

Калмыцкий научный центр РАН
Институт системного программирования
им. В. П. Иванникова РАН

**Платформа LingvoDoc:
цифровизация языков народов России
и инструменты для исследования:
Тезисы докладов Круглого стола
(г. Элиста, 13 декабря 2025 г.)**

Элиста
2025

УДК 811.512.3
ББК 81.11 + 81.116
П 37

Электронное издание

Утверждено к печати
Ученым советом Федерального государственного
бюджетного учреждения науки
«Калмыцкий научный центр Российской академии наук»

Подготовка и издание осуществлены при поддержке РНФ в рамках проекта «Разработка инструментария и комплексные исследования монгольских языков и их диалектов (с применением технологий анализа больших массивов данных словарных и корпусных материалов)» (№ 25-78-20008).

Ответственный редактор:
В. В. Куканова

К84 Платформа LingvoDoc: цифровизация языков народов России и инструменты для исследования: Тезисы докладов Круглого стола (г. Элиста, 13 декабря 2025 г.). – Элиста: КалмНЦ РАН, 2025. – 51 с.

ISBN 978-5-907659-41-4

В настоящем издании представлены тезисы докладов и сообщений, которые прозвучали на Круглом столе «Платформа LingvoDoc: цифровизация языков народов России и инструменты для исследования» (г. Элиста, 13 декабря 2025 г.).

УДК 811.512.3
ББК 81.11 + 81.116

ISBN 978-5-907659-41-4

© КалмНЦ РАН, 2025

ПРОГРАММА

Место проведения: Малый зал КалмНЦ РАН, 3 эт.

Дата и время проведения: 13 декабря 2025 г., начало в 11:00

Открытие заседания Круглого стола

*Куканова Виктория Васильевна, канд. филол. наук, директор
Калмыцкого научного центра РАН*

Доклады и сообщения

Дыбо Анна Владимировна, д-р филол. наук, профессор, член-корреспондент РАН, Институт языкознания РАН
Диалектные наименования деревьев и кустарников в якутском языке: лингвогеографический аспект (на основе лингвистической платформы «ЛингвоДок»)

Норманская Юлия Викторовна, д-р филол. наук, Институт системного программирования им. В. П. Иванникова РАН
Степень близости между дунсянским, баоаньским и шираюгурским языками с точки зрения глоттохронологии и системы регулярных фонетических соответствий

*Куканова Виктория Васильевна, канд. филол. наук, директор,
Калмыцкий научный центр РАН*

*Баянова Александра Тагировна, канд. филол. наук, ст. науч. сотр.,
Калмыцкий научный центр РАН*

Сарт-калмыцкий и калмыцкий язык: фонетические соответствия (на материале полевых записей)

Нуриева Фануза Шакуровна, д-р филол. наук, профессор,
Казанский (Приволжский) федеральный университет;
Галиуллина Гульшат Раисовна, д-р филол. наук, профессор,
Казанский (Приволжский) федеральный университет;
Юсупов Айрат Фаикович, д-р филол. наук, доцент, Казанский
(Приволжский) федеральный университет
Графо-фонетическое описание материалов татарского языка
XVIII в. на платформе ЛингвоДок: по записям участников
Великой Северной экспедиции

Галиуллина Гульшат Раисовна, д-р филол. наук, профессор,
Казанский (Приволжский) федеральный университет;
Юсупов Айрат Фаикович, д-р филол. наук, доцент, Казанский
(Приволжский) федеральный университет;
Нуриева Фануза Шакуровна, д-р филол. наук, профессор, Ка-
занский (Приволжский) федеральный университет
*Сравнительный анализ лексики словника Г. Ф. Миллера и совре-
менного татарского языка*

Гончарова Оксана Владимировна, канд. филол. наук, доцент,
Российский университет дружбы народов им. П. Лумумбы
*Автоматизация этимологического анализа: алгоритмические и
нейросетевые методы*

Муратова Римма Талгатовна, канд. филол. наук, ст. науч. сотр.,
Институт истории, языка и литературы Уфимского федераль-
ного исследовательского центра Российской академии наук;
Вильданов Ильдар Радикович, мл. науч. сотр., Институт истории,
языка и литературы Уфимского федерального исследователь-
ского центра РАН
Названия для мясного бульона и супа в тюркских языках: поиск
этимологий и картографирование

Музраева Деляш Николаевна, д-р. ист. наук, доцент, вед. науч.
сотр., Калмыцкий научный центр РАН

*К изучению языка письменных источников на ойратском языке
(на материале параллельного тибетско-ойратского корпуса
текстов)*

Тушинов Баир Луданович, канд. ист. наук, науч. сотр., Калмыцкий научный центр РАН

Хадмал-тексты как прототип современного тибетско-ойратского параллельного корпуса

Долева Айса Олеговна, мл. науч. сотр., Калмыцкий научный центр РАН

«Море притч» (üligeriindalai) как источник изучения ойратской лексики XVIIв. (по материалам параллельного тибетско-ойратского корпуса)

Танаев Владислав Владимирович, мл. науч. сотр., Калмыцкий научный центр РАН

Применение методов машинного обучения для векторизации тибетского и ойратского текстов

Ковылин Сергей Васильевич, канд. филол. наук, ст. науч. сотр., Институт системного программирования им. В. П. Иванникова РАН

Опыт цифровой обработки селькупских и хантыйских языковых материалов

Манджиева Инна Бембеевна, мл. науч. сотр., Калмыцкий научный центр Российской академии наук

Записи дагурского языка начала и конца XX в.: фонетические соответствия

Степень близости между дунсянским, баоаньским и шира-югурским языками с точки зрения глоттохронологии и системы регулярных фонетических соответствий¹

Вопрос о классификации монгольских языков распространенных в провинции Китая Ганьсу: баоаньского (носители живут в провинциях Ганьсу и Цинхай), дунсянского (провинция Ганьсу и Дунсянский автономный округ), шира-югурского (провинция Ганьсу) многократно по-разному решался учеными. Даже за последние десятилетия высказаны три различные точки зрения. Степень близости этих языков также является предметом дискуссий. С одной стороны, с точки зрения глоттохронологии авторы считают, что они имеют только 4-5% различий в базисной лексики, с другой, они весьма значительно отличаются фонетически. Ю. Янхунен высказывал точку зрения, что эти языки являются примером «стремительных» изменений. Возможно ли это? Или время их распада, в действительности, было гораздо раньше? Материалы и методы. Для анализа были взяты большие словари по дунсянскому языку (11 914 слов), по баоаньскому (5079 слов), по шира-югурскому (3 618 слов). Эти словарные базы данных созданы З. И. Чушкаевой на основе китайских печатных изданий. Они обчисляются с помощью формулы глоттохронологии, разработанной С. А. Старостиным. В результате их анализа на ЛингвоДоке строится график степени близости времени распада языков и диалектов в 2D и 3D-формате с точки зрения времени распада. Было установлено, что с точки зрения глоттохронологии и уточненным значения слов из базисной лексики по китайским источник можно говорить о том, что шира-югурский и дунсян-

¹ Исследование проведено при финансовой поддержке РФФИ в рамках проекта «Возможности искусственного интеллекта для сравнительно-исторического изучения малоресурсных языков народов РФ» (№ 25-78-20002).

ский языки разделились около 1 500 лет назад, поскольку они имеют более 20 % различий в базисной лексике. Интересно, что это имеет определенную корреляцию с историческими фактами. В конце IV – начале V вв. на севере «варварских государств» на севере Китая стало уже 19, и 8 из них были сяньбийскими, что говорит о разделении монгольских племен в Китае на разные государства, как раз 1 500 лет назад.

*В. В. Куканова,
А. Т. Баянова
Калмыцкий научный центр РАН*

**Сарт-калмыцкий и калмыцкий язык:
фонетические соответствия
(на материале полевых записей)¹**

1. Введение

В 2025 г. в рамках проекта, поддержанного Российским научным фондом, была осуществлена экспедиция в Киргизию, в Иссык-Кульский район, для записи информантов — носителей сарт-калмыцкого говора. Сарт-калмыки проживают компактно в селах Челпек, Бёрю-Баш, Бурма-Суу, Таш-Кыя, при этом села Челпек, Бурма-Суу, Таш-Кыя стали частью города Карапалкак. Село Бёрю-Баш пока еще не включено в состав города.

В 1989 г. калмаков в Иссык-Кульской области насчитывалось 4 593 чел., в 1999 г. — 5 314 чел., в 2009 г. — 3 801 чел. [Перепись 2009: 48]. В ряде источников их численность оценивается в пределах от 12 тыс. до 20 тыс. человек [Нанзатов, Содномпилова 2012: 28]. Согласно полевым материалам, полученным в ходе экспедиции, почти все население четырех сел Челпек, Бурма-Суу, Таш-Кыя и Бёрю-Баш составляют сарт-калмаки, а это почти 9 тыс. чел. По переписи 2009 г. в с. Челпек проживало 3 157 чел., в с. Бурма-Суу — 1 939 чел., в с. Таш-Кыя — 1 802 чел., в Бёрю-Баш — 2 100 чел. [Перепись 2009: 220]. Такое несоответствие объясняется тем, что большинство калмаков в паспортах записаны киргизами [ПМА]. Кроме того, в других селах, например, в с. Нарыне, Черике, также проживают калмаки, которые записаны как киргизы, по мнению информантов.

¹ Исследование проведено при финансовой поддержке РФФ в рамках проекта «Разработка инструментария и комплексные исследования монгольских языков и их диалектов (с применением технологий анализа больших массивов данных словарных и корпусных материалов)» (№ 25-78-20008).

Сарт-калмыцкий язык практически исчез, все молодое поколение говорит свободно на киргизском языке, некоторые еще и на русском. Старшее поколение может продуцировать отдельные простые предложения, причем достаточно простые по содержанию и структуре. Отметим также, что письменности и, следовательно, каких-то произведений на сарт-калмыцком языке нет, что во многом влияет на витальность языка.

