуата аффикса образуется долгий гласный. Ср. хака. *маг-аа, саг-аа, аг-аа* и др. [Грамматика хакасского языка 1975, 146–156]. Поразительно, но факт, что это доказывается также и данными Жебраильского, Нахичеванского, Закатал-Кахского, Кедабекского и Карягийского говоров азербайджанского языка, в которых дат. падеж личных местоимений выступает в следующих формах: *magha*<*mag + ha, sagha*<*sag + ha, agha*<*ag + ha* и др. [Сравнительно-историческая грамматика тюркских языков 2002, 148-149]. Надо полагать, что эти факты вместе с карачаево-балкарскими формами типа *маңа, саңа* [Тыныстанов 1928, 83] важны не только с точки зрения принятия данного решения, но и с точки зрения историко-лингвистического обоснования миграции древнетюркских племен в Евразийском пространстве.

Таким образом, кыргызский и турецкий языки в одинаковой мере характеризуются наличием идентичных правил, реализующих типы склонения именных основ. Их различие проявляется только в условиях реализации этих правил: в турецком они действуют только в фонологическом контексте, в кыргызском – как в морфологическом, так и в лексическом контекстах.

В других тюркских языках наблюдается примерно та же картина, что и в кыргызском. Они различаются между собой не количеством типов склонения, а условиями, допускающими действия правил нетривиального склонения. В одних языках сфера действия этих правил сужается. Например, в узбекском под действие правил нетривиального склонения подпадают лишь некоторые личные и указательные местоимения. А в других она расширяется, как, например, в кыргызском, кроме личных и указательных местоимений ед. числа, сюда попадают и основы на -*ым, -иң, -ы/-сы*, а также основы на аффикс -*ныкы*, собирательные числительные на аффикс -*оо* и др. Характерно, что во всех тюркских языках организующим центром для нетривиального склонения является местоименная па-

радика, а для тривиального склонения – простая именная парадигма. Все остальные именные группы распределяются вокруг этих двух центров. Часть из них может иметь в склонении характер тенденции и колебаться между тривиальными и нетривиальными типами. Задача морфонологии здесь заключается в том, чтобы правильно выявить эти тенденции и предложить соответствующие рекомендации для формирования кодифицированной нормы языка.

В заключении, следует сказать несколько слов о значении результатов наших исследований для компьютеризации лингвистических исследований в тюркологии вообще и в кыргызском языкознании в частности. Эти результаты имеют прямое отношение к решению данной глобальной проблемы и могут использоваться в следующих направлениях;

- 1) для создания электронной базы тюркских языков как многоцелевой системы по комплексной автоматизации лингвистических исследований и инженерных разработок,
- 2) для создания программы лемматизации текстовых словоформ в реестровые,
- 3) для создания систем анализа и синтеза звучащей речи в качестве лингвистической базы,
- 4) для создания систем машинного перевода с тюркских или на тюркские языки,
- 5) для создания обучающихся программ со входными тюркскими языками др.

Если верно, что тюркские языки обладают уникальным устройством и характерным для него логически четким основанием, то благодаря компьютеризации тюркское языкознание в ближайшие годы могло бы претендовать на передовые позиции в мировой лингвистике.

### Литература

- Абдувалиев И., Садыков Т. Азыркыкыргыз тили: морфология. – Бишкек: Сорос-Кыргызстан, 1997.
- Арабайуулу И., Сарсекеуулу Х. Алифба йаки төтө окуу. – Уфа, 1911.
- Байтұрсынұлы А. Тілтаглымы. – Алматы, 1992.
- Богородицкий В.А. Введение в татарское языкознание в связи с другими тюркскими языками. – Казань, 1953.
- Грамматика хакасского языка. – М.: Наука, 1975.
- Жунисбек А. Введение в сингармоническую фонетику. – Алматы, 2009.
- Жұбанов Қ. Қазақ тілі жөніндегі зерттеулер. – Алматы, 2010.
- Лайонз Дж. Введение в теоретическую лингвистику. – М.: Прогресс, 1977.
- Поливанов Е.Д. Тыныстанов К. Проект новой орфографии киргизского литературного языка (рецензия) // Рук.фонд НАН КР. Инв. № 124 (18).
- Реформатский А.А. Иерархия фонологических единиц и явления сингармонизма // Исследования по фонологии. – М.: Наука, 1966.
- Реформатский А.А. О соотношении фонетики и грамматики // Вопросы грамматического строя. – М.: Наука, 1955.
- Садыков Т. Муунүндөштүгүжөнүндө // Қазақтілбілімі: тарих, тәжірибе, болашақ. – Алматы, 2011, 141–145 б.
- Садыков Т. О сегментации падежных форм местоимений в кыргызском языке // Тюркологические исследования. – Бишкек: Илим, 1986.
- Садыков Т. Основы кыргызской фонологии и морфологии. – Бишкек: Илим, 1992.
- Садыков Т. Профессор К.Жубанов илими жана морфостатистика жөнүндө // Қазақ тіл білімінің өзекті маселелері. Проф. Құдайберген Жұбановтың 110 жылдыгына арналған халықаралық ғылыми-теориялық конференция материалдары. – Алматы, 2009, 10–13 б.
- Садыков Т., Шаршембаев Б. Кыргыз тилтаанымында // Эпос тілінзерттеудіңөзектімаселелері. Проф. Колдонмо лингвистиканын башаттары Есет Жұбановтың туғанына 80 жыл толуына арналған халықаралық дөңгелек үстел материалдары. – Алматы, 2011, 226–229 б.
- Сравнительно-историческая грамматика тюркских языков. Региональная реконструкция. – М, 2002.
- Трубецкой Н. Некоторые соображения относительно морфонологии // Пражский лингвистический кружок. – М.: Наука, 1987.
- Трубецкой Н. Основы фонологии. – М., 2000.
- Тыныстанов К. Кыргыз адабий тилинин жаңы ымласынын долбору. – Фрунзе, 1934 (1).
- Тыныстанов К. Кыргыз тили. Морфология. – Фрунзе, 1934 (2).
- Тыныстанов К. Лексикон киргизского языка. – Фрунзе: 1934 (3). // Рук.фонд НАН КР. Инв. 177/37-182/42; 79–84.
- Тыныстанов К. Энетилибиз. – Фрунзе, 1928.
- Языки мира: тюркские языки. – Бишкек, 1997.

## ПРАВОПИСАНИЕ ЗАИМСТВОВАННЫХ СЛОВ В БАШКИРСКОМ ЯЗЫКЕ

Статья посвящена изучению вопросов правописания заимствованных слов в башкирском языке. Авторами изучены случаи аффиксации русских заимствований в башкирском языке. Анализу подвергнуты слова с конечным 'ь' и слова, имеющие в конце два или несколько согласных звуков.

**Ключевые слова:** башкирское языкознание, заимствованная лексика, орфография, аффиксация.

Правописание заимствованных слов представляет собой специфическую область орфографии каждого языка. Во всех национальных языках России основную долю заимствований составляют русские заимствования. В их число входят заимствования из европейских языков (слова английского, немецкого, французского и др. языков), которые перешли через русский язык. Кроме них, имеются древние заимствования из арабского и персидского языков. По подсчетам Г. Г. Кагарманова, если все заимствования в башкирском языке составляют 37 %, то 23,2 % из них – это русские заимствования, 11,4% – арабизмы, 2,5% – персидские заимствования [Кагарманов 2002, 52].

Русские заимствования, вошедшие в башкирский язык через письменную речь, можно разделить на несколько групп:

– слова, которые полностью сохранили звуко-буквенные характеристики, имеющиеся в русском языке: *документ, парта, компьютер, интернет, схема, троллейбус, слюда, стена, судья, тахта, кафедра, кофта, сумка, ручка, салфетка, тумбочка, люстра и т. д.*;

– слова, утратившие последнюю букву (как правило, гласный звук): *цифра – цифр, минута – минут, папироса – папирос, сигарета – сигарет, котлета – котлет, гардина – гардин, каникулы – каникул, галоши – галош, макароны – макарон, нервы – нерв и т. д.*;

– слова, изменившие последнюю букву (как правило, гласный звук): *консервы – консерва, ботинки – ботинка, тапочки – тапочка и т. д.*

В данной статье рассмотрены случаи аффиксации русских заимствований, имеющих в ауслауте 'ь', а также слова, имеющие в конце сочетание двух или более согласных звуков. В орфографии 1981 г. их правописание регламентировано в нескольких параграфах. Рассмотрим их.

§35. В. 1. а) если слово оканчивается на -ль, -нь или на два согласных с мягким знаком (-рзь, -фть, -рмь, -рфь – добавлено Х.Л.), то к этим словам прибавляется аффикс с гласными переднего ряда.

б) если слово оканчивается на -рь, -сь, -чь, -шь, то к этим словам прибавляется аффикс с гласными заднего ряда.

При присоединении аффиксов -ь опускается.

Так, согласно правилам, слова, оканчивающиеся на -ль, -нь, присоединяют аффикс с гласным переднего ряда, а слова, оканчивающиеся на -рь, -сь, -чь, -шь, -сть, присоединяют аффикс с гласным заднего ряда.

Вне поля зрения остались слова, оканчивающиеся на -зь (мазь), -кль (спектакль), -мль (кремль), кроме того, среди современных заимствований могут быть примеры на -пь, -ть и т. д.

По данным орфографического словаря 1998 г., к слову 'мазь' прибавлены аффиксы с гласным заднего ряда (мазь, маздың (форма род. п.), мазға (на мазь), маздан (форма исх. п.), мазы(на) (на его мазь), к словам 'спектакль', 'кремль' – аффиксы с гласным переднего ряда (спектаклдең (форма род. п.), спектаклгә (на спектакль), спектаклдән (из спектакля), спектакле(нә) (на его спектакль); кремлдең (кремля), кремлгә (кремлю), кремлдән (из кремля), кремле(нә) (на его кремль).

В результате в разных подпунктах оказались похожие примеры:

аффиксы с гласными заднего ряда	аффиксы с гласными переднего ряда
мазь	ферзь
киноварь, календарь, аптекарь, токарь, слесарь, писарь, вратарь, секретарь, инвентарь, кустарь, рыцарь, лагерь, егерь, панцирь, якорь, глазурь, пластырь, монастырь	декабрь, ноябрь, октябрь, сентябрь
власть, ведомость, волость, крепость, промышленность, повесть, санчасть,	нефть, юфть

В пункте F. 1 дано правило, согласно которому в заимствованных словах с конечным -ск, -нг, -нк между корнем слова и аффиксом ставится гласный звук (это явление в башкирском языке получило название эпитезы): отпуск – отпускыла (в отпуске), диск – дискины (форма вин. п.), пропуск – пропускыға (в пропуске); лозунг – лозунгылар, лозунгыһы (лозунги, его лозунг); митинг – митингыға, митингыһы (на митинг, его митинг); танк – танкыны, танкыһы (форма вин. п.; его танк); банк – банкыла, банкыһы (в банке, его банк).

Слова, оканчивающиеся на -рк, -лк, -рг под это правило не попадают.

в инлауте ставится -ы-	аффиксы присоединяются к корню слова
корень слова оканчивается на сонорный звук+к	
банк, бланк, франк, танк, эвенк, цинк	а) парк, зоопарк, фейерверк, очерк, почерк, цирк, б) полк, тальк
корень слова оканчивается на сонорный звук+г	
фланг, шланг, ранг, орангутанг, тюбинг, крекинг, шиллинг, блюминг, браунинг, демпинг, ринг, митинг, дифтонг, лозунг, тунг	комсорг, парторг, металлург, хирург, драматург, профорг
корень слова оканчивается на глухой звук+к	
диск, прииск, обелиск, оттиск, киоск, пропуск, отпуск	

В артикуляционном плане явление эпитезы (вставочный гласный звук в инлауте между согласными) способствует работе речевого аппарата, так как гласный звук вставляется между тремя согласными звуками: отпусктан – отпускынан (с отпуска), лозунгтар – лозунгылар (лозунги). Но мы уже привыкли выговаривать такие слова как парктың, фейерверктар, очерктың или же драматургтың. По аналогии в речи все чаще употребляются слова киосктан, танктан, обелисктан.

В пункте 2 части F этого параграфа дано правило присоединения аффиксов к заимствованным словам, оканчивающимся на -сть, -ст, -фт, -зд. Согласно правилу, конечный -сть, -т, -д опускается при прибавлении аффикса. Таких слов в башкирском языке оказалось довольно много:

-сть: власть, санчасть, повесть, волость, ведомость, промышленность, крепость;

-ст: энтузиаст, балласт (балласка?), гимнаст, асбест, арест, трест, протест, манифест, славист (слависка?), лингвист, рецидивист, позитивист, активист, коллективист, объективист, штангист, радист, садист, буддист, энциклопедист, торпедист (торпедиска?), пропагандист, текст, контекст, пост (поска?), компост, тост (тоска?), август и др.

-фт: ландшафт (ландшафтка?), лифт (лифтка?), шрифт (шрифтка?);

-зд: поезд, бронепоезд, подъезд, разъезд, съезд.

В данном случае слова ландшафт, лифт, шрифт на практике не теряют конечный согласный основы при аффиксации: ландшафтка (к ландшафту), лифтка (в лифт), шрифтка (к шрифту).

Башкирский язык заимствовал из русского языка и другие слова, оканчивающиеся на сочетание согласных звуков, например:

-йд: рейд;

-нд: дивиденд, зонд, радиозонд, фонд, раунд, секунд, карборунд;

-рд: бакенбард, ломбард, авангард, арьергард, миллиард, нард, рекорд, лорд, курд, абсурд, бильярд;

-кт: акт, пакт, антракт, контракт, абстракт, такт, контакт, факт, диалект, комплект, проект, конспект, проспект, дефект, эффект, объект, субъект, реликт, конфликт, инстинкт, пункт, агитпункт, продукт;

-лт: болт;

-мт: почтампт;

-нт: бант, квант, гигант, педант, комендант, интендант, секундонт, сержант, вариант, репатриант, официант, кант, фабрикант, практикант, музыкант, лейтенант, рант, транспарант, иммигрант, эмигрант, спирант, аспирант, квартирант, лаборант, докторант, оркестрант, демонстрант, курант, преискурант, франт, десант, диверсант, коммерсант, курсант, экскурсант, дилетант, диктант, диссертант, протестант и др.;

-рт: стандарт, азарт, плацкарт, март, старт, конверт, эксперт, десерт, концерт, спирт, борт, аборт, порт, рапорт, импорт, суппорт, спорт, паспорт, экспорт, транспорт, велоспорт, курорт, сорт, торт, комфорт;

- вт: стратонавт, терапевт, фармацевт;
- пт: рецепт, акцепт, эвкалипт, концепт.

Если о таких словах не сказано в правилах, следовательно, при аффиксации конечный -т не должен утрачиваться. Однако и в орфографическом словаре Р. Г. Азнагулова, и в орфографическом словаре Н. Ф. Суфьяновой, З. К. Ишмухаметова даны примеры на -пт, -вт, где при склонении слова используются названные выше правила: рецепт – рецептың (рецепта), рецепка (на рецепт), рецептан (из рецепта), рецебы(на) (на его рецепт); терапевт – терапевтың (у терапевта), терапевка (терапевту), терапевтан (от терапевта), терапевы (его терапевт), хотя слово ‘эвкалипт’ не теряет конечную букву: эвкалиптың, эвкалиптка (эвкалипту), эвкалипттан (из эвкалипта), эвкалипты(на) (на его эвкалипт).

В башкирском языке в последнее время получило широкое употребление слово ‘концепт’, которое не вошло ни в один словарь. По аналогии со словом ‘рецепт’ оно должно быть записано как концептың (концепта), концепка (концепту), концептан (из концепта), концебы(на) (на его концепт), а не концептың, концептка, концептан, концептына.

Пункт 4 § 35 регламентирует правописание заимствованных слов, оканчивающихся на -пп, -кк, которые при присоединении аффиксов, во-первых, теряют последнюю букву, согласно пункту 3 этого же параграфа (3) (заимствованные слова, оканчивающиеся на удвоенный согласный, при присоединении аффиксов теряют один согласный звук (кросс – кроска (на кросс), металл – металы (его металл), во-вторых, оставшийся глухой согласный звук -п, -к в интервокальном положении озвончается, согласно §21 и §24. Здесь рассматриваются в основном медицинские термины: грипп – грибы, гонококк – гонокогы, стафилококк – стафилокогы, стрептококк – стрептокогы, элеутерококк – элеутерокогы и др. На наш взгляд, в качестве примера в данном случае можно привести правило из кыргызского языка, которое гласит, что в «кыргызском языке не допускается написание трех одинаковых согласных, поэтому при соединении аффиксов на ‘п’, ‘т’, ‘с’, ‘м’ к словам, которые заканчиваются на двойные согласные типа ‘-пп’, ‘-тт’, ‘-сс’, ‘-мм’ один из этих согласных выпадает: киловатт+да – киловатта, киловатт+дан – киловаттан, класс+сыз – классыз и т. п.» [Правила правописания кыргызского языка].

Таким образом, в современном башкирском языке чаще всего наблюдаются спорные вопросы в оформлении заимствованных слов из русского языка. Решить этот вопрос однозначно невозможно уже потому, что процесс заимствования новых слов продолжается. В данном случае нужно попытаться выработать общие принципы оформления таких единиц. Для современного башкирского языка можно предложить следующие правила правописания:

- согласно правилам, слова, оканчивающиеся на -ль, -нь, присоединяют аффикс с гласным переднего ряда. Здесь необходимо добавить, что слова на -рзь, -рмь, -рфь, -кль, -мль, -мь, -фть (кроме слова нефть) подчиняются этому правилу;

- согласно правилам, к словам, оканчивающимся на -рь, -сь, -чь, -шь, -сть, присоединяются аффикс с гласным заднего ряда (кроме названий месяцев). Этому правилу подчиняются слова на -зь (а также слова на -пъ, -тъ, которые могут быть среди современных заимствований);

- при прибавлении аффиксов к заимствованным словам на -сть, -ст, -зд конечный -ть, -т, -д не слышится и не пишется;

- при прибавлении аффиксов к заимствованным словам на -фт(ь), -фт, -йд, -нд, -рд, -кт, -лт, -мт, -нт, -рт, -вт, -пт сохраняются оба согласных: лифт – лифттан (из лифта), шрифт – шрифтты (шрифта); исключение – рецепт – рецебы (его рецепт).

### Литература

- Азнагулов Р. Г. Орфографический словарь башкирского языка. – Уфа: Китап, 1998. – 368 с.  
Алфавит и орфография башкирского литературного языка. – Уфа: ИИЯЛ, 2007. – 28 с.  
Кагарманов Г. Г. Лексика и терминология башкирского языка. – Стерлитамак: СГПИ, 2002. – 242 с.

Правила правописания кыргызского языка: [Электронный ресурс]. URL: <http://bizdin.kg/static/media/pdf/Pravila-pravopisaniia-kyrgyzskogo-iazyka-Kyrgyz-tilinin-zhazuu-erezheleri.pdf> (дата обращения: 14.10.2019).

Суфьянова Н. Ф., Ишмухаметов З. К. Орфографический словарь башкирского языка. – Уфа: Башкирская энциклопедия, 2002. – 352 с.



## ТЕОЛИНГВИСТИЧЕСКИЙ АНАЛИЗ СЛОЖНЫХ ГЛАГОЛОВ, ОБРАЗОВАННЫХ С УЧАСТИЕМ ТЕОНИМОВ В СОВРЕМЕННОМ КАЗАХСКОМ ЯЗЫКЕ

В настоящее время актуальным является изучение прагматических особенностей языковых единиц с точки зрения национального самосознания. В связи с этим статья посвящена анализу устойчивых глагольных словосочетаний, образованных с помощью теонимов, наиболее часто используемых в современном социуме. Авторами рассматривается генезис сложных устойчивых единиц на материале религиозных источников.

**Ключевые слова:** теолингвистика, архетип, теоним, языковое сознание, религиозное познание.

Б. М. Сүйерқұл, А. Мурсал  
Алматы қ., Қазақстан

## ҚАЗІРГІ ҚАЗАҚ ТІЛІНДЕГІ ТЕОНИМДЕРДІҢ ҚАТЫСУЫМЕН ЖАСАЛҒАН КҮРДЕЛІ ЕТІСТІКТЕРДІ ТЕОЛИНГВИСТИКАЛЫҚ ТАЛДАУ

Қазіргі қазақ қоғамында қолданылып жүрген тілдік бірліктердің прагматикалық ерекшеліктерін ұлттық сана тұрғысынан қарастыру өзекті мәселе болып отыр. Осымен байланысты мақалада теонимдердің қатысуымен жасалған тұрақты сипат алған етістікті тіркестер діни таным мен тілдік сана сабақтастығы шеңберінде талданған. Авторлар қазіргі тілдік ұжымда жиі қолданылатын күрделі тілдік бірліктердің түпкі мәнін діни деректер негізінде ашуға тырысқан.

**Тірек сөздер:** теолингвистика, архетип, теоним, тілдік сана, діни таным

**Кіріспе.** Адам қоршаған әлемді және өзін өзі саналы түрде өзгертетін процесс, яғни жеке тұлғаның немесе топтың әлеммен жасайтын өзара байланысы әрекет арқылы жүзеге асырылады. Іс-әрекет – адамның дүниемен қарым-қатынас тәсілі. Іс-әрекет барысында адам табиғатпен өзара әрекетте болып қана қоймайды, оны бірте-бірте өзінің материалдық және рухани мәдениетінің құрамына кіргізеді. Сыртқы дүниені өзгерту адамның өзін дамытуы үшін қажетті жағдай және алғышарт болып табылады. Міне, осыдан іс-әрекеттің қаншалықты маңызды екенін көре аламыз. Әрбір жасаған әрекетіміз үшін жаратушыдан берекетін сұраймыз.

Тілімізде әрекетті таңбалайтын белгілі бір сөздер тобы бар, яғни әрекет етістік арқылы жеткізіледі. Етістіктер құрамына қарай *дара* және *күрделі* болып екіге бөлінеді. Күрделі етістік құрамындағы сыңарлар лексикалық мағынаның жасалуына бірдей қатысады, бір сөз қызметін атқарады. Көрнекті ғалым А.Ысқақов күрделі етістік құрамына құранды етістік, құрама етістік (жай күрделі етістік), тұрақты етістіктерді жатқызады [Маманов 2007, 41].

Осы күрделі етістіктермен ұқсас құбылыстарға – аналитикалық етістіктер мен көсемшелі сөз тіркестері жатады. Құранды етістік күрделі етістік пен құрлымы жағынан әр басқа болғанымен (бірі есім+етістік, екіншісі, етістік+етістік) аналитикалық тәсілмен жасалып, қимылды білдіруі жағынан ұштасып жатады [Маманов 2007, 42].

Аналитикалық етістіктерді де күрделі етістіктерден ажырату қиынға түспейді. Етістіктің аналитикалық формасы – грамматикалық құбылыс, ол сөзжасамға жатпайды. Аналитикалық формада тұрған етістіктер екі бөлімнен тұрады: лексикалық мағына беретін негізгі морфема және грамматикалық мағына беретін көмекші морфема мен аналитикалық формант. Демек, аналитикалық формалы етістіктерді күрделі етістіктерден ажыратудағы негізгі критерий – лексикалық және грамматикалық мағыналарында.