2. Материалы

В докладе мы затронем только те особенности, которые не получили освещения в работе [Дыбо и др. 2025]. Результаты носят предварительный характер, после расширения материала транскрипции и записи информантов можно будет говорить более подробно об особенностях языка сарт-калмыков. Материалом для доклада выступили полевые записи от одного информанта общим объемом 1,5 часа. Записанные слова на сарт-калмыцком языке сопоставлялись с калмыцким материалом [КРС 1977], для сравнения и выявления архаичных или инновационных черт выступили реконструкции, произведенные Х. Нугтереном [Nugteren 2011]. Запись сарт-калмыцких слов дается в транскрипции МФА, впоследствии после обработки в программе Praat будет уточнена.

3. Особенности сарт-калмыцкого языка в области вокализма

3.1. В современном калмыцком языке отсутствует губная гармония в пределах одного слова, напротив, сарт-калмыцкий язык достаточно отчетливо демонстрирует действие губной гармонии. Как правило, если в первом слоге сарт-калмыцкого слова произносится лабиализованный звук, то в последующих слогах будет представлен подобный звук. Однако в словах, где в финале ранее имелся *a_i, отсутствует гармония гласных, хотя в некоторых ойратских диалектах отмечается уподобление гласному первого слога по лабиализованности. Например: *поха* ‘собака’, *това* ‘змея’.

Перевод	Калмыцкий	Сарт-калмыцкий	Реконструкция
‘лепешки’	<i>bortʃrk</i>	<i>bortʃuk</i>	
‘вчера’	<i>øtsktəɖər</i>	<i>øtskøɖur</i>	*öčügedür
‘одежда’	<i>xuvʃɣn</i>	<i>kʰubʃeu</i>	*kubčasun

3.2. о — у. В монгольских языках действовал и в некоторых языках продолжает действовать закон сингармонизма и по огубленности, и по ряду [Санжеев 1953: 83]. Данная черта [Кичиков 1963: 4; Убушаев 1979: 31] относится к диагностическим признакам дербетского и торгутского диалектов калмыцкого языка. В сарт-калмыцком языке у восходит к *и, является архаизмом, поэтому не может рассматриваться как признак данного диалекта.

Перевод	Калмыцкий	Сарт-калмыцкий	Реконструкция
‘крутит’	<i>moʃkulna</i>	<i>muʃkulna</i>	*muʃki-

3.3. а — ә. В полевых записях было обнаружено множество примеров с упереждением гласных, причем этот процесс, по-видимому, связан с регрессивной ассимиляцией под влиянием гласной второго или последующих слогов.

Перевод	Калмыцкий	Сарт-калмыцкий	Реконструкция
‘яблоко’	<i>aʃmɣn</i>	<i>æʃim</i>	*alīma

3.4. е — а. Данный переход встречается в монгольском языке [EDAL 2003: 152], в других не обнаружен. Для калмыцкого языка указанная инновация не характерна.

Перевод	Калмыцкий	Сарт-калмыцкий	Реконструкция
‘открывает’	<i>seknæ</i>	<i>sakna</i>	

3.5. і — ø. В сарт-калмыцком языке данная черта относится к архаичным: восходит к *ö.

Перевод	Калмыцкий	Сарт-калмыцкий	Реконструкция
‘лягает’	<i>iʃkɣʃnæ</i>	<i>øʃkøʃnæ</i>	*öskel- (?hisköl-)

3.6. e — ø. Данная черта является распространенным явлением в калмыцком языке, в настоящее время сохранилось подобное произношение среди дербетов, однако материалы словарей показывают, что такой переход носил наддиалектный характер, так как отмечен в словарях и исторических работах, в которых зафиксирован язык калмыков не только в регионах, где традиционно проживали дербеты-калмыки. В сарт-калмыцком языке данное соответствие относится к инновационным.

Перевод	Калмыцкий	Сарт-калмыцкий	Реконструкция
‘сегодня’	<i>endər</i>	<i>əndər</i>	*ene üdür

4. Особенности сарт-калмыцкого языка в области консантизма

4.1. t — d. Подобный переход встречается в ордосском, монгорском, шира-югурском и дунсянском языках [EDAL 2003: 151]. Для калмыцкого языка не характерен данный переход.

Перевод	Калмыцкий	Сарт-калмыцкий	Реконструкция
‘арбуз’	<i>tarvʏs</i>	<i>darbus</i>	—
‘живой’	<i>æmdə</i>	<i>æmtə</i>	*amīdu

4.2. d — t. Здесь отмечается два процесса: переход *d в t и сохранение *t. В калмыцком языке в первом случае происходит сохранение исконной *d, а во втором случае — озвончение исконной *t.

Перевод	Калмыцкий	Сарт-калмыцкий	Реконструкция
‘мочевой пузырь’	<i>davʏʁk</i>	<i>tavʏʁʁk</i>	*dabasuj
‘вместе’	<i>xamdan</i>	<i>xamtan</i>	*kamtu

4.3. s — ɛ. Такой переход не зафиксирован в [EDAL 2003: 151]. Такое произношение встречается среди ойратских диалектов Синьцзяна, в частности у олетов.

Перевод	Калмыцкий	Сарт-калмыцкий	Реконструкция
‘кошка’	<i>mis</i>	<i>miɛ</i>	—

4.4. tɛ — ɛ. Данный переход не встречается среди ойратских диалектов. Возможно, локальная особенность языка сарт-калмыков.

Перевод	Калмыцкий	Сарт-калмыцкий	Реконструкция
‘костный мозг’	<i>teimɣ</i>	<i>eiɣɣ</i>	*čimegen (?čimögen)

4.5. ts — ʂ. Исконное *č переходит в ʂ в сарт-калмыцком языке. Данное явление можно отнести к инновационным, при этом такой же переход зафиксирован в бурятском языке.

Перевод	Калмыцкий	Сарт-калмыцкий	Реконструкция
‘лает’	<i>xutsna</i>	<i>xoʂna</i>	*kuča-

4.6. b — m. В сарт-калмыцком языке согласный b переходит в m, такое явление было характерно и для калмыцкого языка, например: *möiʂr*. T = Ö bitʂr, bütʂr, D bütʂr ‘ветка’ [Ramstedt 1935: 267].

Перевод	Калмыцкий	Сарт-калмыцкий	Реконструкция
‘печка’	<i>beʂə</i>	<i>meʂə</i>	–
‘картофель’	<i>bodɣntʂɣ</i>	<i>moduntɕuk</i>	–

4.7. b — p. Переход b → p в начале слова происходит, думается, перед и после узкой гласной. Такой переход, видимо, связан с придыхательностью исконной *b.

Перевод	Калмыцкий	Сарт-калмыцкий	Реконструкция
‘пишет’	<i>bitɕə</i>	<i>pitɕə</i>	*biči-
‘нельзя’	<i>biteɕə</i> <i>bitekə</i>	<i>pitekə</i>	*bütegei

4.8. b — p. Данное явление характерно было и для калмыцкого языка, что отмечал еще Г. Й. Расмтедт: сказки в его записи изобилуют такими формами [Калмыцкие сказки 2018; Куканова 2016: 114].

Перевод	Калмыцкий	Сарт-калмыцкий	Реконструкция
‘лежит’	<i>kebtɕə</i>	<i>keptɕə</i>	
‘рожок’	<i>uvdɕə</i>	<i>ɣpteø</i>	

4.9. r — z. Данный переход обусловлен, скорее всего, влиянием киргизского языка, для которого характерен подобный процесс [EDAL 2003: 143].

Перевод	Калмыцкий	Сарт-калмыцкий	Реконструкция
‘мастер’	<i>urun</i>	<i>uzda</i>	*uran

‘коренной зуб’	<i>aran</i>	<i>azu</i>	*aḡia
-------------------	-------------	------------	-------

4.10. *ɪ — j* Аналогичный переход отмечен и в работах Н. Н. Убушаева, Б. Х. Борлыковой, Б. В. Меняева, однако указывается, что соответствие встречается только в начале слова [Убушаев 2006: 41; Борлыкова, Меняев 2020: 64].

Перевод	Калмыцкий	Сарт-калмыцкий	Реконструкция
‘плачет’	<i>u:tna</i>	<i>u:jna</i>	*uila-
‘плеть’	<i>matæ</i>	<i>majæ</i>	*mīlaa

4.11. *g / k — ɸ*. Данный переход встречается в шира-югурском и дунсянском языках [EDAL 2003: 151]. В сарт-калмыцком языке является инновацией.

Перевод	Калмыцкий	Сарт-калмыцкий	Реконструкция
‘недоуздок’	<i>noḡtə</i>	<i>noɸtə</i>	*noḡta
‘друг’	<i>nøkəɾ</i>	<i>nøɸt</i>	*nöker

5. Заключение

Полевой материал, полученный в ходе экспедиции, позволяет поставить вопрос о статусе сарт-калмыцкого языка. С одной стороны, язык сарт-калмыков сближается с цаатанским говором калмыцкого языка по ряду диагностических признаков, но в то же время сближается и с другими диалектами калмыцкого языка; с другой стороны, в нем достаточно много фонетических отличий от калмыцкого языка, что не позволяет нам отнести его однозначно к диалекту калмыцкого языка. Очевидно, что язык каракольских калмыков является одним из диалектов ойратского языка, только при расширении материала по его диалектам можно установить его языковую близость с тем или иным диалектом ойратов.