Күрделі етістіктерде тұтас бір ғана лексикалық мағына болады да, қосымша грамматикалық мағына болмайды. Аналитикалық формалы етістіктерде лексикалық мағына жетекші (алдыңғы) сыңарда ғана болады да, екінші сыңар грамматикалық мағына үстейді [Маманов 2007, 43]. Сондай-ақ “Күрделі етістік етістіктің барлық формаларымен түрленеді және оның құрамындағы көмекші етістіктер әртүрлі грамматикалық мағыналарды білдіреді. Күрделі етістік көсемше формаларымен көмекші етістіктердің тіркесуі арқылы жасалады. Күрделі етістіктің құрамындағы алдыңғы көсемше түріндегі сыңары лексикалық мағынаға ие болады да, соңғы сыңары көмекші етістік оған грамматикалық мағына үстейді және етістіктің басқа формаларымен түрленетін конструктивтік қызмет атқарады” (М.И-КТТ етістік., 55 б) [Оралбаева 1975, 136].

Мақалада теонимдердің қатысуымен жасалатын, яғни бір сыңары абстрактылы есім (оның ішінде дінмен байланысты түсіндірілетін), екінші сыңары етістік болып келетін күрделі етістіктер қарастырылады. Ондай етістіктерді дәстүрлі мәдениетімізде әбден орныққан және аға буынның құлағы үйренген, санасына сіңген тілдік бірліктер деуге болады.

Қазақтың рухани адамгершілік тәрбиесінде баланы отаншылдыққа, әдептілікке, ұқыптылыққа, мейірімділікке, табиғатты сүюге, отбасын сыйлауға үйретудің маңызы зор екені мәлім. Оларға жалпыадамзаттық құндылықтардың мәнін түсіндіруде жанында жүрген адамдарды сыйлауға, қадірлеуге, түсіністікпен, ізгілікпен қарауға, риясыз жақсылық жасауға баулуда, адамгершілік принциптерді ғұмыр бойы басшылыққа алатын тұлға ретінде қалыптастыру үшін жиі қолданылатын сөздер: “обал болады”, “ұят болады”, “сауап болады” және “жаман болады”.

Бұлар – бірінші сыңары *обал*, *ұят*, *сауап*, *жаман* абстрактылы есімдерінің, екінші сыңары *бол* сөзінің қатысуымен жасалған діни мәнді тұрақты сипатты күрделі етістіктер.

**Обал** мен **сауап** – мағынасы жағынан бір-біріне ұқсас ұғымдар. *Обал* лексемасы жақсы нәрсенің қадірін біл, зияныңды тигізбе, болмаса түбінде оның зардабын тартасың деген ұғымды білдіреді. Оның қатысуымен жасалған тілдік қолданыстарда, міндетті түрде *тыйым* мағынасы актуалданады. Сондықтан *обалды* тыйым сөз деп түсінген жөн. Мысалы:

Көк шөпті жұлма, ОБАЛ БОЛАДЫ.

Ақты төкпе, ОБАЛ БОЛАДЫ.

Ұяны бұзба, ОБАЛ БОЛАДЫ.

Нанды баспа, ОБАЛ БОЛАДЫ.

Құсты өлтірме, ОБАЛ БОЛАДЫ.

Кісі ақысан жеме, ОБАЛ БОЛАДЫ т.б.

**Сауап** адамға жақсылық, табиғатқа қамқорлық жасау, өзіне де, өзгеге де екі дүниеде пайдалы болатын жақсы іс істеу дегенді білдіреді. *Сауап* сөзінің семантикалық аясында *жөн*, *шын*, *лайық* семалары өзектеледі. Ислам дінінде *сауап* – жасалған игілікті іс үшін Алла тағала тарапынан берілетін сый. Кімде-кім ислам парыздары мен бұйрықтарын орындап, Алла разылығы үшін жақсылықтарға лайық істер атқарса, сауап жинайды, құдайдың шексіз мейірім-шапағатын иеленеді. Діни таныммен астасқан тілдік санамызда орныққан осындай тұрақты бірліктер алынған көптеген аяттар адам баласын Аллаға сенуге, әрдайым күнәдан сақтанып, сауап жинауға шақырады. Сөзімізді діни дереккөздерден алынған үзінділермен дәйектейік: «Алланың құлдарының оянатын әрбір күні көктен екі періште түсіп, олардың бірі: «Я, Алла, Сенің жолыңда қайырымдылық етіп, садақа берушіге несібенді арттыр» десе, екіншісі: «Я, Алла, сараңды опат қыл», – деп тілейді екен (Бұхари). Бұл – садақа беріп, өзгелерге жақсылық жасаған адамның ризығы артып, несібесі молаятынының, өмірінің берекелі болатынының айқын дәлелі. Кейде жақсылық жасағысы келген кісінің қолын жоқтық байлайды, яғни қалтасынан ақша шығарып, өзгелерге садақа беруге мүмкіндігі болмай жатады. Біз әуел бастан-ақ жарты құрмамен болса да, өзімізді тозақ азабынан арашалап қалуға бұйырылғанбыз. Мұндай жағдайда не істеуге болады? Қалтамызды жұқартпай-ақ жақсылық жасап, сауап табудың жолдары бар ма? Шындығында, сауап алғысы келген адам үшін игілікке апаратын жол өте көп екен. Қасиетті Құранда бұл туралы: «Қайырдан не істесеңдер де, күдіксіз, Алла оны біледі» («Бақара» сүресі, 215-аят), «Сонда кім заттың ең кішкентай бөлшегінің салмағындай жақсылық істеген болса, ол оны көреді» («Зилзала» сүресі, 7-аят) делінген [Ақпараттық-танымдық порталы]. Осы тұста тіліміздегі *сауап* сөзінің семантикасы *садақа* теонимімен сабақтас екенін айта кеткеніміз жөн. Мысалы, төменде келтірілген хадис мәтіндеріндегі *садақа* сөзін *сауап* лексемасымен алмастырсақ, мағынасы елеулі өзгеріске ұшырай қоймайды.

1. *Жымию да – садақа*. Пайғамбарымыз (оған Алланың игілігі мен сәлемі болсын): «Еш уақытта шариғаттың рұқсат еткенін елемеуші болма, керек десең, ол бауырыңды күлімдеп қарсы алу тұрғысынан болса да», – деп күлімдеудің де өмірде өзіндік маңызы бар екенін ескертіп кеткен. (Мұсылим)

2. *Жолдан кедергіні алып тастау да – садақа*. Себебі, Пайғамбарымыз (оған Алланың игілігі мен сәлемі болсын): «Мен Жәннәтта, осы дүниеде мұсылмандардың жүруіне кедергі жасағандығы үшін, кесіп тастаған ағашының көлеңкесінде рақаттанып жатқан адамды көрдім», – деген. (Мұсылим)

3. *Жақсы сөз сөйлеу – садақа*. Өзінің тақуалығымен, шыншылдығымен танылған Әбу Бәкір (оған Алла разы болсын) жиі-жиі тілін қолымен ұстап тартады екен. Хазіреті Омар (оған Алла разы болсын) мұның сырын сұрағанда: «Бұл – маған қауіп төндіреді» деп жауап беретін көрінеді. Шындығында, Әли ибн әбу Талиб айтқандай, «Жүрек таза болғанда ғана тілден әдемі сөз шығады». Демек, біз күнделікті сөйлеген сөздеріміз арқылы да сауап, не күнә арқалай аламыз. Мәселен, біреуге қошемет сөз /комплимент/ айту. Compliment (қошемет, марапат) – адам көңіліне қуаныш сыйлайтын, марқайтатын сөз. Оны тіпті мотивацияға да айналдыруға болады. Өйткені, өзі туралы жақсы пікірді

естіген адам одан да жақсы болуға тырысады. Бұл, әсіресе, ерлі-зайыптыларға аса қажет әдістердің бірі. Бүгінгі адамдардың бәрі де мейірімге, ықыласқа, жылы сөзге шөлдеп жүр. Сондықтан комплиментті де шынайы ықыласпен айтар болсақ, бұл артында сауап тұрған құлшылық болып жазылады. Себебі, «Жақсы сөз – садақа» (Ибн Хиббан) [Қазақстан мұсылмандары діни...].

4. *Алланы зікір ету мен дұға намазын оқу – садақа.* Алла елшісі: «Күн шыққан әр күні әрқайсысың денелеріңде неше буын болса, сонша садақа беруге тиістісіңдер. Әрбір айтылған: «Субханаллаһ» және «Әлхамду лиллаһ» садақа болып табылады және әрбір айтылған: «Лә иләһә иллаллаһ» садақа болып саналады, әрбір айтылған: «Аллаһу Әкбар» садақа және жақсылыққа шақыру садақа болып табылады, жамандықтан тыю садақа, бірақ осының бәрін таңертеңгілік оқыған екі бас нәпіл намаз (духа) ауыстыра алады», – деген (Муслим).

5. *Пісірген асыңнан көршілерге бөліп беру де – садақа.* Себебі, Алла елшісі әйел адамға қойдың тұяғын берсе де, көршісіне жақсылық жасауда арланып, ұялмауды бұйырған. Бірде сахаба Абдулла ибн Омар (р.а.) қатты сырқаттанып қалады. Әйелі одан нені қалайтынын сұрайды. Ал ол балық жегісі келетінін айтады. Әйелі дереу балық сатып әкеліп, дәмді ас пісіреді де, алдына қояды. Абдулла ибн Омардың енді жей бергені сол еді, кенет есік қағылады. Арғы жақтан: «Ей, Алланың құлдары, бізді тамақтандырыңдар!» деген дауыс шығады. Ол дереу ұлына балықты ұстатып, әлгілерге беруді бұйырады. Сонда әйелі: «Өзің біраздан ауырып, аш жатырсың. Бұл тағам да өзіңнің жақсы көретінің. Мен де оған аз қаржы жұмсамадым» дейді налып. Сонда Абдулла: «Мен тағамды міскіндерге беруді бұйырдым. Өйткені, Раббым: «Сендер жақсы көрген нәрселеріңнен жұмсамайынша, игілікке жетпейсіңдер» деді емес пе?» – дейді. Мұны естіген ұлы көз жасын ұстай алмай жылап жібереді. Расында, біздің садақамызға, кедейлер емес, өзіміз мұқтажбыз [ҚМДБ-ның Батыс Қазақстан облысы ... интернет ресурсы]. Біз садақа беру арқылы оларды емес, ең алдымен, өзімізді құтқарамыз. Оларға емес, өзімізге жақсылық еткен боламыз...

6. *Әрбір тірі нәрсеге жасалған жақсылық та – садақа.* «Бірде өзінің жолымен келе жатқан адам қатты шөлдей бастайды. Ол бір құдық тауып, онда түсіп шөлін қандырып шыққанда, шөлдегеннен лай топырақты жеп жатқан бір итті көреді. Осыны көрген ол өзіне-өзі: «Бұл ит мен шөлдегендей қатты шөлдеп жатыр», – дейді де, қайтадан құдыққа түсіп аяқ киіміне су толтырып, оны тісімен тістеп, сыртқа көтерілгенше түсірмейді. Сосын сыртқа шығысымен иттің шөлін қандырып су береді. Осы амалы үшін Алла ол адамның бүкіл күнәларын кешіріп, алғысын білдіреді». Мұны естіген адамдар: «Иә, Алланың елшісі, бізге хайуандар үшін де сауап беріле ме?», – деп сұрайды. Сонда Пайғамбарымыз: «Сауап барлық тірі жан үшін беріледі», – деп жауап берген (Бұхари, Муслим) [Алматы қаласының орталық мешіт сайты].