Литература

- EDAL 2003 — Starostin S. A., Dybo A. V., Mudrak O. A. An Etymological Dictionary of Altaic Languages. Leiden: Brill, 2003. 1556 p. (In Eng.)
 Nugteren 2011 — Nugteren H. Mongolic Phonology and the Qinghai-Gansu Languages. Utrecht: LOT, 2011. 563 p. (In Eng.)
 Борлыкова, Меняев 2020 — Борлыкова Б. Х., Меняев Б. В. О некоторых

- фонетических особенностях сарт-калмыцкого языка // Известия РАН. Серия литературы и языка. 2020. Т. 79. № 4. С. 58–66.
- Дыбо и др. 2025 — *Дыбо А. В., Грунтов И. А., Мушаев В. Н.* Некоторые фонетические особенности языка каракольских калмыков: 50 лет лингвистического изучения // *Oriental Studies*. 2025. Т. 18. № 3. С. 695–719. DOI: 10.22162/2619-0990-2025-79-3-695-719
- Калмыцкие сказки 2018 — *Калмыцкие сказки в записи Г. Й. Рамстедта: тексты и исследования. В 3 ч. Ч. II: Переложение и перевод / сост. А. Т. Баянова, Б. Б. Горяева, А. О. Долеева, В. В. Куканова; пер. на русский язык Т. А. Михалевой, Б. Б. Горяевой, А. Т. Баяновой, А. О. Долеевой, В. В. Кукановой.* Элиста: КалмНЦ РАН, 2018. 621 с.
- Кичиков 1963 — *Кичиков А. Ш.* Дербетовский говор (фонетико-морфологическое исследование). Элиста: Калмгосиздат, 1963. 88 с.
- КРС 1977 — *Калмыцко-русский словарь / отв. ред. Б. Д. Муниев.* М.: Русский язык, 1977. 768 с.
- Куканова и др. 2016 — *Куканова В. В., Горяева Б. Б., Баянова А. Т., Долеева А. О.* Фонетические явления и процессы калмыцкой речи начала XX в. (на примере сказок, записанных Г. И. Рамстедтом) // *Известия Дагестанского государственного педагогического университета. Общественные и гуманитарные науки.* 2016. Т. 10. № 4. С. 109–118.
- Нанзатов, Содномпилова 2012 — *Нанзатов Б. З., Содномпилова М. М.* Сарт-калмаки в современном Кыргызстане // *Культурное наследие народов Центральной Азии. Вып. 3: сб. ст. Улан-Удэ: Бурятский научный центр СО РАН,* 2012. С. 28–49.
- Перепись 2009 — *Перепись населения и жилищного фонда Кыргызской Республики 2009 года. Книга III: Регионы Кыргызстана. Иссык-Кульская область. Бишкек: [б. и.], 2010. 246 с.*
- ПМА — *Полевые материалы В. В. Кукановой, А. Т. Баяновой.*
- Санжеев 1953 — *Санжеев Г. Д.* Сравнительная грамматика монгольских языков. М.: Наука, 1953. 240 с.
- Убушаев 1979 — *Убушаев Н. Н.* Фонетика торгутского говора калмыцкого языка. Элиста: Калм. кн. изд-во, 1979. 196 с.
- Убушаев 2006 — *Убушаев Н. Н.* Диалектная система калмыцкого языка. Элиста, КИГИ РАН, 2006. 255 с.

Ф. Ш. Нуриева,
Г. Р. Галиуллина,
А. Ф. Юсупов,
Казанский (Приволжский) федеральный университет

Графо-фонетическое описание материалов татарского языка XVIII в. на платформе LingvoDoc: по записям участников Великой Северной экспедиции¹

В современной отечественной лингвистике одной из признанных платформ является платформа LingvoDoc. На данной платформе есть возможность создания, хранения и анализа словарей, материалов по всем языкам и диалектам народов Российской Федерации.

Реализуя новые тренды в отечественной науке, мы присоединились к научной группе пользователей платформы LingvoDoc. Нами размещены в онлайн доступе ранее неисследованные материалы Г. Ф. Миллера и Й. Э. Фишера по татарским диалектам XVIII в., которые загружены в Корпус письменных памятников татарского языка.

Ученые-востоковеды Г. Ф. Миллер (1733–1743) и Й. Э. Фишер (1739–1747) были участниками Великой Северной экспедиции. В 1733 г. Г. Ф. Миллер прибыл в Казанскую губернию (1733), где приступил к изучению языка, жизни и быта поволжских народов. Результатом их пребывания в пределах губернии стали записки, положенные в основу его работ «Описания трех языческих народов в Казанской губернии», в приложении разместил словарные записи на кириллице [Миллер 1791: 81–99]. В фонде Российского государственного архива древних актов сохранились две рукописные словари с материалами татарского языка: № 1125 (портфель Миллера 199) с делами 513 и 513 на латинской графике. Все три словаря Г. Ф. Миллера обчисланы с помощью инструментария платформы LingvoDoc, на ней же

¹ Результаты получены с использованием услуг Центра коллективного пользования Института системного программирования им. В.П. Иванникова РАН — ЦКП ИСП РАН

строится график степени близости словарей к современному татарскому литературному языку и его диалектам.

В процессе графо-фонетического анализа татарского материала словарей выявлены следующие особенности: 1) все процессы перехода в области согласных в татарском языке в XVIII в. уже завершились; 2) перелом поволжских гласных к этому периоду еще для всех гласных не завершился, а для [*o] и [*u] данный процесс даже не начался; 3) в XVIII в. в татарском языке были уже выделены диалектные зоны; 4) словарь 1791 г. на кириллической графике отличается инновациями от современного татарского литературного языка и словарей на латинской графике; 5) словари на латинской графике более близки к современному татарскому литературному языку [Нуриева 2025: 191].

Й. Э. Фишер участвовал в Великой Северной экспедиции в 1739–1747 гг. Рукописные словарные материалы ученого, которые ранее не были известны исследователям, разместил в свой сборник трудов «Ältere westeuropäische quellen zur kalmückischen sprachgeschichte» [Дерфер 1965: 40–182] немецкий тюрколог, алтаист Герхард Дерфер. В словаре зафиксированы материалы по семи говорам татарского языка, которые размещены в отдельные столбцы: Bucharice; Tatarorum [obiumTschulim]; Tatarorum [Tschat]; Tatarorum Cazanensium [aburbenCafan]; Tatarice [TatarorumTobolensium]; Tatarorum Kuznezensium [ad Tomum, Tschumysch]; Tatarorum Katschensium&Krasnojariensi]. Нами исследован словарь «Tatarorum Cazanensium [aburben Cafan]».

Графо-орфографический анализ материала словаря Й. Э. Фишера позволяет сделать следующие выводы: 1) все процессы перехода в области согласных в XVIII в. уже завершились; 2) перелом поволжских гласных к этому периоду еще для всех гласных не завершился, а для [*o] и [*u] данный процесс даже не начался; 3) в XVIII в. в татарском языке были уже выделены диалектные зоны. Ниже приводятся некоторые сравнительные данные из словарей указанных авторов.

[o]		[o]	
on on-şun	'десять десять тыс...	on-şung on tjumen	'десять тысяч деся...
on-bärg	'одинадцать'	on-bir	'одинадцать'
on-ika	'двенадцать'	on-iky on-ike	'двенадцать двенад...
ot-ujs	'тридцать'	otos otus	'тридцать тридцать'

[i]		[e]	
ira	'утро'	erten erte ertiä	'утро утро утро'
işhek	'дверь'	eshik	'дверь'
it	'мясо'	et	'мясо'
işärek	'льнян'	esryk	'льняной'
illi	'пятьдесят'	eli	'пятьдесят'

Выявлен системный переход по анлаутному гласному [i] в словнике Й. Э. Фишера, а в словаре Г.Ф. Миллера еще сохраняется древнетюркское [*e], который показывает перелом происходил в разных диалектных зонах неравномерно.

Анализ графических и фонетических особенностей позволяет утверждать, что в XVIII в. в татарском языке были уже выделены диалектные зоны; можем датировать определенные фонетические процессы, происходящие в диалектах татарского языка.

Литература

- Миллер 1791 — *Миллер Г. Ф.* Описание живущих в Казанской губернии языческих народов, яко то черемис, чуваш и вотяков: с показанием их жительства, политического учреждения СПб.: Имп. Акад. наук, 1791, 99, [2] с., [8] л.
- Нуриева 2025 — *Нуриева Ф. Ш., Галиуллина Г. Р., Юсупов А. Ф.* Словари Г.Ф. Миллера как источник для изучения истории диалектов татарского языка (на материале платформы LingvoDoc) // *Oriental Studies*. 2025. Т. 18. № 1. С. 191–211.

¹ В таблицах на первом столбце размещены примеры из словаря Й.Э. Фишера, во втором — перевод на русский язык; третий столбец — примеры из словаря Г.Ф. Миллера (1759), четвертый столбец — перевод на русский язык.

*Г. Р. Галиуллина,
А. Ф. Юсупов,
Ф. Ш. Нуриева,
Казанский (Приволжский) федеральный университет*

Сравнительный анализ лексики словника Г. Ф. Миллера и современного татарского языка¹

История фиксации, накопления материалов по диалектам татарского языка начинается с середины XVIII в., когда шло выявление элементов духовной и материальной культуры и одновременно накопление лексического материала по отдельным зонам расселения татар различными деятелями, учеными-историками, такими как Г. Ф. Миллер, И. Георги, Ф.-И. Страленберг, П. Паллас и др. Следует отметить, что в этом аспекте большое научное значение имеет словарь Г. Ф. Миллера «Описание живущих в Казанской губернии языческих народов, яко то черемис, чуваш и вотяков: с показанием их жительства, политического учреждения...» [Миллер 1791], который в татарском языкознании до сих пор не получил комплексного научного освещения. Актуальность исследования определяется возрастающим интересом к исторической лингвистике и необходимостью сохранения языкового наследия. Словарь Г.Ф. Миллера содержит ценнейшие материалы по языкам народов Поволжья XVIII века.

В настоящей статье представлены первичные результаты сравнительного анализа лексики, зафиксированной в словнике Г. Ф. Миллера XVIII в. и современного татарского языка. Источниковую базу составляет указанный словарь исследователя, размещенный на платформе LingvoDoc, словари и корпусные данные современного татарского языка, а также для этимологических уточнений, сравнения с пратюркскими корнями мы использовали базу данных СтарЛинг [СтарЛинг].

¹ Результаты получены с использованием услуг Центра коллективного пользования Института системного программирования им. В. П. Иванникова РАН — ЦКП ИСП РАН.

По Русск.	По Татарск.	По Черемисск.	По Чувашск.
Водуль. =	Давла	Пушь Меленитъ	Сналь
Вода. -	Суу	Пушь	Шина, Шина
Земля. -	Давра	Меленитъ	Сирь, Сирь
Гора. -	Аш	Кургуль	Туль
Вугора. -	Иш-Аша	Иш-Кургуль	Конш-Туль
Поля. -	Басуу	Нурь	Он
Степь. -	Кюрю	Ирь	Гарь
Дорога. -	Давла	Карна	Сюдь
Аш. -	Урмайтъ	Кашлакъ	Вурмайтъ
Дерево. -	Агашъ	Иу	Ишусь
Древо. -	Ошашъ		Водла
Море. -	Денгизъ	Тенгизъ	Тингизъ, Дингизъ
Сыр. -	Куль	Ерь	Кюль
Рѣка большая	Илея	Инь	{ Алая, Сирь
малая. -	Суу	Иньгиръ	
Источникъ	Чырма	Памашъ	Сельва
Колодець.	Кую	Мюрбъ	Посъ

Словарь Г. Ф. Миллера «Описание...» включает 325 лексем, представленных в рамках лексико-тематических групп о явлениях природы, металлах и минералах, животных и птицах, названиях органов и частей человеческого тела.

Основной целью данного исследования является выявление семантических трансформаций слов лексико-тематической группы «Водные объекты», зафиксированных в словаре «Описание...» [Миллер 1791: 84].

В словаре представлены следующие лексические единицы из данной тематической группы:¹

Денгизъ	Море	Дингез	teñiz (МК I)
Куль	Озеро	Күл	köl (Orkh., OUygh.)

¹ В таблицах на первом столбце размещены примеры из словаря Г. Ф. Миллера, во втором — перевод на русский язык; третий столбец — примеры на современном татарском литературном языке; четвертый — древнетюркский по данным СтарЛинг и ДТС.

Идель ссу	Река большая малая	Елга, зур су, Идел	šub (Orkh.), sub, suv (OUygh.)
Чишма	Источник	Чишмә	персидское заимствование
Кую	Колодезь, колодец	Кое	qud- 1(OUygh.), quduy 3 (OUygh.)

Примеры показывают, что лексемы в целом сохранили свои первоначальные значения, зафиксированные в древнетюркской эпохе.

Лексемы *денгись* ‘диңгез’, *куль* ‘күл’, *кую* ‘кое’ претерпели лишь фонетические изменения, а семантика слов сохранилась без изменений до настоящего времени.

Вызывает интерес употребление лексемы *Идель ссу* ‘елга, зур су’ в значении ‘большая вода’. Еще в словаре М. Кашгари лексема *Etil* зафиксирована в качестве географического наименования [ДТС 1969: 187]. В современной татарской разговорной, а также в диалектной речи значение ‘большая вода’ сохранилось, хотя в современном татарском литературном языке слово *Идел* ‘Волга’ употребляется только в качестве гидронима.

Функционирующая в разговорной речи татар XVIII в. заимствованная лексема из персидского языка *чишма* ‘чишмә’, сохранила свою внешнюю и внутреннюю формы до современного татарского литературного языка.

Сравнительный анализ лексики словника Г. Ф. Миллера и современного татарского языка позволил нам выявить, что лексемы тематической группы «Водные объекты» сохранили номинативные значения, восходящие к древнетюркской эпохе, лишь претерпев некоторые фонетические изменения.

Литература

- ДТС 1969 — Древнетюркский словарь. Л.: Наука, ЛО, 1969. 676 с.
Миллер 1791 — *Миллер Г. Ф.* Описание живущих в Казанской губернии языческих народов, яко то черемис, чуваш и вотяков: с показанием

их жительство, политического учреждения ... СПб.: Имп. Акад.
наук, 1791. 99, [2] с., [8] л.

СтарЛинг — СтарЛинг. Тюркская этимология // <https://starlingdb.org/>
(дата обращения: 11.11.2025).

*О. В. Гончарова,
Российский университет дружбы народов им. П. Лумумбы*

Автоматизация этимологического анализа: алгоритмические и нейросетевые методы

Алгоритмические (rule-based) методы стали первыми попытками автоматизировать выявление заимствований. Данные методы основываются на расчетах фонетического сходства между словами разных языков и на заданных порогах, при превышении которых слова считаются вероятно родственными из-за заимствования. Одним из простейших показателей является расстояние Левенштейна между фонетическими формами слов. Метод Левенштейна и подобные методы способны выявлять очевидные случаи заимствований, особенно при отсутствии родства между языками. Однако их надёжность ограничена, так как простое совпадение звуковой формы ещё не гарантирует заимствованного происхождения, а выбор порога сходства зачастую ненадежен.

Для повышения точности были разработаны алгоритмы, учитывающие системные фонетические соответствия и классы звуков. Важным шагом стало введение фонетического выравнивания (alignment) и сравнение не отдельных букв, а групп сходных звуков. Например, метод Sound Class Alignment (SCA) для выравнивания слов с преобразованием фонем в обобщённые звуковые классы (по месту и способу образования) [List, Forkel 2022], превосходящий классические методы редакционного расстояния по точности [Miller, List 2023: 2599–2605]. Другой подход — анализ фонетических соответствий на уровне групп языков, основанный на идее, что заимствования нарушают регулярные звуковые соответствия внутри языковой семьи. Если в списке родственных языков определённое слово выбивается из регулярных звуковых переходов, возможно, оно не наследовано, а заимствовано [Miller, List 2023: 2599–2605].

Современные нейросетевые методы применяются к задаче определения заимствований двумя основными способами: (1) для

классификации слов по их происхождению и (2) для оценки парного сходства слов разных языков. Нейросети позволяют учесть нелинейные комбинации признаков и работают напрямую с символьным (фонемным) представлением слова, минимизируя потребность в ручной обработке признаков.

Один из подходов — обучить нейросетевой классификатор, который по записи слова (например, в транскрипции) предсказывает, является ли оно заимствованием. Для этого слово подаётся как последовательность символов (букв или звуков) в однослойную или многослойную сеть. Сверточные сети способны выучивать характерные n -граммы звуков, которые могут служить маркерами заимствований (например, сочетания звуков, не свойственные родной фонологии языка). Рекуррентные сети могут учитывать структуру последовательности целиком [Silva, Ahmadi 2025]. Отдельно стоит класс подходов, где задача формулируется как оценка степени родства (этимологического соответствия) между двумя словами из разных языков. Здесь эффективны сиамские нейросети — архитектуры, принимающие на вход пару слов и возвращающие вероятность того, что слова связаны. Сиамская сеть обычно состоит из двух ветвей, каждая из которых вычисляет эмбединг для слова своего языка; затем сходство эмбедингов измеряется некоторой функцией, на основе которой делается вывод о родстве [Rama 2016: 1018–1027]. Сиамские сети применялись прежде всего для автоматического поиска родственных слов (когнатов) [Rama 2016: 1018–1027], но их можно использовать и для поиска заимствованных пар, особенно если обучать сеть на известных парах «слово в языке А — то же слово в языке В». Сиамская архитектура удобна тем, что её можно обучать на примерах заведомо связанных и несвязанных слов (по данным этимологических словарей), после чего применять к новым парам. При достаточном объёме таких данных сеть способна учитывать сложные соответствия звуков лучше, чем фиксированные метрики. Ещё одно направление — применение последовательных (seq2seq) моделей для генерации предполагаемой исходной формы по данной заимствованной

форме или наоборот. Идея состоит в том, чтобы смоделировать процесс фонетической адаптации при заимствовании как некое преобразование строки (слова). Например, можно попытаться обучить нейросеть, которая по французскому слову генерирует его адаптированный вариант на английском (или наоборот) — по сути, моделируя заимствование [Ciobanu, Dinu 2019:667-704]. В 2020-е гг. исследователи обратились к использованию больших языковых моделей (LLM) и многоязычных трансформеров (например, mBERT, XLM-R). Их можно либо применять напрямую (предоставляя модели слово и прося решить, заимствовано оно или нет), либо настроить (fine-tune) на задаче классификации заимствований, так как без обучения большие LLM не способны надёжно отличить заимствованные слова от исконных [Silva, Ahmadi 2025].

В целом нейросетевые архитектуры представляют наиболее продвинутую группу методов. Преимущество их в способности выявлять сложные паттерны — как фонетические, так и контекстуальные. Кроме того, они легко переносятся на новые языки при наличии данных: например, можно обучить одну модель на нескольких языковых парах сразу. Недостатки же нейросетевых методов — высокая требовательность к данным и вычислительным ресурсам, а также меньшая интерпретируемость. Часто трудно понять, по каким признакам сеть решила, что слово — заимствование, без дополнительного анализа. Поэтому в случаях, где это важно, нейросети используются в сочетании с более прозрачными методами: сеть может предложить кандидатов, а затем лингвист проверит соответствия.

Помимо факта заимствования, важнейшим аспектом этимологического анализа является определение, из какого языка произошло заимствование. Идентификация языка-донора часто рассматривается как задача множественной классификации: для данного слова в языке А выбрать наиболее вероятный язык-источник из некоторого набора кандидатов. Традиционные признаки для этого — специфические звуковые соответствия, характерные для данной пары языков. Автоматически такие соответствия

можно выявлять, обучив модель различать, какие фонетические n-граммы указывают на тот или иной язык-донор [Miller, List 2023: 2599–2605]. Одним из подходов к определению донора является многоязычный поиск похожего слова: взяв слово-кандидат на заимствование, можно сравнить его со словами из множества потенциальных языков-доноров и выбрать язык, чей словарь содержит наиболее похожее слово с тем же значением. Автоматически такой подход может быть реализован через вычисление расстояний (например, Левенштейна или с помощью эмбедингов) между словом и словами из разных языков.

Одной из целей этимологического анализа является разграничение заимствованных слов и слов-когнатов. Например, чтобы отличить заимствование от исконного слова, система может использовать информацию о распределении слова в языковой семье. Если предполагается, что слово исконное, то его когнаты должны обнаруживаться в родственных языках. Если же аналогичных слов нет ни в одном родстве, кроме подозреваемого донора — аргумент в пользу заимствования. Такой подход реализован в некоторых алгоритмах на базе филогенетических деревьев. Например, Automatic Tree Reconciliation — метод, где слова распределены по предполагаемым семьям, и если какое-то слово «выпадает» из общего древа и лучше объясняется вставкой ветви из другого дерева, то оно считается заимствованием [Köllner 2021].