7. *Шаригат рұқсат еткен барлық нәрселер – садақа!* Шындығында «ақшамыз жоқ болса, қалай садақа береміз?» деген сұрақ тек бізді ғана емес, ілгерідегі ізгі буынның бәрін де толғандырған. Бұған мына оқиға дәлел. Бірде Пайғамбарымыздан сахабалары: (бұлар кедей сахабалар еді): «Уа, Алла Елшісі! Байлар (көп) сауап алады. Олар біз намаз оқығандай намаз оқиды, біз ораза тұтқандай ораза тұтады, әрі олар мал-дүниелерінің артығынан садақа береді», – дейді. Сонда Пайғамбарымыз: «Расында Алла сендер үшін садақа қылатын нәрсе жасады емес пе? Шындығында, әрбір тәсбих («Субханаллаһ» деп айту) – садақа, әрбір тәкбір («Аллаһу әкбар» деп айту) – садақа, әрбір тахмид («Әлхамдулилләһ» деп айту) – садақа, әрбір тәһлил («Лә иләһә илләллаһ» деп айту) – садақа, жақсылыққа бұйыру – садақа, теріске шығарылғаннан қайтару – садақа» деген екен. Демек, Алла разылығы үшін жасалған әрбір жақсы іс сауап болып есептеледі [«Жайық» ... орталығының веб-сайты]. Бұл ұғым қазақ тілінде мынадай қолданыстар арқылы вербалданады:

Құлшылық жасасаң, САУАП БОЛАДЫ.

Сабақ оқысаң, САУАП БОЛАДЫ.

Ғылыммен айналыссаң, САУАП БОЛАДЫ.

Біреудің қажеттілігін өтесең, САУАП БОЛАДЫ.

Қарыз берсең, САУАП БОЛАДЫ.

Бірінші болып сәлемдесең, САУАП БОЛАДЫ.

Аш адамды тойдырсаң, САУАП БОЛАДЫ.

Шөлдегенге сусын берсең, САУАП БОЛАДЫ.

Кішілерге қамқор болсаң, САУАП БОЛАДЫ.

Қарияның жүгін көтеріссең, САУАП БОЛАДЫ.

Торға түскен құсты босатсаң, САУАП БОЛАДЫ.

Аталғандардың бәрі де *сауап істер* қатарына жатады. Сондықтан елімізде «сауап үшін» дегенде, жақсылық жасаушы ақша, төлемақы күтпейді. Өйткені ол ісінің бұл дүниедегі көзге көрінбес берекесі мен ақыреттегі шексіз мол сыйын Алладан күтеді.

Ұят. Этимологиясына қарап, түбін тексергенде, *хая* деген сөзге тіреледі. *Хая* /хайиат/ деген *өмір*. Туыстас өзбек тілінде осы сөздің бастапқы мағынасы сақталған. Оларда «Хая – иманнан» деген мәтел бар. Ұят – иманнан. Сонда *өмір* мен *ұят* түбірлес, мәндес сөздер болып шығады. *Өмір* деген ұғымға *ұят* деген сөз балама бола алады екен. Басқа сөз емес. Яғни, ұяты болмаса, адамның бұл өмірде жүрген-тұрғанынан еш пайда жоқ екен. Қазақтың «Өлімнен ұят күшті» дегені де осыдан болса керек. Бір сөзбен айтқанда – тәрбие. Бұл ұғымның қазақы тілдік санадағы көрінісіне күнделікті сөзжұмсадан мынадай мысалдар келтіруге болады:

Ашық-шашық киінбе, ҰЯТ БОЛАДЫ.

Кісіге қарап керілме, ҰЯТ БОЛАДЫ.

Айғайлап күлме, ҰЯТ БОЛАДЫ.

Уәдеңде тұрмасаң, ҰЯТ БОЛАДЫ.

Жалған сөйлеме /өтірік айтпа/, ҰЯТ БОЛАДЫ.

Ата-бабаларымыз қашанда адамның бір әдепсіз қылығын көрсе: «Ұят болады» деп бір-біріне ескертіп, ол істің күнә немесе жамандық, әдепсіздік екендігін меңзеп, қоғамды рухани күйреуден сақтандырып отырған. Ұятсыздықтың қашанда жақсылыққа апармайтынын білген. Арсыздық аса жаман қасиет деп есептеліп, адалдық, ар-намыс, ұяттылық қашанда жоғары бағаланған. Бұлар адамды бақытқа, табысқа жетелейді деп иланған. Осы тұрғыда түркі халықтарының данышпаны Жүсіп Баласағұнның мына өлеңін келтіргеніміз жөн:

Пәле-жала жолын ұят кеседі,

Ұятсыздық – ердің емсіз кеселі.

Ашық мінез, пәктік, ұят, серік боп,

Жарасады қуанышқа көрік боп.

Адал ар мен ақ көңіл де еленер –

Екі жалған бақытына бөленер, – деп [Ислам және Өркениет ... діни-танымдық газетінің сайты] ортағасырлық ғұлама ұятсыздықты жазылмайтын дертке теңейді.

**Жаман болады.** Бұл тіркестің мағынасы «Жақсы сөз – жарым ырыс» дегенмен сабақтас, яғни жамандықтан аулақ жүр дегенді білдіреді. Тілімізде сондай-ақ «Жақсының шарапаты, жаманның кесапаты тиеді» деген паремия да бар.

Қазақта «жаман болады» деген тыйымдар өте көп. Олардың сыры, шын мәнінде, халықтың танымымен байланыстыра қарастырғанда ашылады. Мысалы, үйге жүгіріп кіре бергенде әжеміздің тоқтататыны – бала кезден таныс сурет... «Балам, үйге жүгіріп кірме, жаман болады». «Сонда не болады?» ... «Жаман болады...». Шынында, бұрынғы заманда елге жау шапқанда немесе оқыс кісі өлімі орын алған жағдайда, сол қаралы хабарды жеткізушілер атқа қамшы басып, екпіндеп жеткен бойы үйге жүгіріп кіретін болған...

«Бейітке қарай қолыңды созба // мазарға қолыңды шошайтпа, жаман болады». Тіпті байқамай қолын созған адам «жаман болмауы үшін» саусағын аямай үш рет тістеуі тиіс болатын... Себебі, бұрынғылар, молаға қолын созған адам жуық арада бір жақынынан айырылады, деп сенген. «Күн батып бара жатқанда ұйықтама, жаман болады». Себебі, екі кештің арасында күллі жын-жыбыр жиналып, жер бетіне шығып, бейқам жатқан адамдарды айналдырады, *жын соғады, албасты басады, шайтан азғырады* т.б. деген негатив мәнді нанымдар бар.

«Тізеңді құшақтама, жаман болады». Себебі, қазақ ұғымында жақынын жоғалтып жалғыз қалған адам «қу тізесін құшақтап отыратын» болған.

«Екі қолыңды төбеңе қойма, жаман болады». Себебі, ел арасында жаман аты шығып, жасаған ауыр қылмысы, күнәсі әшкере болған адамды, өзгелерге жұғып кетпеуі үшін, яғни жамандық атаулы көпке жайылмай, соның өзімен бірге кетуі үшін, екі қолын төбесіне қойғызып, елден аластап, яғни қуып жіберетін болған.

«Екі бүйіріңді таянба, шашыңды жайма, жаман болады». Себебі, бұрынғы кезде жақынынан айрылып, ауыр күйге түскен әйелдер, шаштарын жайып жіберіп, екі бүйірін таянып, беттерін осып, тырнап отырып жылайтын, яғни марқұм болған кісіні жоқтайтын болған.

«Киімді желбегей жамылма, жаман болады». Себебі, екі қолы жоқ адам ғана киімді осылай киеді деп түсінген. Көріп отырғанымыздай, «жаман болады» деп аяқталатын тыйымдардың әрқайсысының өзіндік себебі бар екен. Қазіргі кезде олардың көпшілігі ұмытылған. Дегенмен, қай-қайсысы да адамның басына түсетін қайғы мен азаптану кезінде жасалатын әрекеттер болғандықтан, ырымшыл бабаларымыз *жамандықты шақырмау* үшін, осындай көптеген әрекеттерге тыйым салған.

Біздің айтпағымыз, осындай тыйымдарды құлағына сіңіріп өскен ұрпақтың расында да көптеген жаманшылықтардан алыс болғаны. Яғни ол сөздерді ұрпақ тәрбиелеуде ерекше тәлім-тәрбие, ереже іспетті қолданған қазақ жаман болған жоқ. Қазақтың бала тәрбиесі жөнінде арнайы жазған заңы жоқ.

Бірақ «жаман болады» деген бір ауыз жақсы сөзі бар! Сол – заң. Сол – ереже. Өйткені, жас балаға жағымсыз нәрселерді түсіндіру оңай емес. Сондықтан «жаман болады» деп келте қайырады. Саналы ұрпақ оның мәнін өсе келе, көкірек көзімен ұғынады. Алайда қазіргі балаларымыз бен жасөспірімдерімізге бабалардан қалған осындай «жазылмаған ережелерді» ашық айтып, мәнін түсіндіріп отыру – баршамыздың міндетіміз. Немқұрайдылықтан бас тартып, әрқайсымыз айналамыздағы жастарға осындай тәлімді бере бастасақ, еліміздегі рухани, мәдени ахуал жақсарар еді.

Сонымен тілімізде сақталған тыйымдардың бірқатарын келтірейік:

Түнде күл төксе, ЖАМАН БОЛАДЫ /күлтөбе жын-шайтандардың мекені болып саналады/.

Қол жуған соң сілкісе, ЖАМАН БОЛАДЫ /ырысы кетеді/.

Қатар отырған екі адам ет кессе, ЖАМАН БОЛАДЫ /дау-шарға ұрынады/.

Түнде тырнақ алса, ЖАМАН БОЛАДЫ /тырнақ ұшып, ас-суға түссе, оның астындағы кір /ол да шайтанның бір мекені деп есептеледі/ немесе қазіргі тілмен айтқанда – микробтар түрлі аурулар тудырып, адам ағзасына зиянын тигізеді.

Кеште /қараңғыда/ мал табынын араласа, ЖАМАН БОЛАДЫ /түрлі бәлеге ұшырауы мүмкін/.

Туыстар арасында ит алысса, ЖАМАН БОЛАДЫ /ұрыс-керіс болуы мүмкін/.

Саусақпен адам санаса, ЖАМАН БОЛАДЫ /кісі басы азаюы мүмкін/.

Түнде түс айтса, ЖАМАН БОЛАДЫ /көрген түс жақсы болса, алдымен «өңім екен...» деген сөзді айтып, сосын жақын, тілектес кісілерге баяндап беруге болады. Ал жаман түсті ешкімге жария етпей, ұмыту керек.../.

**Қорытынды.** Сонымен «Обалды біліп өссең – сауапқа кенелесің» деген халық даналығында үлкен мән бар екенін көрдік. Үлкендер адам баласы жасауға болмайтын жаман қылықтарды «обал болады» деп ескерте отырып, кішілерді, балаларды жаман әдеттерден сақтандырады. «Сауабын аласың» дей отырып, жақсы істер жасауға жетелейді. «Сауабын аласың» дегеннің өзі *жақсы іс істеп, жақсылыққа жетесің, алғысқа бөленесің* дегенді білдіреді. Сондықтан үлкендердің «обал болады», «сауабын аласың» дегеніне құлақ асып, тілін алу үлкен жақсылықтарға жеткізеді.