Литература

- Ciobanu & Dinu 2019 — Ciobanu A.-M., Dinu L. P. Automatic Identification and Production of Related Words for Historical Linguistics // Computational Linguistics. 2019. Vol. 45(4). Pp. 667–704.
- Köllner 2021 — Köllner M. Automatic Loanword Identification using Tree Reconciliation. Dissertation (Dr. phil.). Tübingen: Eberhard Karls Universität Tübingen, 2021.
- List, Forkel 2022 — List J.-M., Forkel R. Automated identification of borrowings in multilingual wordlists // Open Research Europe. 2022. 1:79. DOI: 10.12688/openreseurope.13843.3.
- Miller, List 2023 — Miller J. E., List J.-M. Detecting Lexical Borrowings from Dominant Languages in Multilingual Wordlists. In: Proceedings

- of the 17th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2023). 2023. Pp. 2599–2605.
- Rama 2016 — Rama T. Siamese Convolutional Networks for Cognate Identification. In: Proceedings of COLING 2016, Osaka, 2016. Pp. 1018–1027.
- Sousa Silva, Ahmadi 2025 — Sousa Silva M., Ahmadi S. Language Models Are Borrowing-Blind: A Multilingual Evaluation of Loanword Identification across 10 Languages. arXiv preprint arXiv:2510.26254, 2025. URL: <https://arxiv.org/abs/2510.26254>.

Р. Т. Муратова,
И. Р. Вильданов
Институт истории, языка и литературы
УФИЦ РАН

Названия для мясного бульона и супа в тюркских языках: поиск этимологий и картографирование¹

Цель настоящего исследования —изучить лексемы, обозначающие мясной бульон и суп в тюркских языках, и выявить общее и специфическое в их развитии и функционировании. Поиск этимологий и картографирование проводились на лингвистической платформе ЛингвоДок.

Для начала внесем ясность в понятия «бульон» и «суп». В толковых словарях русского языка (например, в словаре Ожегова) даются следующие определения: бульон — это отвар мяса (а также может быть отваром рыбы, грибов, овощей); суп —это отвар из мяса, рыбы, заправленный мучными изделиями, овощами, крупами и т. д. Другими словами, бульон — это просто отвар, в нашем случае — мясной, а суп — это бульон, заправленный различными ингредиентами.

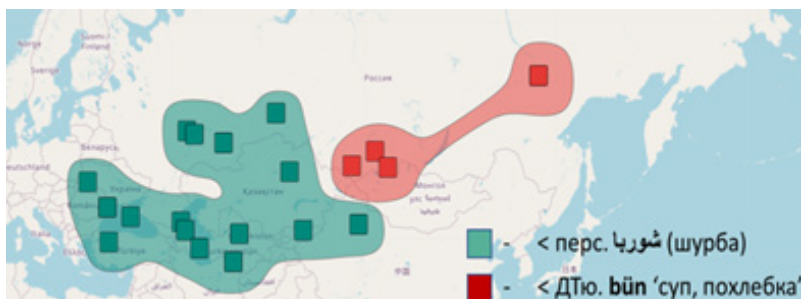
Бульон. Мясной отвар (или бульон) наравне с мясом считался основным блюдом у тюркских народов. Для обозначения бульона в тюркских языках употребляются 2 типа наименований:

1) названия, восходящие к перс. شوربا (шурба) ‘бульон; суп; похлебка’: башк. *һурпа*, тат. *шулпа*, кар., кум. *шорпа*, кбалк., кирг. *сорпо*, ног., ккалп., каз. *сорпа*, сиб.-тат. *шурба*, уйг. *шорва*, узб. *шўрва*, тур. *çorba*, туркм. *чорба*, чув. *шўрпе*;

2) лексемы, восходящие к ДТю. *bün* ‘похлебка’ (<сино-кор. *tjeep* ‘суп’): алт., тув., хак. *мун*, як. *эт мине* ‘букв. мясной суп’.

Обозначим эти лексемы на карте:

¹ Исследование выполнено за счет гранта Российского научного фонда № 24-28-01562, <https://rscf.ru/project/24/-28-01562/>. Результаты получены с использованием услуг Центра коллективного пользования Институтом системного программирования им. В. П. Иванникова РАН — ЦКП ИСП РАН.



На карте можно увидеть, что область распространения персидского заимствования занимает западную и центральную части территории расселения тюрков. У восточных (сибирских) тюрков, в которые не проник данный персизм, для обозначения мясного бульона функционируют лексемы, восходящие к древнетюркскому слову. Эта же лексема в форме *bün* зафиксировано в словаре М. Кашгари в значении 'суп, похлебка' с пометками *огуз., кыпчак., сувар* (то есть оно употреблялось у предков современных тюркских языков огузской, кыпчакской, болгарской подгрупп).

Примечательно, что именно в перечисленных Махмудом Кашгари языках это слово сегодня не употребляется для обозначения бульона. Можно предположить, что оно, возможно, функционировало в этих языках, но в результате языковых контактов было вытеснено персидским заимствованием.

Суп. Наряду с бульоном, традиционным блюдом у тюркских народов был суп, приготовленный на мясном бульоне. Для обозначения мясного супа в тюркских языках употребляются три типа лексем:

1) названия, восходящие к ПТю. **(i)al'* 'еда, пища': башк., тат., сиб.-тат., кр.-тат. уйг. *аш*, тур. *aş*;

2) названия, совпадающие с лексемой для бульона:

а) персидское заимствование شوربا (*шурба*) 'бульон; суп; похлебка': ног. *сорпа*, каз. *сорпа*, ккалп. *шорпа*, *сорпа*, кум. *шорпа*, кар. *шорпа*, тур. *çorba*, туркм. *чорба*, узб. *шўрва*, чув. *шўрпе*;

б) восходящие к ДТю. *bün* 'похлебка: алт., тув., хак. *мүн*, шор. *мўн*, як. *эт мине* 'букв. мясной суп'.

К изучению языка письменных источников на ойратском языке (на материале параллельного тибетско-ойратского корпуса текстов)¹

Письменные источники, включенные в параллельный тибетско-ойратский корпус, размещенный на платформе LingvoDoc, представляют большой интерес в плане сравнительно-сопоставительного текстологического анализа письменных памятников на классическом тибетском и ойратском языках. Важное место в корпусе отведено сочинениям из числа переводов известного ойратского просветителя XVII в. Зая-пандиты Намкай Джамцо (1599–1662), которые записаны на созданную им письменность «тодо бичиг» («ясное письмо») (1648). Эта письменность сохранялась у калмыков до 1924 г., сыграла особую роль в их судьбе. На ней создавались не только переводы буддийских памятников с тибетского языка, но также и оригинальные религиозные и светские сочинения. Среди этих произведений можно отметить разноплановые, разножанровые произведения, большей частью, из состава буддийского канона, но также и внеканонические сочинения.

Переводы Зая-пандиты и его учеников дают возможность изучить язык памятников XVII в. и более ранних текстов, включенных в состав канона, а также принципы и правила, которых придерживался ойратский просветитель в работе по переводу с тибетского языка на монгольский и ойратский. Эти вопросы в определенной степени освещались в литературе по монголоведению и ойратоведению. Среди этих работ можно отметить труды отечественных и зарубежных монголоведов Г. Ц. Цыбикова, Н. С. Яхонтовой, А. Д. Цендиной, Д. Н. Музраевой, Н. В. Ям-

¹ Исследование проведено при финансовой поддержке РФФ в рамках научного проекта № 23-18-00978 «Параллельный корпус текстов на тибетском языке и ойратском “тодо бичиг” (“ясном письме”): разработка инструментария и текстологические исследования».

польской, С. В. Мирзаевой и др., в которых рассматривались принципы перевода с тибетского языка на монгольские языки с опорой на примеры из таких сочинений, как «Лам-рим чэн-по», «Сутра Золотого блеска», «Книга сына», «Сутра о мудрости и глупости», «Бхадрачарья» и др. Тем не менее, на сегодняшний день еще недостаточно работ, которые бы раскрывали принципы перевода Зая-пандиты и его школы на основе большого массива текстов. Решению этого вопроса и будет способствовать параллельный тибетско-ойратский корпус. В его создании были использованы современные информационные технологии платформы LingvoDoc, разработанные сотрудниками ИСП РАН, при этом программный код создавался в том числе с помощью нейронных сетей.

Описываемый параллельный корпус является инновационным проектом в отечественной тибетологии и монголистике, поэтому неудивительно, что до недавних пор отсутствовали алгоритмы по созданию параллельных корпусов на восточных языках, в особенности, при использовании текстов на языках, существенно отличающихся по своей лингвистической структуре. Как известно, ойратский является одним из монгольских, т. е. агглютинативных, языков, а тибетский относится к сино-тибетской языковой группе, особенностью которого является слоговое письмо. В решении вопроса сопоставления текстов на этих двух языках важная роль отведена научной транслитерации на латинице, которая собственно и вводится в специально разработанную программу распараллеливания двух текстов — на тибетском и ойратском. Отличительной особенностью последних является то, что до наших дней дошли многочисленные копии и списки известных сочинений, иногда фрагментарные, с большими утратами, составление которых было сопряжено с отходом от классического «тодо бичиг». Подобные тексты весьма затрудняли набор транслитерации. В этой связи актуальным стало создание программы по распознаванию ойратского вертикального письма. В рамках проекта была создана программа по распознаванию «тодо бичиг», что способствовало составлению научной транс-

литерации текстов на ойратском языке. Среди них можно отметить такие хорошо известные памятники, как «Алмазная сутра» (или «Дорджи Джодва»), «О пользе чтения „Дорджи Джодва“», «Пророчество Будды», «Амитаюс» (или «Ценд-Аюуш»), «Манджушри-нама-сангити» (или «Истинное произношение имен Манджушри»), «Субхашита» (или «Сокровищница назидательных речений»), «Тарпа-ченпо» («Великий освободитель»), 1-я часть «Мани-камбума» (или «Собрания изречений о мани»), «Море притч» и др.

В процессе создания параллельного корпуса разработанная программа проходила тестирование, в алгоритм выравнивания текстов вносились коррективы.