Қорыта келгенде, қазақ халқы бала тәрбиесін төрт ауыз сөздің төңірегіне топтастырған: *жаман болады, ұят болады, обал болады, сауап болады*. «Жаман болады» деу арқылы балаға ненің жаман, ненің жақсы екенін ажыратып беріп, тыйып отырған. Ал оң-солын танып үлгермеген сәбиін ұятты істерге ұрындырмай, арсыздық пен ашкөздіктен, түрлі қалауларынан «ұят болады» деу арқылы тыйған. Сондай-ақ, перзентін ойын баласы кезінен бастап ұя бұзып, жан-жануарға жәбір көрсетуден, асты қорлап, ысырапшылдыққа жол беруден «обал болады» деп сақтандырған. Бұлардан бөлек, балаға қамқорлық көрсетуге, жақсылық жасауға үйрету «сауап болады» деген тұрақты тілдік бірлік арқылы таңбаланған. Бұл жерде үлкен тәрбие жатыр. Ересектердің әрбір әрекеті мен әрбір сөзі сәби санасына жазылып жатады. Кейін олар соны қайталайды. Осы себепті, бала көзінше бейәдеп сөздер айтып, жағымсыз қылықтар мен жөнсіз әрекеттер жасамау да – маңызды тәрбие өзектерінің бірі. Қазақ сөзге тоқтай білген заманда Бөлтiрiк шешен: «Байлық – ерiп кететiн мұз, Бақ – ұшып кететiн құс, Бала – артыңда қалар iз» деген. Ендеше, артымызда қалатын iздiң iзгiлiкке толы болуын тiлесек, баланы сөзiмiз арқылы жастан тәрбиелейiк. Ол үшiн өмiрдегi қандай жағдай болса да, ойымызды сыпайы, терең, ықшам, жинақы жеткiзуге болатын тұрақты тiлдiк бiрлiктердi пайдаланғанымыз жөн. Мақалада сондай қолданыстардың төртеуiне ғана тоқталдық. Ал ана тiлiмiздiң сарқылмас байлығын дiни таныммен ұштастыра отырып тереңiрек зерттеу келешек еншiсiнде!

### Әдебиеттер

«Жайық» деструктивтік діни ағымдардан жапа шеккендерге көмек орталығының веб-сайты: [Электронный ресурс]. URL: <http://www.antisektazko.kz> (шағымдану күні: 17.10.2019).

Ақпарттық-танымдық порталы: [Электронный ресурс]. URL: <http://www.e-islam.kz> (шағымдану күні 17.10.2019).

Алматы қаласының орталық мешіт сайты URL: <http://www.azan.kz> (шағымдану күні: 17.10.2019).

Ислам және Өркениет. Қазақстан мұсылмандарының діни-танымдық газетінің сайты: [Электронный ресурс]. URL: <http://www.kazreligiya.kz> (шағымдану күні: 17.10.2019)

Қазақстан мұсылмандары діни басқармасы ресми интернет ресурсы: [Электронный ресурс]. URL: <http://www.muftayat.kz> (шағымдану күні: 17.10.2019).

ҚМДБ-ның Батыс Қазақстан облысы бойынша филиалының ресми интернет ресурсы: [Электронный ресурс]. URL: <http://www.oral-islam.kz> (шағымдану күні: 17.10.2019).

Маманов Ы. Қазақ тіл білімінің мәселелері. –Алматы: Арыс, 2007. –488 б.

Оралбаева Н. Қазіргі қазақ тіліндегі етістіктің аналитикалық форманттары. – Алматы: Мектеп, 1975. – 136 б.

## **ПРОБЛЕМЫ КОГНИТИВНОЙ ЛИНГВИСТИКИ В ОБУЧЕНИИ КАЗАХСКОМУ ЯЗЫКУ**

Когнитивная лингвистика, новое направление в казахской лингвистике, имеет всего десятилетнюю историю развития и находится на стадии становления. В статье дается обзор развития направления когнитивной лингвистики в казахском языкознании. Проводится критический анализ научных публикаций, отмечается, что монография Оразалиевой является фундаментальным исследованием в этой области.

**Ключевые слова:** новое направление, когнитивная лингвистика, интеллект, мысль, познание, концепт.

*А. Н. Турлыбекова, М. К. Амангазиева,  
Ақтобе қ, Қазақстан*

## **ҚАЗАҚ ТІЛ БІЛІМІНДЕГІ КОГНИТИВТІК ЛИНГВИСТИКА МӘСЕЛЕЛЕРІ**

Мақалада қазақ тіл біліміндегі жаңа бағыттардың, соның ішінде когнитивтік лингвистиканың қалыптасып, дамуы жайлы және осы бағытта жұмыстанған зерттеушілер еңбектеріне жалпы шолу жасалынды.

**Түйін сөздер:** жаңа бағыт, когнитивтік лингвистика, ақыл, ой, таным, концепт.

Қазіргі жаһандану үдерісі ғылым салаларының тоғысуы мен бірігуіне ғана ықпал етпей, сонымен қатар тың ғылыми салалардың қалыптасуы мен дамуына да әсер етті. Жаңа бағыттағы лингвистика – тіл білімінің қоғамдық-әлеуметтік орны мен рөлінің күшеюіне орай, адамзаттың ой мен тіл, сана мен сөйлеу, ым-қимыл, таным мен тілдік қарым-қатынас үдерістерінің бірлескен әрекеті нәтижесінде тілші-ғалымдар алдына жаңа идея, тың ізденісті қажет еткен сала. Мұндай бағыттар қазіргі өмір талабына сәйкес пайда болып отырғаны белгілі.

Бұл жаһандану үдерісі когнитивтік лингвистиканың қалыптасуына да әсер етті. Когнитивтік лингвистика – тілдің танымдық теориясы мен танымдық қызметін зерттеп, адамзат білімінің түзілуі мен оның қызмет болмысын қарастырады. Бұл сала – жалпы тіл білімі проблемаларымен да, жеке тіл білімі мәселелерімен де тығыз байланысты [Жаманбаева 1998].

Қазақ тіл білімінде когнитивтік лингвистика дербес сала болып қалыптасқалы көп уақыт өткен жоқ. Жоспарлы түрде зерттеу мен оқытудың өзі тек кейінгі он шақты жыл. Қазақ тілінде атап көрсетерліктей іргелі арнайы зерттеу еңбегіне Оразалиеваның монографиясын жатқызуға болады. Қазақ тіл білімінде когнитивтік лингвистика бойынша біраз зерттеу еңбектері жарық көрді, оларда сөйлеу мен танымның ықпалдастығы туралы, когнитивтік лингвистика мен когнитологияның терминдері, оның басқа ғылымдармен байланысы, жеке тұлғаның сөз саптауы жайында тұжырымдар табуға болады.

Тілді антропоцентристік парадигма бағытында зерттеудің бір саласы – ұлттық мәдениетімен, салт-дәстүрімен, рухани болмысымен сабақтастыра қарастыру. Осы бағыттың басты көрінісі тіл білімінің этнолингвистика, лингвомәдениеттану, лингвоелтану тәрізді салалармен негізделді. Бұл мәселемен шұғылданған ғалымдар жалпы, қазақ ғылымында когнитивистика жайында көзқарас білдіргендер деп Ә. Қайдар, Е. Жанпейісов, М. Копыленко, Р. Сыздықова, Ж. Манкеева сияқты ғалымдарды атауға болады. Өйткені бұл ғалымдар этнолингвистика шеңберінде концептілер туралы да пікірлер айтқан.

Қазақ тіл біліміндегі когнитивтік бағыттың қалыптасуында Э.Д. Сүлейменова, Н.Ж. Шаймерденова, Г.Г. Гиздатов, Ж. Манкеева, Қ. Жаманбаева, Г. Смағұлова, Б. Хасанов, Б. Қасым, А. Ислам еңбектерінің маңызы зор.

Қазақ тілін когнитивтік зерттеуге арнаған алғашқы жұмыстардың бірі ретінде Қ.Ә. Жаманбаеваның «Тіл қолданысының когнитивтік негіздері: эмоция, символ, тілдік сана» атты монографиясын атауға болады. Бұл жұмыста тілдік сана құрылымы, гештальт теориясы, тілдік сана және мәтін құрылымы, тілдік модельдер, мұң концептісі, т.б. когнитивтік мәселелер қарастырылған [Манкеева 2004, 156].

Профессор Ж. Манкееваның «Қазақ тілін зерттеудің когнитивтік негіздері» атты мақаласында М. Бирвиш, Хомский, Лакофф, Коок, Г.В. Колшанский, Ю.Д. Апресян және тағы басқалардың

есімдерімен байланысты когнитивтік лингвистиканың пайда болуы мен дамуына қысқаша шолу жасалады [Жаманбаева 1998, 40].

Кейінгі онжылдықтар белесіндегі зерттеулер этносты, ұлт мәдениетін тіл арқылы анықтауды негізгі қағида етеді (Ә. Қайдар, Е. Жанпейісов, Р. Сыздық, Н. Уәли, Ж. Манкеева, Қ. Жаманбаева т.б.). Нақты айтқанда, тілді тұтынушы ұлтты алғашқы орынға қоятын лингвистикадағы антропоэзектік парадигма өз аясында жаңа бағыттарға жол ашуда. Осымен байланысты Г. Снасапова, Елемесова, М. Күштаева, Н. Аитова, А. Әмірбекова, Қ. Жаманбаева Э. Оразалиева, А. Ислам, Ә. Айтпаева тәрізді жас ғалымдардың жұмыстарында тілтаным кілті деп қаралып, когнитивтік лингвистика мәселелері этнолингвистика, лингвомәдениеттану, психоллингвистика, әлеуметтік лингвистикамен тығыз байланыста зерттелген.

М.Т. Күштаеваның «Тары концептісінің семантикалық құрылымы мен лингвомәдени мазмұны» атты зерттеуі «Тары» сөзіне қатысты тұрақты тіркестер, коннотациялық мағыналарды айқындауға негізделіп, тілді ұлттық психологиялық таныммен сабақтастыра қарастыруға бағытталған еңбек болып саналады.

Келесі еңбектің бірі – Ә. Әмірбекованың «Концептілік құрылымдардың поэтикалық мәтіндегі вербалдану ерекшелігі (М. Мақатаев поэзиясы бойынша)» атты диссертациясы. Ғалым өз жұмысына концептілік құрылымдарға және ақын М. Мақатаевтың танымдық болмысын «Өмір» концептісінің негізінде талдау жасап, «Көркемдік сана», «Когнитивтік модель», «ассоциация», «бейне» т.б. ұғымдарға поэтикалық мәтіннен мысалдар келтіре отырып саралады.

2007 жылы Э.Н. Оразалиеваның «Когнитивтік лингвистика: қалыптасуы мен дамуы» атты монографиясы жарық көріп, зерттеуші өз еңбегінде қазақ тіл біліміндегі когнитивтік бағыттың қалыптасу, даму алғышарттарын ғылыми теориялық тұжырымдарға сүйене отырып қарастырады.

Қорыта айтқанда, когнитивтік лингвистикада бәріне бірдей ортақ мәселе «ақыл – ой – тіл – репрезентациялау – концептуализациялау – категоризациялау – қабылдау» ұғымдарының когнитивтік шынжырдағы өзара қарымқатынасын анықтау мәселесі болып отыр [Маслова 2004, 225].

Қазақстандағы когнитивтік лингвистиканың бағдары этномәдени немесе мәскеулік бағыттан өріс алған деп тұжырымдауға болады.

#### **Әдебиеттер тізімі:**

Ақбердиева Б. Когнитивтік лингвистика. – Алматы, 2009.

Жаманбаева Қ.Ә. Тіл қолданысының когнитивтік негіздері. – Алматы, 1998.

Манкеева Ж. Қазақ тілін зерттеудің когнитивтік негіздері. Тілтаным, 2004. №1.

Маслова В.А. Когнитивная лингвистика. – Минск: Тетра Системе, 2004. – 225 с.

Оразалиева Э. Когнитивтік лингвистика: қалыптасуы мен дамуы. – Алматы, 2007.

## КОНЦЕПТ “РЕЛИГИЯ” В РОМАНЕ РИНАТА КАМАЛА “ТАНЯ-ТАНХЫЛЫУ”

Данная статья посвящена исследованию особенностей языкового воплощения концепта «религия» на материале романа Р. Камала «Таня-Танхылу». Семантическое пространство концепта «религия» включает в себя лексемы, фразеологические единицы, пословицы и поговорки.