Таким образом, создание тибетско-ойратского корпуса имеет важное значение в решении многих вопросов современного монголоведения и тибетологии, является одним из перспективных направлений в корпусной лингвистике, а также в источниковедении, текстологии, литературоведении, демонстрирует новый подход в изучении языка ойратских переводных памятников в сопоставлении с их первоисточниками на тибетском языке.

Хадмал-тексты как прототип современного тибетско-ойратского параллельного корпуса¹

В процессе создания тибетско-ойратского параллельного корпуса значимым этапом является решение задачи автоматического выравнивания текстов, заключающегося в разбиении обоих языковых вариантов на небольшие сегменты, предпочтительно предложения. Выравнивание подразумевает сопоставление фрагментов исходного текста с соответствующими переводными сегментами, которые должны располагаться рядом друг с другом. Этот процесс включает идентификацию и согласование текстовых единиц, что позволяет формировать структурированные и синхронизированные данные, необходимые для дальнейшего анализа и обработки в рамках корпусной лингвистики.

В обсуждении вопросов выравнивания в тибетско-ойратском параллельном корпусе нельзя не учитывать историческую традицию монгольского книгопечатания, в частности феномен так называемых тибетско-монгольских билингвальных текстов, известных в монгольской терминологии как «хадмал». Их можно рассматривать как исторические аналоги современных параллельных корпусов, которые используются для изучения межъязыковых соответствий и практик перевода. Они представляют важный пласт монгольской буддийской литературной традиции и отражают специфические особенности взаимодействия тибетской и монгольской письменных культур. С точки зрения корпусных исследований, хадмал-тексты обладают уникальными структурными характеристиками, а их билингвальное оформление служит эмпирической базой для изучения переводных практик и установления языковых соответствий между тибет-

¹ Исследование проведено при финансовой поддержке РФФИ в рамках научного проекта № 23-18-00978 «Параллельный корпус текстов на тибетском языке и ойратском “тодо бичиг” (“ясном письме”): разработка инструментария и текстологические исследования».

ским и монгольским языками. Изучение этих текстов позволяет выявить типичные лексические и грамматические маркеры, а также устойчивые словоформы, актуальные для задач автоматизированного выравнивания и формирования параллельных корпусных данных тибетско-ойратских текстов. Эти материалы предоставляют богатую основу для формализации лингвистических инструментов, используемых в корпусной лингвистике, и способствуют глубокому изучению механизмов адаптации тибетского языка в ойратской книжной традиции.

В рассматриваемых текстах применяется практика дословного или подстрочного перевода, что обусловлено сходством структуры предложений в тибетском и монгольском языках, позволяющим выполнять перевод с минимальными искажениями. Например, в первых строках поклонения Зеленой Таре: «‘phagspaspyanrasgzigsdbangphyugthugsrje’igterlaphyg ‘tshallo — qutuγ tu nidü-ber üjegčī erketü nigülesküi-yin sang-dur mörgümüi», перевод выполнен слово в слово, и при этом на монгольском языке звучит корректно. Например, в дословном переводе на русский язык, где фраза может восприниматься тяжеловесно: «*Благородному Авалокитешваре Владыке, сострадания кладезю поклоняюсь*», такой перевод, несмотря на свою буквальность, в контексте церковной или религиозной лексики, воспринимается уместно и будет соответствовать торжественному стилю служебных текстов.

Поскольку хадмал-тексты представляют буквалистический стиль перевода, они обеспечивают четкое и однозначное отображение параллельных маркеров и индикаторов в обоих языках.

Начальный этап предполагает использование базовых пунктуационных маркеров — в тибетских текстах это шеды (одинарные, двойные, четыре вертикальные черты), а в ойратских соответствующие — точка, две точки и четыре точки. Несмотря на то, что данный подход позволяет идентифицировать очевидные сегменты текста, он приводит к выделению избыточно крупных блоков, что снижает точность выравнивания и требует дальнейшей оптимизации методологии посредством введения

более тонких и специфичных лингвистических индикаторов. В решении этой задачи перспективным является исследование хадмал-текстов на предмет выявления соответствующих друг другу лексических и прочих маркеров. Это будет способствовать совершенствованию автоматизированных методов выравнивания текстов и развитию корпусного анализа за счет создания более точной и детальной базы данных лингвистических индикаторов.

**«Море притч» (üligeriyin dalai)
как источник изучения ойратской лексики XVII в.
(по материалам параллельного тибетско-ойратского
корпуса)¹**

«Море притч» (*üligeriyin dalai*) является классическим образцом тибето-монгольской повествовательной литературы XVII в., включенным в состав Ганджура. Тибетскую версию сутры исследователи датируют 632 г. н. э. Значительный вклад в изучение этого текста внесли российские востоковеды, среди которых особо следует отметить Я. И. Шмидта, А. А. Бобровникова, А. В. Попова, К. Ф. Голстунского и А. М. Позднеева.

На сегодняшний день науке известно пять монгольских переводов этого сочинения, каждый из которых имеет свое название. Труд одного из самых известных монгольских переводчиков XVI–XVII вв. Ширээт-гуши-Цорджи (Гуши Цорджи, монг. *Širegetü güüsi čorji*) из Хух-Хото имел название «*Siluyun onol-tu sudur*». Другой монгольский перевод этого сочинения, озаглавленный как «*Silayun budayun üy-e onoqui*», был выполнен Тойн гуши. Автор третьего из известных переводов — Зуун-Авгинский Цултэмлодой. Его работа имеет название «*Sayin mayu üyle-yi uqayulaqui*» и была обнаружена профессором Д. Цагаан в Хубсугульском аймаке. По мнению академика Ц. Дамдинсүрэнэ, этот перевод был осуществлен в XVI в. В XVII в. это сочинение перевел на ойратский язык известный просветитель Зая-пандита Намкай Джамцо (1599–1662). Хронологически наиболее поздним переводом сутры «Море притч» (*üligeriyin dalai*) считается работа калмыцкого буддийского священнослужителя

¹ Исследование проведено при финансовой поддержке РФФ в рамках научного проекта № 23-18-00978 «Параллельный корпус текстов на тибетском языке и ойратском “тодо бичиг” (“ясном письме”): разработка инструментария и текстологические исследования».

Тугмюд-гавджи (О. М. Дорджиева, 1887–1980), записанная на ойратском «ясном письме» (тодо бичиг) в 1968 г.

Особо следует остановиться на переводе Зая-пандиты, хорошо известном под названием *Medētei mede ügeiyigi ilya qui ketēkii sudur* («Сутра, различающая мудрого и глупого»). По мнению монгольского ученого Х. Лувсанбалдана, перевод данной рукописи был выполнен в период между 1650 и 1661 гг. [Лувсанбалдан 1968: 118–119]. Текст ойратской рукописи «Üligeriyin dalai» («Море притч») был опубликован Б. Ринченом. В предисловии к изданию он отмечал, что «„Книга Улгэр-ун далай” — „Океан сказаний” была одной из самых распространенных среди читателей старой Монголии книг. Не было стойбища, в котором не было бы этой книги, существовавшей в Монголии в нескольких переводах» [Biography ... 1967: VII]. Списки перевода Зая-пандиты в настоящее время имеются в известных мировых собраниях восточных книг [Сазыкин 2001: 116; Uspensky 2001: 30; Gerelmaa 2005: 232; Бичеев 2011: 241]. В 1970 г. была осуществлена публикация рукописи перевода Зая-пандиты под названием «**Oirat Version of Damamūkonāmasūtra. Üligeriyin dalai**» в издании «Corpus Scriptorum Mongolorum».

При предварительном анализе ойратского перевода «Моря притч» («Üligeriyin dalai») были выявлены следующие лексические пласты:

1. Тюркские заимствования, представленные такими лексемами, как *keseq* ‘часть’, *bölöq* ‘глава’, *biliq* ‘знание’.

2. Санскритская лексика, проникшая в монгольские языки в связи с распространением буддизма. Изначально заимствование религиозных терминов осуществлялось через уйгурский язык, а впоследствии — через тибетский, поскольку именно с тибетского переводилось большинство буддийских трактатов. В качестве примеров можно привести слова *sudur* ‘сутра’, *zandan* ‘сандал’, *aṣṣan* ‘святая вода’ и *bodhi* ‘просветление’.

3. Тибетские заимствования, которые включают не только религиозную, но и бытовую лексику: *tonting* ‘сапфир’, *klu* ‘нага, дракон’, *bumba* ‘бутылка’.

4. Заимствования из греческого и арабского языков: *nom* ‘книга’ (из греческого через сирийские языки); *araki* ‘вино’ (из арабского).

5. Слова китайского происхождения, относящиеся в основном к области административной терминологии: *noyon* ‘князь’, ‘господин’, *šang* ‘казна’.

Таким образом, ойратский перевод «Моря притч» («*üligeriün dalai*»), выполненный Зая-пандитой в XVII в., представляет собой уникальный и комплексный источник для изучения лексического состава ойратского языка.

Применение методов машинного обучения для векторизации тибетского и ойратского текстов¹

Актуальность. Проблема цифрового неравенства языков представляет собой значительное методологическое ограничение в области обработки естественного языка (NLP). Множество языков, обладающих глубоким историко-культурным наследием, относятся к категории низкоресурсных. К их числу принадлежат, в частности, тибетский язык и ойратский язык, использующий классическую монгольскую письменность («Тодо бичиг»). Критический дефицит машинно-читаемых текстовых корпусов, особенно параллельных и размеченных, существенно затрудняет разработку и применение для них современных лингвистических технологий. В этом контексте ключевой исследовательской задачей становится генерация качественных семантических эмбедингов — векторных представлений текста, кодирующих смысловые отношения и пригодных для решения широкого спектра задач, таких как информационный поиск, классификация и кластеризация текстов, а также кросс-языковой перенос (cross-lingualtransfer).

Методология и данные. В рамках данного исследования работа проводилась с использованием параллельного тибетско-ойратского корпуса, размещенного на платформе Litgovodoc. Обучающий датасет состоит из 2 516 выровненных пар предложений. Для целей тестирования и оценки из корпуса был экстрагирован специальный лексический подкорпус, содержащий 114 пар ключевых буддийских терминов на обоих языках. На базе этого набора данных была выполнена адаптация (fine-tuning) многоязычной модели LaBSE (Language-agnostic BERT

¹ Исследование проведено при финансовой поддержке РНФ в рамках научного проекта № 23-18-00978 «Параллельный корпус текстов на тибетском языке и ойратском „тодо бичиг“ („ясном письме“): разработка инструментария и текстологические исследования».

SentenceEmbedding) с целью получения контекстуализированных векторных представлений для исследуемых языков.

Результаты и их визуализация. Для анализа и интерпретации полученных эмбеддингов была применена методика снижения размерности с помощью алгоритма UMAP (Uniform Manifold Approximation and Projection) с последующей кластеризацией. Визуализация результатов выявила различную структурную организацию семантических пространств двух языков. Так, для ойратских терминов наблюдается выраженная кластерная структура, визуально идентифицируется пять компактных групп. В то же время, векторные представления тибетских терминов не формируют столь же четко дифференцированных кластеров, однако визуальный анализ подтверждает, что семантически близкие единицы проецируются в близкие точки на плоскости, сохраняя локальные смысловые связи.

Вывод. Проведенный эксперимент демонстрирует, что даже ограниченный по объему, но качественно подготовленный параллельный корпус (в данном случае ~2.5 тыс. предложений) является достаточным для успешной векторизации текстов низко-ресурсных языков на основе адаптации предобученных языковых моделей. Полученные семантические пространства обладают интерпретируемой внутренней структурой, что подтверждает практическую применимость предлагаемого подхода для начальных этапов интеграции таких языков в цифровую среду.

Литература

- Feng et al. 2022 — *Feng F. et al.* Language-agnostic BERT sentence embedding // Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2022. Pp. 878–891.
- McInnes et al. 2018 — *McInnes L., Healy J., Melville J.* Umap: Uniform manifold approximation and projection for dimension reduction // URL: arXiv preprint arXiv:1802.03426. 2018.

*С. В. Ковылин,
Институт системного программирования
им. В. П. Иванникова РАН*

Опыт цифровой обработки селькупских и хантыйских языковых материалов

В ходе глобализации с нарастающей интенсивностью происходит исчезновение малых языков или сокращение численности их носителей. В качестве примеров можно привести центральные и южные селькупские диалекты, практически прекратившие свое существование, если не считать ряда носителей, чей язык был ревитализован на базе академических ресурсов. Или, например, отдельные диалекты хантыйского языка, где васюганский также практически прекратил свое функционирование, в то время как для ваховского выделяется еще около 100 активных носителей. В связи с этим большое значение приобретает цифровизация данных, которая способствует сохранению, исследованию и ревитализации языков и культур в целом, путем аккумулирования имеющейся информации и предоставления быстрого доступа к ней.

Существуют различные инструменты для цифровой обработки текстовых, аудио- и видеоматериалов, среди которых можно выделить такие программы как FieldworksLanguageExplorer (FLEx), Toolbox, Praat, ELAN, Audacity. Для анализа текстов также разрабатываются различные парсеры. В 2012 г. появилась платформа ИСП РАН Lingvodoc¹, предназначенная для хранения и многоуровневой обработки языковых данных. Эти и другие инструменты использовались автором для обработки памятников письменности и создания корпусов текстов.

Далее приведем пример процедуры создания языковых баз данных по селькупским и хантыйским диалектам при помощи разных программ.

¹ Результаты получены с использованием услуг Центра коллективного пользования Института системного программирования им. В. П. Иванникова РАН — ЦКП ИСП РАН.

Изначально для внесения текста и проработки материалов используется программа FLEx (применяется полуавтоматический парсинг данных), которая достаточно хорошо подходит для аннотации бесписьменных языков. Часто из-за отсутствия описания морфологии отдельных диалектов приходится создавать отдельные корпуса по каждому информанту.

Далее, файлы из FLEx конвертировались в программу ELAN, однако это требует дополнительных процедур по адаптации материала, привязке его к иерархическим слоям. Если текст сопровождается аудио, то требуются дополнительные действия для привязки его к временной шкале.

Полученные в программе ELAN материалы конвертируются на платформу Lingvodoc (<http://lingvodoc.ispras.ru>), которая позволяет создавать лексические и морфологические словари согласно разметке (в том числе с автоматической нарезкой аудиофайлов). Все корпусные данные сохраняются в неизменном виде с возможностью их редактирования и объединения. Обработка материалов возможна при помощи инструментов доступных на платформе, а сама работа отличается автономностью (без постоянной привязки к программистам), быстрым решением возникающих вопросов (есть прямой чат для общения программистов ИСП РАН и пользователей), а также открытостью данных. Совместимость разных программ и понятный процесс конвертации позволяют быстро создавать разные типы баз данных, использовать все их инструменты, оперативно выкладывать в открытый доступ полученные материалы. Главной задачей в таком случае является постоянное накопление и улучшение качества информации.

В результате плодотворной работы были созданы и выложены в открытый доступ десятки текстовых и аудиокорпусов по различным селькупским и хантыйским диалектам, созданы лексические и морфологические словари (см. соответствующие разделы на [Lingvodoc 3.0: <http://lingvodoc.ispras.ru>]). Созданные корпуса позволяют заново пересмотреть более ранние грамматические описания, сделанные на базе ограниченного количества материалов, но уже опираясь на анализ

больших данных. Так, в качестве одного из примеров корпусного анализа можно представить исследование памятников письменности XIX в. в сопоставлении с данными за XX в. (см. [Ковылин, Норманская 2022]).

Перспективы корпусных исследований с применением инструментов компьютерной лингвистики довольно широки. Необходимо и далее накапливать данные, выкладывать их в открытый доступ и продвигать политику открытости обработанных по грантам материалов для использования их в научных, учебных и просветительских целях.

Литература

Ковылин, Норманская 2022 — *Ковылин С. В., Норманская Ю. В.* Кириллические памятники на уральских и алтайских языках. II том. Памятники селькупской письменности XIX в., созданные святителем Макарием (Невским). М.: Издательская группа «Альма Матер», 2022. 301 с.

Lingvodoc 3.0. Institute of Linguistics Russian Academy of Sciences, Ivannikov Institute for System Programming of the Russian Academy of Sciences. URL: <https://lingvodoc.ispras.ru> (accessed: 08.11.2025).

Записи дагурского языка начала и конца XX в.: фонетические процессы¹

Дагурский язык относится к монгольской группе алтайских языков и является родным языком этнического меньшинства дауров (дахуры, дагуры), проживающих преимущественно в Китае, в основном во Внутренней Монголии и в Синьцзян-Уйгурском автономном районе. Согласно переписи 2020 г. дауры также проживают в России: в Амурской области (200 человек) и Хабаровском крае (89 человек) [см. <https://ru.wikipedia.org/wiki/Дауры>].

Дауры ведут свою историю с древнейших времен, однако исследования дагурского языка в России начались в конце XIX в. с небольшой публикации А. О. Ивановского.

Отдельного внимания заслуживают исследования Н. Н. Поппе относительно дагурского языка. Н. Н. Поппе провёл глубокие исследования лексики, грамматической структуры и фонологии языка, выявив специфику его развития и связи с соседними языками Центральной Азии. В работе «Дагурское наречие» Н. Н. Поппе (1930) подчёркивал, что дагурское наречие обладает рядом архаичных особенностей, сближающих его с древнейшими формами монгольского языка.

Б. Х. Тодаева внесла существенный вклад в исследование фонетики, морфологии и синтаксиса этого языка, а также исследовала особенности взаимодействия дагурского языка с китайским языком и культурой.

Настоящее исследование выполнено на основе лексических материалов дагурского языка, представленного в словаре

¹ Исследование проведено при финансовой поддержке РФФ в рамках проекта «Разработка инструментария и комплексные исследования монгольских языков и их диалектов (с применением технологий анализа больших массивов данных словарных и корпусных материалов)» (№ 25-78-20008).

Б. Х. Тодаевой (бутхаский говор) (1986) и словаре Н. Н. Поппе (хайларский говор) (1930). Дополнительно использовались реконструкция Х. Нугтерена (2011), опубликованная в его монографии о южномонгольских языках, а также краткий дагурско-русский словарь современного дагурского языка, подготовленный Г. Тумурдзем и Б. Д. Цыбеновым (Бурятский научный центр СО РАН, 2014). Инициатива составления краткого дагурско-русского словаря принадлежит Г. Тумурдэю, который начиная с 2007 года активно собирал лексику дагурского языка и осуществил самостоятельный перевод ряда выражений на русский язык. Так как сам Г. Тумурдэй владеет хайларским диалектом, именно этот вариант дагурской речи лёг в основу представленного словаря.

Исследование проведено с помощью платформы «Lingvodoc», где были размещены словники. Представленные данные содержат анализ соответствий звуков в дагурском языке, основанных на сравнении четырех разных источников лексики.

Вот некоторые выводы анализа.

В следующих примерах видим переход, при котором палатализованный назальный согласный *ɲ* превращается в обычный зубной назальный согласный *n* в конце слова:

– ун 'ч'й̆ «сирота» [Поппе 1930: 95]: ПМонг. **onecin* 'сирота' [Nugteren 2011: 474]; унчин 'сирота' [Тодаева 1986: 171]; совр. лит. унчин 'сирота' [КДРС 2014: 170];

– к'ӗ̆̆ «кто» [Поппе 1930: 83]: ПМонг. **ken* 'кто' [Nugteren 2011: 410]; кэн 'кто' [Тодаева 1986: 149]; совр. лит. хэн 'кто' [КДРС 2014: 198];

– хал̆ў «жаркий, горячий» [Поппе 1930: 97]: ПМонг. **kalaun* 'горячий' [Nugteren 2011: 402]; хал̆он 'жаркий, жарко' [Тодаева 1986: 173]; совр. лит. халуун 'горячий, жаркий' [КДРС 2014: 177].

Таким образом, здесь представлен процесс упрощения артикуляции, т. е. депалатализации. Это довольно распространённый процесс в истории многих языков.