**Ключевые слова:** лингвокультурология, языковая картина мира, концепт, лексема, концепт «религия».

Г. Р. Юмагулова  
Өфө к., Рәсәй

### РИНАТ КАМАЛДЫҢ «ТАНЯ-ТАҢҢЫЛЫУ» РОМАНЫНДА «ДИН» КОНЦЕПТЫ

Тел ғилемдә лингвокультурология фәненә иғтибар арткандан-арта. В. А. Маслова тикшеренеүҙәрҙә ярашлы, “лингвокультурология – лингвистика һәм культурология сигендә барлыкка килгән һәм телдә сағылыш тапкан, нығынып калған халыҡ мәҙәниәтенә күренештәрҙән тикшеренеүсә фән. Был дисциплиналар фән өлкәһе, үҙенең максаты, мәсьәләһе, алымдары һәм тикшеренеү объекты буларак үз аллы тармак булып үсешә” [Маслова 2007, 28]. Уның төп берәмектәрҙенә берәһе булып концепт һанала. Ғалимдар “концепт” терминына төрлө билдәләмә биргән, уларҙың билдәләмәләре мәғәнә яғынан окшаш, шулай за айырмалыҡтар бар. Профессор Л. Х. Сәмситова билдәләүенсә, концепт аша кешелә туған тел образы һәм уның концептосфераһы барлыкка килә. Туған тел аша милли мәҙәниәттә өйрәнеү укыусыларҙың аңы формалашыуында ғына түгел, ә уларҙың мәҙәниәткә берәһеүендә мөһим роль уйнай [Сәмситова 2013, 3]. Академик М. В. Зәйнуллин фекеренсә, “донъяның тел картинаһы мәсьәләһен тикшергәндә төп конструкт булып концепт төшөнсәһе алына. Тел һәм мәҙәниәт араһындағы бәйләнештәрҙә асыҡлағанда теге йәки был этник үҙенсәлектә сағылдырыусы концепттарҙы анализлау максатка ярашлы” [Зәйнуллин 2008, 33].

Телдәге концепттар йыйылмаһы концептосфераны барлыкка килтерә. Башҡорт теленә концепттар системаһы Ринат Камал ижадында зур сағылыш тапкан. Шуларҙың берәһе – “дин” концепты.

Башҡорттар бик борондан ислам динен кабул иткән, шәриғәт канундарына буйһоноп, уға яраклашып йәшәгән милләттәрҙең берәһе булған. Әлегә көндә лә ислам дине башҡорт халқында киң таралған. Бик күптәр намазға басқан, бәләкәй йәштән бик күптәр доғалар өйрәнә. Башҡорт донъя тел картинаһында ла ислам диненә, ислам тәғлимәттәрҙенә кағылышлы лексикон ныкты үсешкән һәм “иман” концепты менән тығыз бәйләнештә тора. Л. Х. Сәмситова “дин” концептын ике яклы итеп билдәләй: “Дин – кешенә Алла менән аралашыу ихтыяжын сағылдырыусы төшөнсә. Фольклорҙа “дин” концепты бер яктан кешенә Аллаға яҡын булыу теләге менән билдәләнгән, икенсә яктан Алланың кешенә рухи башланғысына йоғонтоһон аңлата” [Сәмситова 2013, 364].

Һүз “Хәҙергә башҡорт теленә аңлатмалы һүзләге”ндә шулай бирелә: “Донъяның барлыкка килеүен һәм булмышын тәбиғәттән тыш көскә генә кайтарып калдырған караш һәм шуға нигезләнгән йолалар системаһы” [Хәҙергә башҡорт теленә аңлатмалы һүзләге 2004, 101].

“Дин” концепты *динле, иман, иманлы, мөфтөй* лексемалары аша тормошка ашһа, уның юклығы *иманһыз, динһез* һ.б. һүзҙәр менән белдерелә.

Ринат Камалдың “Таня-Таңһылыу” романында борондан калған “дин” концептына кағылышлы ырым-ышаныуларҙы осратырға мөмкин. Мәсәлән: *Мөрдәнән сыккан төтөнгә ен килә* (123-сө б.).

Ринат Камалдың “Таня-Таңһылыу” романында ислам һәм христиан дине тураһында бергә һүз бара, сөнки төп герой Таңһылыу рус егетә Петрикка кейәүгә сыға. Үзе лә башҡорт телен бик белмәгән, ислам дине канундарынан йырақта йөрөгән был кыз асылы менән христиан диненә яҡын, әммә уны сит милләттәр, сит дин кешеләре генә үҙһенмәйҙәр. Был романда “дин” концепты тураһында һүз барғанда ошолар хақында бәйән ителә.

“Дин” концепты тураһында һүз йөрөткәндә, әлбиттә, “дини караш буйынса бөтә донъяны яраткан, уның менән идара иткән юғары зат – Аллаһы Тәғәлә” [Шәкүр 2004, 24] исеме һәр ерҙә телгә йыш алына. Романда ла шулай: *Кайтам, Алла бойорһа* (153-сө б.); *Алла бирһә, быйыл Балтиканан китәм, етте* (170-сө б.).



Аллаһы Тәғәләнең синонимы *Хозай*, *Тәңре* һүзәрә менән алынған миҫалдар романда ярайһы уҡ урынды ала: *Тағы дөрөсөн Хозай үзе генә белә* (152-се б.); *Һаҡланғанды Хозай һаҡлармын тигән. Һин гонаһһың йән, изге фәрештәләр затынанһың. Хозай, әйе, һезең Тәңрегеҙ, һинең ишеләрҙе аяр, ярлыҡар, игелегенән мәхрүм итмәҫ. Ышан һин үзегеҙең Тәңрегеҙгә, ышан!* (117-се б.).

– *Бирһен Хозай, бирһен Хозай... – Марфа әбей бик дини ине, шулай үз алдына һөйләнде* (100-се б.).

– *Ха-ха-ха... Тәңреһенән юрай... (105-се б.).*

Шәриғәт канундарына ярашлы башкарылған *доға кылыу, фәрештә, иман* һ.б. кеүек һүзәрҙе лә “дин” концептының бер өлөшө буларак карарға мөмкин. “Таня-Таңһылыу” романында ул һүзәрҙең бирелеше түбәндәгесә: *Әбейем көлә: “Сукынма, доға кыл, үзеңдекәнән айбанма!” – ти* (101-се б.); *Әллә улар күктән төшкән фәрештәмә?* (116-сы б.); *Көн дә теләйем, кис ятқан һайын доғаларымды укып ятам* (129-сы б.). *Вәғәҙә – иман. Иманына хыянат итмәгән Полина Сергеевна, ниһайәт, һүзен тотто* (153-се б.).

Ислам динен инкар иткән, был диндән, уның йола-канундарынан көлгән, мыҫкыл иткән орактарҙы ла осратырға мөмкин Ринат Камалдың “Таня-Таңһылыу” романында. Мәҫәлә: *Динебез ҙә, гәҙәттәребез ҙә башка, бер ҙә генә йәбешмәй, юк менән булышмайыҡ, килешмәгәнде, – тип торалар, ти* (108-се б.); *Петяға армиянан “шул башкирка-басурманкаға өйләнһең, “ағай” тип һанамайым, – тип язып ятқан* (125-се б.); – *Дини бәндә, мусулманка!* (105-се б.).

Шулай уҡ был романда “дин” концепты йыш телгә алынып, уларҙы кулланып әрләүҙәрҙе лә йыш осратырға мөмкин: *Иманын онотқан, диненә төкөргәнде? Иблескә йәнән дә, тәнән дә һатқан. Үз илендә үзенекен тапмаған, сит-ят дин кешенә ҡапланған* (135-се б.).

Ринат Камалдың “Таня-Таңһылыу” романында “дин” концептына бәйләнгән, “мәғәнәһе менән төзөлөшө яғынан ойошқан, нығынған, телмәргә эҙер килеш алына торған, экспрессив бизәкле, образлы фразеологик берәмектәр”зе лә [Ишбаев 2012, 77] осратырға мөмкин. Мәҫәлә, кемделер аҡылға ултыртыу максатында *иманын укытыу* тигән фразеологик берәмек осраны:

– *Николай Григорьевич, һез мине Таисия Алексеевнаға эләктергән өсөн тағы шелтәләргә саҡырҙығыҙмы? Бының иманын укытайым әле тиһегеҙме?* (115-се б.).

*Иманһың* тигән әрләү һүзе лә кинәйәләп, йүнһез, алама кеше тигәнде белдерә: *Теге иманһың ни арағы таптырып йәнәмде кыя* (128-се б.); – *Тағы теге иманһыңдың, шул ғына ошо көнгә төшөрҙө һине* (116-сы б.).

Романдың геройҙары ике дин кешеләре булғас, әлбиттә, христиан диненә ҡағылышлы концепттарҙы ла йыш осратырға мөмкин. Мәҫәлә: *Мына крест, мин әйтте тиерһең* (117-се б.); *Әбейем: “Ярраббым, бер үзем нишләр инем икән был төндә?” – тип укына ла сукына* (123-се б.); *Марфа әбей әйтмешләй, “да Бог с ними со всеми”* (155-се б.).

Шулай итеп, Ринат Камалдың “Таня-Таңһылыу” романында “дин” концептына ҡараған һүзәрҙәр геройҙарҙың ниндәй милләт кешенән һәм уларҙың дингә ни тиклем яҡын булыуҙарын билдәләй.

### Әҙәбиәт

Зайнуллин М.В. Общие проблемы лингвокультурологии. – Уфа: ред. БГУ, 2008. – 342 с.

Камал Р. Таня-Таңһылыу: Роман һәм повестар. – Өфө: Китап, 1996. – 368 б.

Маслова В.А. Лингвокультурология. – М., 2007. – 357 с.

Санъяров Ф.Б. Башкорт телен лингвокультурологик аспектта укытыу // Башкортостан укытыусыһы. – 2013, № 2. – 15–18-се б.

Самситова Л.Х. Культурные концепты в башкирской языковой картине мира. – Уфа: Гилем, 2015.

Сәмситова Л.Х., Әхмәтшина Г.С., Ижбаева Г.Р., Байназарова Г.М., Килмәкова Г.Н., Ташбулатова Р.М. Башкорт фольклоры концептосфераһы. Лингвокультурологик һүзлек. – Өфө: Китап, 2013.

Хәҙерге башкорт теленәң аңлатмалы һүзлегә / Г.Д. Зәйнуллина, З.Ә. Сиразитдинов, Н. Ф. Суфьянова, З.Ф. Уразбаева, З. Ф. Ураксин. – Өфө, 2004.

## СВЕДЕНИЯ ОБ АВТОРАХ

**Абдигази Сандугаш Хасенкызы**, старший преподаватель Актюбинского регионального государственного университета им. К. Жубанова Республики Казахстан, г. Актобе (Республика Казахстан).

**Абдирова Шолпан Гайдаровна**, докторант Актюбинского регионального государственного университета им. К. Жубанова Республики Казахстан, г. Актобе (Республика Казахстан).

**Абдурахмонова Нилуфар**, кандидат филологических наук, старший преподаватель Ташкентского государственного университета узбекского языка и литературы имени Алишера Навои, г. Ташкент (Республика Узбекистан).

**Алнайяр Халид**, докторант Государственного университета Хельсинки, г. Хельсинки (Финляндия).

**Амангазиева Меруерт Кусаиновна**, доктор филологических наук, декан факультета педагогики Актюбинского университета им. С. Баишева Республики Казахстан, г. Актобе (Республика Казахстан).

**Андрианова Инна Николаевна**, старший лаборант-исследователь Центра инновационных языковых технологий Дома дружбы народов Республики Коми, г. Сыктывкар.

**Ароянц Аэлита Акоповна**, кандидат филологических наук, доцент кафедры маркетинга и бизнес-коммуникаций Кубанского государственного технологического университета, г. Краснодар.

**Бочкарев Василий Васильевич**, старший лаборант-исследователь отдела якутского языка Института гуманитарных исследований и проблем малочисленных народов Севера Сибирского отделения Российской академии наук, г. Якутск.

**Буранкулова Эльмира Тимербаевна**, докторант Актюбинского регионального государственного университета им. К. Жубанова Республики Казахстан, г. Актобе (Республика Казахстан).

**Бускунбаева Лилия Айсовна**, кандидат филологических наук, старший научный сотрудник лаборатории лингвистики и информационных технологий Института истории, языка и литературы Уфимского федерального исследовательского центра Российской академии наук, г. Уфа.

**Валиева Гульназ Дауратовна**, кандидат филологических наук, доцент Башкирского государственного университета, г. Уфа.

**Валиуллина Зайнаб Рахимьяновна**, кандидат философских наук, доцент кафедры этики, культурологии и связей с общественностью Башкирского государственного университета, г. Уфа.

**Габдрахманова Фануза Хайдаровна**, кандидат филологических наук, доцент кафедры общего языкознания и тюркологии ИФМК Казанского (Приволжского) федерального университета, г. Казань.

**Гайнуллина Гульназ Фоатовна**, кандидат филологических наук, старший научный сотрудник, доцент Института языка, литературы и искусства им. Г. Ибрагимова Академии наук Республики Татарстан, г. Казань.

**Гатауллин Рамиль Раисович**, кандидат технических наук, научный сотрудник Института прикладной семиотики Академии наук Республики Татарстан, г. Казань.

**Гатиатуллин Айрат Рафизович**, заведующий отделом интеллектуальных информационных систем Института прикладной семиотики Академии наук Республики Татарстан, г. Казань.

**Гафарова Вилюза Робертовна**, ведущий научный сотрудник отдела когнитивных исследований Института прикладной семиотики Академии наук Республики Татарстан, г. Казань.

**Гейт Маргарита Вячеславовна**, магистрант Башкирского государственного педагогического университета им. М. Акмуллы, г. Уфа.

**Гильмуллин Ринат Абрекович**, кандидат физико-математических наук, заместитель директора Института прикладной семиотики Академии наук Республики Татарстан, г. Казань.

**Даржинова Любовь Владимировна**, докторант Образовательного университета Гонконга (Гонконг).

**Дюсенов Багдат Даниярович**, научный сотрудник Института востоковедения им. Р. Б. Сулейменова Комитета науки Министерства образования и науки Республики Казахстан, г. Алматы (Республика Казахстан).

**Ералханқызы Молдир**, магистрант Актюбинского регионального государственного университета им. К. Жубанова Республики Казахстан, г. Актобе (Республика Казахстан).

**Жанабекова Айман**, доктор филологических наук, ведущий научный сотрудник Института языкознания им. А. Байтурсынова Министерства науки и образования Республики Казахстан, г. Алматы (Республика Казахстан).

**Жубанов Аскар Кудайбергеноулы**, доктор филологических наук, главный научный сотрудник отдела прикладной лингвистики Института языкознания им. А. Байтурсынова Министерства образования и науки Республики Казахстан, г. Алматы (Республика Казахстан).

**Жукова Мария Александровна**, студент Нижегородского филиала Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации, г. Нижний Новгород.

**Ижбаева Гульназ Ришатовна**, кандидат филологических наук, доцент кафедры башкирского и сравнительно-сопоставительного языкознания Башкирского государственного педагогического университета им. М. Акмуллы, г. Уфа.

**Ишкильдина Линара Камиловна**, кандидат филологических наук, старший научный сотрудник отдела языкознания Института истории, языка и литературы Уфимского федерального исследовательского центра Российской академии наук, г. Уфа.

**Ишмухаметова Анита Шаукатовна**, младший научный сотрудник лаборатории лингвистики и информационных технологий Института истории, языка и литературы Уфимского федерального исследовательского центра Российской академии наук, г. Уфа.

**Кадирова Фирюза Ильфировна**, студент Казанского федерального университета, г. Казань.

**Карбозова Булбул Давлетханкызы**, докторант Института языкознания им. А. Байтурсынова Министерства образования и науки Республики Казахстан, г. Алматы (Республика Казахстан).

**Каршыгаева Айнур**, кандидат филологических наук, научный сотрудник Института языкознания им. А. Байтурсынова Министерства образования и науки Республики Казахстан, г. Алматы (Республика Казахстан).

**Кирчанов Максим Валерьевич**, доктор исторических наук, профессор кафедры регионоведения и экономики зарубежных стран Воронежского государственного университета, г. Воронеж.

**Коваль Екатерина Сергеевна**, старший лаборант-исследователь Центра инновационных языковых технологий Дома дружбы народов Республики Коми, г. Сыктывкар.

**Кожаметова Актоты Кожаметкызы**, младший научный сотрудник Института языкознания им. А. Байтурсынова Министерства образования и науки Республики Казахстан, г. Алматы (Республика Казахстан).

**Кондратьева Наталья Владимировна**, доктор филологических наук, директор Института удмуртской филологии, финно-угроведения и журналистики, г. Ижевск.

**Крижановская Наталья Борисовна**, ведущий инженер-исследователь лаборатории информационных компьютерных технологий Института прикладных математических исследований Карельского научного центра Российской академии наук, г. Петрозаводск.

**Крижановский Андрей Анатольевич**, кандидат технических наук, ведущий научный сотрудник лаборатории информационных компьютерных технологий Института прикладных математических исследований Карельского научного центра Российской академии наук, г. Петрозаводск.

**Крючкова Ольга Юрьевна**, кандидат филологических наук, заведующая кафедрой теории, истории языка и прикладной лингвистики Института филологии и журналистики Саратовского национального исследовательского государственного университета им. Н. Г. Чернышевского, г. Саратов.

**Култанбаева Нургул Калдыгулкызы**, докторант Актюбинского регионального государственного университета им. К. Жубанова Республики Казахстан, г. Актюбе (Республика Казахстан).

**Латыпов Булат Назирович**, главный редактор, заведующий отделом электронных ресурсов и научных изданий Государственного автономного учреждения науки Республики Башкортостан «Башкирская энциклопедия», г. Уфа.

**Левченко Дмитрий Александрович**, ведущий программист Центра инновационных языковых технологий Дома дружбы народов Республики Коми, г. Сыктывкар.

**Леонтьева Александра Филипповна**, младший научный сотрудник отдела филологических исследований Удмуртского института истории, языка и литературы Удмуртского федерального исследовательского центра Уральского отделения Российской академии наук, г. Ижевск.

**Лукьянов Андрей Викторович**, доктор философских наук, профессор кафедры философии и политологии Башкирского государственного университета, г. Уфа.

**Магомедов Даниял Магомедович**, научный сотрудник Института языка, литературы и искусства им. Г. Цадасы Дагестанского научного центра Российской академии наук, г. Махачкала.

**Магомедов Магомед Ибрагимович**, доктор филологических наук, профессор, главный научный сотрудник Института языка, литературы и искусства им. Г. Цадасы Дагестанского научного центра Российской академии наук, г. Махачкала.

**Миннуллина Эльвина Маратовна**, студент Казанского (Приволжского) федерального университета, г. Казань.

**Мурсал Айкерим**, магистрант Казахского национального университета им. Аль-Фараби, г. Алматы (Республика Казахстан).

**Муслу Али**, соискатель Кыргызского государственного университета им. К. Карасаева, г. Бишкек (Республика Кыргызстан).

**Мухалиев Адилбай**, аспирант Кыргызского государственного университета им. И. Арабаева, г. Бишкек (Республика Кыргызстан).

**Николаев Егор Револьевич**, кандидат филологических наук, младший научный сотрудник Института гуманитарных исследований и проблем малочисленных народов Севера Сибирского отделения Российской академии наук, г. Якутск.

**Остапов Николай Вениаминович**, кандидат исторических наук, преподаватель истории, философии и эстетики Сыктывкарского колледжа сервиса и связи, г. Сыктывкар.

**Остапова Елена Васильевна**, кандидат филологических наук, доцент кафедры финно-угорской филологии и национального образования Сыктывкарского государственного университета им. Питирима Сорокина, г. Сыктывкар.

**Партанен Нико**, докторант Государственного университета Хельсинки, г. Хельсинки (Финляндия).

**Пушкарёва Марина Алексеевна**, доктор философских наук, доцент кафедры этики, культурологии и связей с общественностью Башкирского государственного университета г. Уфа.

**Родионова Александра Павловна**, кандидат филологических наук, научный сотрудник сектора языкознания Института языка, литературы и истории Карельского федерального исследовательского центра Российской академии наук, г. Петрозаводск.

**Рюйтер Джек Майкл**, доктор философии, профессор Государственного университета Хельсинки, г. Хельсинки (Финляндия).

**Сагинова Жанат Кабиевна**, магистрант Актюбинского регионального государственного университета им. К. Жубанова Республики Казахстан, г. Актобе (Республика Казахстан).

**Садыков Ташполат Садыкович**, доктор филологических наук, профессор кафедры кыргызского языка Кыргызского государственного университета им. К. Карасаева, г. Бишкек (Республика Кыргызстан).

**Сайфуллин Шамиль Саяхович**, научный редактор Государственного автономного учреждения науки Республики Башкортостан «Башкирская энциклопедия», г. Уфа.

**Салимова Акерке**, магистрант Евразийского университета им. Гумилева, г. Нур-Султан (Республика Казахстан).

**Самситова Луиза Хамзиновна**, доктор филологических наук, профессор кафедры башкирского и сравнительно-сопоставительного языкознания Башкирского государственного педагогического университета им. М. Акмуллы, г. Уфа.

**Сиразитдинов Зиннур Амирович**, кандидат филологических наук, ведущий научный сотрудник, заведующий лабораторией лингвистики и информационных технологий Института истории, языка и литературы Уфимского федерального исследовательского центра Российской академии наук, г. Уфа.

**Сооссаар Свен-Эрик**, доктор философии, старший научный сотрудник Института эстонского языка, г. Таллинн (Эстонская Республика).

**Ставропольский Юлий Владимирович**, кандидат социологических наук, доцент кафедры общей и социальной психологии Саратовского национального исследовательского государственного университета им. Н. Г. Чернышевского, г. Саратов.

**Степанов Вячеслав Андроникович**, ведущий специалист Центра инновационных языковых технологий Дома дружбы народов Республики Коми, г. Сыктывкар.

**Суйеркул Ботагез Мырзабайкызы**, доктор филологических наук, главный научный сотрудник Института языкознания им. А. Байтурсынова Министерства образования и науки Республики Казахстан, г. Алматы (Республика Казахстан).

**Сулейманов Джавдет Шевкетович**, директор Института прикладной семиотики Академии наук Республики Татарстан, профессор кафедры информационных систем Казанского (Приволжского) федерального университета, г. Казань.

**Сычева Юлия Сергеевна**, учитель коми языка и литературы Муниципального автономного общеобразовательного учреждения Средняя школа №38, г. Сыктывкар.

**Титова Ольга Владимировна**, кандидат филологических наук, научный сотрудник отдела филологических исследований Удмуртского института истории, языка и литературы Удмуртского федерального исследовательского центра Уральского отделения Российской академии наук, г. Ижевск.

**Ткаченко Кирилл Станиславович**, инженер 1-й категории Севастопольского государственного университета, г. Севастополь.

**Тлегенова Гульден Бакытказыевна**, докторант Института языкознания им. А. Байтурсынова Министерства образования и науки Республики Казахстан, г. Алматы (Республика Казахстан).

**Торекеев Бахытжан Асылханович**, докторант Казахского университета международных отношений и мировых языков им. Абылай хана, г. Алматы (Республика Казахстан).

**Торекеева Раушан Сейсенбаевна**, учитель английского языка школы им. С. Сейфуллина, Туркестанской области Республики Казахстан (Республика Казахстан).