Дифтонги, подобные *ай* в конце слова, подвержены изменениям и даже полной утрате со временем. Подобные тенденции

наблюдаются практически во всех языках мира. Часто второй элемент дифтонга (например, *й*) теряется, оставляя первый компонент неизменённым или слегка видоизменённым. Например:

– *мобој* «змея» [Поппе 1930: 86]; ПМонг. **mogai* ‘змея’ [Nugteren 2011: 445]; *мого*, *мог* ‘змея’ [Тодаева 1986: 154]; совр. лит. *могу* ‘змея’ [КДРС 2014: 112];

– *ноб* «собака» [Поппе 1930: 87]; ПМонг. **no kai* ‘собака’ [Nugteren 2011: 462]; *ного*, *ног* ‘собака’ [Тодаева 1986: 158]; совр. лит. *ногу* ‘собака’ [КДРС 2014: 125];

– *шарай* «лицо» [Поппе 1930: 101]; ПМонг. **čirai* ‘лицо’ [Nugteren 2011: 303]; *шари*, *шарā* ‘лицо’ [Тодаева 1986: 183]; совр. лит. *шар* ‘лицо’ [КДРС 2014: 214].

В следующих примерах дифтонг *ай* встречается в середине слова и благодаря своей позиции сохранился с древней эпохи до наших дней:

– *хайлас* «вяз» [Поппе 1930: 96]; ПМонг. **kailasun* ‘улица вязов’ [Nugteren 2011: 400]; совр. лит. *кайлс* ‘вяз’ [КДРС 2014: 85];

– *сайхяң* «прекрасный» [Поппе 1930: 90]; ПМонг. **saikan* ‘красивая, симпатичный’ [Nugteren 2011: 480]; *саикан* ‘(уменьш. от *саин*) прекрасный, красивый’ [Тодаева 1986: 161]; совр. лит. *сайкэн* ‘красивый’ [КДРС 2014: 139];

– *хай’ч’* «ножницы» [Поппе 1930: 96]; ПМонг. **kaiči* ‘ножницы’ [Nugteren 2011: 400]; *хайчи*, *кайчи* ‘ножницы’ [Тодаева 1986: 148]; совр. лит. *кайч* ‘ножницы’ [КДРС 2014: 85].

В приведенных ниже примерах наблюдается характерная замена согласной *б* на *в*, произошедшая в ходе исторических изменений в монгольских языках:

– *шавяр* «глина» [Поппе 1930: 101]; ПМонг. **sibar* ‘грязь, суглинок’ [Nugteren 2011: 488]; *шавар* ‘глина’ [Тодаева 1986: 182]; совр. лит. *шаур* ‘глина’ [КДРС 2014: 215];

– *шово* «ловчая птица» [Поппе 1930: 101]; ПМонг. **sibaun* ‘птица’ [Nugteren 2011: 488].

Несмотря на общую тенденцию к замене *б* на *в*, обе формы продолжают сосуществовать в современных монгольских языках. В указанных ниже примерах видим сохранные формы древ-

него монгольского языка, что подтверждает тезисы Н. Н. Поппе относительно архаизма дагурского языка:

– *ш'ирбүс*, *ш'ирбёс* «сухожилие» [Поппе 1930: 101]: ПМонг. **sirbüsün* (?*sirbösin*) 'сухожилие' [Nugteren 2011: 497]; *ширбэс*, *ширбус* 'сухожилие' [Тодаева 1986: 184]; совр. лит. *ширбэс* 'жила, жилы; сухожилие' [КДРС 2014: 218];

– *дабхур* «двойной» [Поппе 1930: 75]: ПМонг. **dabkur* 'двойной; рядный' [Nugteren 2011: 310]; *дабкур* 'двойной' [Тодаева 1986: 134]; совр. лит. *дабкур* 'слой; ряд; ярус' [КДРС 2014: 52];

– *дёлбур* «веер» [Поппе 1930: 75]: ПМонг. **delbi* 'обмахиваться веером' [Nugteren 2011: 310]; *дэлбур* 'веер, опахало' [Тодаева 1986: 134]; совр. лит. *дэлбур* 'веер' [КДРС 2014: 52].

Начальная позиция звука в слове является сильной позицией, однако в записи Н. Н. Поппе зафиксировано исчезновение начальной согласной *h*, что вызывает вопросы и требует дополнительного рассмотрения. В словарях Б. Х. Тодаевой и Г. Тумурдэя, Б. Д. Цыбенова сохранились формы с начальной согласной *h*, отражающие исходную прамонгольскую форму соответствующих лексем:

– *оддод* «звезда» [Поппе 1930: 88]: ПМонг. **hodun* 'звезда' [Nugteren 2011: 358]; *ходо* 'звезда' [Тодаева 1986: 176]; совр. лит. *ход* 'звезда' [КДРС 2014: 183];

– *ук'ур* «бык» [Поппе 1930: 95]: ПМонг. **hüker* 'крупный рогатый скот' [Nugteren 2011: 367]; *хукур* 'крупный рогатый скот' [Тодаева 1986: 179]; совр. лит. *хукур* 'крупный рогатый скот' [КДРС 2014: 191];

– *юдүс* «перья» [Поппе 1930: 94]: ПМонг. **hödün* 'перо' [Nugteren 2011: 360]; *худус* 'перья, перо' [Тодаева 1986: 179]; совр. лит. *худус* 'перо' [КДРС 2014: 191].

В нижеследующих примерах наблюдаем прямой переход, когда начальная согласная *k* изменилась на менее напряжённую *x*, что соответствует общим тенденциям изменения звука в языках:

– *к'ёң* «кто» [Поппе 1930: 83]: ПМонг. **ken* 'кто' [Nugteren

2011: 410]; *кэн, хэн* ‘кто’ [Тодаева 1986: 149]; совр. лит. *хэн* ‘кто’ [КДРС 2014: 198];

– *к’үцү* ‘шея» [Поппе 1930: 84]: *ПМонг. *kijüün* ‘шея’ [Nugteren 2011: 435]; *куджу, худжү* ‘шея’ [Тодаева 1986: 151, 179]; совр. лит. *хужуу* ‘шея’ [КДРС 2014: 191].

Но также имеются не характерные для монгольских языков примеры возвращения сильного согласного *к* взамен более слабого *х*. Данный переход существенно уступает по масштабу классическому переходу *к* в *х*, поэтому этот переход возможно, носит локальный характер:

– *хаб’л* «черепаха» [Поппе 1930: 96]; *хабил* ‘черепаха’ [Тодаева 1986: 172]; совр. лит. *кабиль* ‘черепаха’ [КДРС 2014: 84];

– *хэйлэс* ‘вяз» [Поппе 1930: 96]: *ПМонг. *kailasun* ‘улица вязов’ [Nugteren 2011: 400]; совр. лит. *кайлс* ‘вяз’ [КДРС 2014: 85];

– *хэй’ч’* «ножницы» [Поппе 1930: 96]: *ПМонг. *kaiči* ‘ножницы’ [Nugteren 2011: 400]; *хайчи, кайчи* ‘ножницы’ [Тодаева 1986: 148]; совр. лит. *кайч* ‘ножницы’ [КДРС 2014: 85].

Дагурский язык представляет собой уникальный лингвистический материал и данное исследование показывает необходимость его дальнейшего изучения для полного понимания структуры и динамики этого языка.

Литература

Nugteren 2011 — *Nugteren H. Mongolic Phonology and the Qinghai-Gansu Languages. Utrecht: LOT, 2011. 563 p. (In Eng.)*

КДРС 2014 — Краткий дагурско-русский словарь. Сост. Г. Тумурдэй, Б. Д. Цыбенков / Улан-Удэ: Изд-во БНЦ СО РАН, 2014, 236 с.

Поппе 1930 — *Ponne H. H. Дагурское наречие. Ленинград, 1930. 178 с.*

Тодаева 1986 — *Тодаева Б. Х. Дагурский язык. Москва, 1986. 189 с.*

СОДЕРЖАНИЕ

Программа	3
<i>Норманская Ю. В.</i> Степень близости между дунсянским, баоаньским, шира- югурским языками с точки зрения глоттохронологии и системы регулярных фонетических соответствий	6
<i>Куканова В. В., Баянова А. Т.</i> Сарт-калмыцкий и калмыцкий язык: фонетические соответ- ствия (на материале полевых записей)	8
<i>Нуриева Ф. Ш., Галиуллина Г. Р., Юсупов А. Ф.</i> Графо-фонетическое описание материалов татарского языка XVIII в. на платформе LingvoDoc: по записям участников Великой Северной экспедиции	15
<i>Галиуллина Г. Р., Юсупов А. Ф., Нуриева Ф. Ш.</i> Сравнительный анализ лексики словника Г.Ф. Миллера и современного татарского языка	18
<i>Гончарова О. В.</i> Автоматизация этимологического анализа: алгоритмиче- ские и нейросетевые методы	22
<i>Муратова Р. Т., Вильданов И. Р.</i> Названия для мясного бульона и супа в тюркских языках: поиск этимологий и картографирование	27
<i>Музраева Д. Н.</i> К изучению языка письменных источников на ойратском языке (на материале параллельного тибетско-ойратского корпуса текстов)	30
<i>Тушинов Б. Л.</i> Хадмал-тексты как прототип современного тибетско-ойрат- ского параллельного корпуса	33
<i>Долеева А. О.</i> Море притч» (üligeriyin dalai) как источник изучения ойрат- ской лексики XVII в. (по материалам параллельного тибет- ско-ойратского корпуса)	36

<i>Танаев В. В.</i>	
Применение методов машинного обучения для векторизации тибетского и ойратского текстов	39
<i>Ковылин С. В.</i>	
Опыт цифровой обработки селькупских и хантыйских языковых материалов	41
<i>Манджиева И. Б.</i>	
Записи дагурского языка начала и конца XX в.: фонетические соответствия	44

Электронное издание

Платформа LingvoDoc: цифровизация языков народов России и
инструменты для исследования:
Тезисы докладов Круглого стола
(г. Элиста, 13 декабря 2025 г.)

Верстка Д. В. Татнинов
Дизайн обложки Д. В. Татнинов

Подписано в печать
Формат 60x84/1/16. Печать офсетная.
Усл. печ. л. 3,37. Заказ 32-25.

Отпечатано в КалмНИЦ РАН:
Республика Калмыкия, 358000, г. Элиста,
ул. им. И. К. Илишкина, д. 8