**Турлыбекова Альбина Нурмуханбетовна**, магистрант Актюбинского университета им. С. Баишева Республики Казахстан, г. Актобе (Республика Казахстан).

**Урдишев Кабил**, Ташкентский государственный университет узбекского языка и литературы имени Алишера Навои, г. Ташкент (Республика Узбекистан).

**Файзуллина Ляйсан Расиховна**, студент 3 курса магистратуры Башкирского государственного университета, г. Уфа.

**Федина Марина Серафимовна**, кандидат филологических наук, заместитель директора Центра инновационных языковых технологий Дома дружбы народов Республики Коми, г. Сыктывкар.

**Хайруллаева Гульшода**, Ташкентский государственный университет узбекского языка и литературы имени Алишера Навои, г. Ташкент (Республика Узбекистан).

**Хакимов Булат Эрнстович**, кандидат филологических наук, ведущий научный сотрудник Института прикладной семиотики Академии наук Республики Татарстан, г. Казань.

**Хаммалайнен Мика**, докторант Государственного университета Хельсинки, г. Хельсинки (Финляндия).

**Хузиахметова Гульнара Айдаровна**, магистрант кафедры общего языкознания и тюркологии ИФМК Казанского (Приволжского) федерального университета, г. Казань.

**Хусайнов Айдар Фаилович**, кандидат технических наук, ведущий научный сотрудник Института прикладной семиотики Академии наук Республики Татарстан, г. Казань.

**Хусайнова Лейла Мидхатовна**, кандидат филологических наук, доцент кафедры башкирского и сравнительно-сопоставительного языкознания Башкирского государственного педагогического университета им. М. Акмуллы, г. Уфа.

**Шадкам Зубайда**, доктор философии, профессор Казахского Национального университета им. Аль-Фараби, г. Алматы (Республика Казахстан).

**Шаехов Марат Рашитович**, кандидат филологических наук, научный сотрудник Института прикладной семиотики Академии наук Республики Татарстан, г. Казань.

**Шакиров Искандер Аликович**, кандидат философских наук, старший научный сотрудник отдела этнологии Института истории, языка и литературы Уфимского федерального исследовательского центра Российской академии наук, г. Уфа.

**Шамсутдинова Гульгина Газнабиевна**, младший научный сотрудник лаборатории лингвистики и информационных технологий Института истории, языка и литературы Уфимского федерального исследовательского центра Российской академии наук, г. Уфа.

**Шибанов Алексей Александрович**, кандидат филологических наук, научный сотрудник отдела филологических исследований Удмуртского института истории, языка и литературы Удмуртского федерального исследовательского центра Уральского отделения Российской академии наук, г. Ижевск.

**Юмагулова Гузалия Рафиковна**, магистрант Башкирского государственного педагогического университета им. М. Акмуллы, г. Уфа.

## СОДЕРЖАНИЕ

### НАЦИОНАЛЬНЫЕ КОРПУСЫ ЯЗЫКОВ. ПРОБЛЕМЫ КОРПУСНОЙ ЛИНГВИСТИКИ

Бускунбаева Л. А. Принципы сбора полевых материалов для диалектного корпуса башкирского языка.....	3
Гатауллин Р. Р., Гильмуллин Р. А., Хакимов Б. Э. Разрешение морфологической многозначности в корпусе татарского языка на основе машинного обучения.....	7
Гафарова В. Р., Гайнуллина Г. Ф. Проблема разграничения частеречной природы слова «якын» в татарском языке (на материале корпуса татарского языка).....	10
Жубанов А. К. Об особенностях составления корпуса разговорной речи казахского языка.....	13
Коваль Е. С. Из опыта создания газетного подкорпуса коми языка.....	17
Крижановский А. А., Крижановская Н. Б., Родионова А. П. Архитектура корпусного менеджера и разметка текстов корпуса ВепКар.....	19
Крючкова О. Ю. Участие студентов в создании лингвистического корпуса как образовательная технология.....	24
Магомедов М. И., Магомедов Д. М. Опыт создания национального корпуса аварского языка.....	27
Николаев Е. Р. Подкорпус диалектной лексики якутского языка в условиях цифровизации (постановка проблемы).....	30
Рютер Дж., Партанен Н. Новые текстовые корпуса миноритарных языков на сервере Corp.csc.fi...	32
Сиразитдинов З. А. О концепции национального корпуса башкирского языка.....	37
Тлегенова Г., Кожажметова А. О полуавтоматической метаразметке казахского корпуса.....	42
Федина М. С. Корпус коми языка как база для научных исследований.....	45

### ЛИНГВИСТИЧЕСКИЕ БАЗЫ ДАННЫХ И ПРИЛОЖЕНИЯ. ФУНКЦИОНИРОВАНИЕ ЯЗЫКА В СЕТИ ИНТЕРНЕТ

Альнайар Х., Хаммалайнен М., Партанен Н., Рюйтер Дж. Инфраструктура открытых словарей уральских языков.....	49
Андрианова И. Н, Левченко Д. А., Степанов В. А. Электронные разработки Центра инновационных языковых технологий ГАУ РК «Дом дружбы народов Республики Коми».....	52
Ароянц А. А. Функционирование литературных сообществ в глобальной сети: коммуникативный аспект.....	56
Бочкарев В. В. Машинный фонд якутского языка.....	59
Валиуллина З. Р., Лукьянов А. В., Пушкарёва М. А. Трудности машинного перевода и обработки речи.....	62
Ишмухаметова А. Ш., Шамсутдинова Г. Г. О лексикографической базе данных машинного фонда башкирского языка.....	64
Кирчанов М. В. Национальные Википедии России и проблемы визуализации национальных языков в Рунете.....	67
Латыпов Б. Н., Сайфуллин Ш. С. Из опыта создания электронных версий «Башкирской энциклопедии» на русском и башкирском языках.....	70
Миннуллина Э. М. Социальные сети как инструмент изучения языков.....	73
Сооссаар С.-Э. Электронные словари финно-угорских языков Института эстонского языка....	74
Сулейманов Д. Ш., Гатиатуллин А. Р., Гатауллин Р. Р. О разработке базы данных правил контекстно-свободных грамматик семантико-синтаксического анализатора татарского языка.....	76
Сычева Ю. С. Персональные сайты учителей коми языка и литературы с точки зрения коми литературного образования.....	79
Титова О. В. Функционирование неологизмов удмуртского языка в сети интернет.....	81
Ткаченко К. С. Управление компьютерным узлом информационного контура лингвистической базы данных для повышения его эффективности.....	84
Хусаинов А. Ф., Гильмуллин Р. А., Сулейманов Д. Ш. Обзор первой нейросетевой системы русско-татарского машинного перевода.....	87
Шаехов М. Р. Принципы построения и возможности использования Wordnet для тюркских языков.....	91
Шибанов А. А. Удмуртский язык в сети интернет.....	95

## ЭЛЕКТРОННЫЕ ОБРАЗОВАТЕЛЬНЫЕ РЕСУРСЫ И МЕТОДИКА ОБУЧЕНИЯ НАЦИОНАЛЬНЫМ ЯЗЫКАМ

Абдурахмонова Н., Хайруллаева Г., Урдишев К. Использование корпуса для обучения синонимии на узбекском языке.....	99
Бускунбаева Л. А., Ишмухаметова А. Ш., Шамсутдинова Г. Г. Машинный фонд как средство обучения башкирскому языку .....	104
Валиева Г. Д. Новые интерактивные филологические технологии в современном информационном пространстве.....	107
Гейт М. В. Проблема оттеснения родных языков в виртуальном социуме.....	110
Жукова М. А. Цифровые образовательные ресурсы: проблемы и перспективы.....	112
Ижбаева Г. Р. Цифровизация образования и новые технологии в обучении родным языкам .....	115
Ильина Н. В., Леонтьева А. Ф. Возможности использования корпусной лингвистики в практике преподавания удмуртского языка.....	118
Кадилова Ф. И. Изучение фразеологизмов в процессе обучения иностранному языку в начальном классе.....	121
Кондратьева Н. В. Междисциплинарные образовательные проекты как условие развития национальных языков.....	123
Остапова Е. В. Изучение информационных технологий в вузе для профессиональной деятельности учителя родного языка и литературы.....	125
Остапов Н. В. Мультимедийные учебные пособия по истории России и обществознанию на коми языке.....	127
Ставропольский Ю. В. Прогресс компьютерного перевода и полемика о преподавании английского языка в Японии.....	130
Торекеев Б. А., Торекеева Р. С., Шадкам З. Цифровизация как условие формирования новых компетенций педагогов.....	132
Файзуллина Л. Р. Лингвокультурологический подход в преподавании.....	137
Хузиахметова Г. А., Габдрахманова Ф. Х. Способы сохранения культурного наследия.....	139
Шакиров И. А. Использование корпусных данных в изучении лексики.....	141

## АКТУАЛЬНЫЕ ПРОБЛЕМЫ СОВРЕМЕННОЙ ЛИНГВИСТИКИ

Абдирова Ш. Г. Организация текстов призывов.....	144
Абдирова Ш. Г., Абдигази С. Х. Особенности казахоязычного идеологического дискурса .....	147
Буранкулова Э. Т. Особенности пословиц и поговорок гендерного значения (на материале казахского и башкирского языков).....	150
Даржинова Л. В. Некоторые особенности лингвистической ситуации в Республике Калмыкия .....	153
Дюсенов Б. Д. Фонетическая система словаря «Китаб ад-Дурра».....	156
Ералханкызы М. Фразеологические единицы со значением «радость» в русском языке .....	160
Жанабекова А., Салимова А. Сравнительный анализ частотных словарей казахского языка .....	163
Ишкильдина Л. К. Языковые особенности башкир Салаватского района РБ.....	171
Карбозова Б. Д., Жубанов А. К. Лексико-семантические классы существительных и глаголов казахского языка.....	174
Каршыгаева А. Значения звуков заимствованных букв казахского алфавита .....	183
Кирчанов М. В. «Сибирский» и «новгородский» языки в современной России как языки, скрипты или артланги.....	187
Култанбаева Н. К. О понятии времени и пространства в сказочном дискурсе (на примере казахских и русских народных сказок).....	190
Мухалиев А. Состав и варианты звуков в казахском языке.....	193
Сагинова Ж. К. Психолингвистические концепции обучения и использования иностранного языка .....	196
Садыков Т., Муслу А. Об истории тюркской фонологии и морфонологии .....	199
Самситова Л. Х., Хусаинова Л. М. Правописание заимствованных слов в башкирском языке. 206	
Суйеркул Б. М., Мурсал А. Теолингвистический анализ сложных глаголов, образованных с участием теонимов в современном казахском языке.....	209
Турлыбекова А. Н., Амангазиева М. К. Проблемы когнитивной лингвистики в обучении казахскому языку.....	214
Юмагулова Г. Р. Концепт «религия» в романе Рината Камала «Таня-Танхылыу».....	216

*Научное издание*

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

Сборник статей

Под общей редакцией  
члена-корреспондента АН РБ *А. И. Акманова*

Редактор: *Ахмадеева А.У.*

Подписано в печать 23.12.2019. Формат 60x84 <sup>1</sup>/<sub>8</sub>. Бумага офисная «Снегурочка».  
Гарнитура «Таймс». Печать на ризографе. Усл. печ. л. 26,04. Уч.-изд. л. 24,27.  
Тираж 500 экз. Заказ № 33.

ГАУН РБ «Башкирская энциклопедия».  
450006, г. Уфа, ул. Революционная, 55. Тел.: (347) 250-06-72.  
<http://www.bashenc.ru>  
E-mail: [gilem@bashenc.ru](mailto:gilem@bashenc.ru)

Отпечатано в типографии  
ООО «Информационно-просветительский центр “Башакадемкнига”».  
450006, г. Уфа, ул. Революционная, 55