

Инновации, технологии

УДК 811.511.142`1:004(045)

И. М. Молданова

ХАНТЫЙСКИЙ ЯЗЫК: ОТ БУМАЖНЫХ РУКОПИСЕЙ ДО ЦИФРОВЫХ ПЛАТФОРМ



В статье рассматривается современная виртуальная среда, в которой, посредством лингвистических платформ сохраняется и изучается хантыйский язык. Научному исследованию хантыйского языка, которое началось в XIX в., предшествовали фрагментарные записи хантыйской лексики, известные с XVII века. На современном этапе изучение и резервирование языковых данных осуществляется с помощью новейших цифровых технологий, при этом материалы исследования находятся в открытом доступе, что повышает ценность языковых баз. Цель: рассмотреть современные онлайн-площадки, где аккумулируется материал по исследованию и документации хантыйского языка. В результате исследования выявлены отечественные и зарубежные цифровые базы данных, в которых представлен также материал по хантыйскому языку. Указанные онлайн-площадки имеют специальные программы с высокими инструментальными возможностями в области компьютерной лингвистики, что значительно облегчает изучение языков на современном этапе, среди них: виртуальная исследовательская среда «Ob-Ugric languages: conceptual structures, lexicon, constructions, categories» и «About Ob-Ugric Database» (2009, 2014), виртуальная лаборатория «ЛингвоДок» (2012), «Типологическая база данных угорских языков» (Typological Database of the Ugric Languages) (2012–2015) и др. Данные ресурсы содержат оцифрованные материалы по хантыйским диалектам с целью систематизации, анализа, а в дальнейшем они могут служить ревитализации языка. Проекты предлагают описание и анализ по всем уровням языковой системы. Ценность также заключается в том, что указанные виртуальные площадки предлагают гибкий пользовательский запрос по различным аспектам лингвистического исследования.

Ключевые слова: хантыйский язык, компьютерная лингвистика, цифровая платформа, информационно-коммуникационные технологии.

DOI: 10.35634/2224-9443-2024-18-1-109-114

С развитием цифровых технологий¹ появляются новые возможности в сфере сохранения языка и развития коммуникационного ландшафта как для хантыйского языка, так и для других миноритарных языков. Сегодня обыденное явление – использование поисковых систем, переводчика текста и голосовых помощников. Благодаря компьютерной лингвистике стало возможным автоматизировать обработку естественного языка. Система искусственного интеллекта выполняет заданные алгоритмом действия по систематизации и анализу языковых данных, а исследователь принимает окончательное решение о корректности эксперимента.

Хантыйский язык – один из миноритарных языков Сибири. Научное изучение хантыйского языка началось в середине XIX в. (см. например, работу М. А. Кастрена [Castrén 1858]), однако этому предшествовала работа по накоплению лексикографических сведений. На современном этапе, как отмечает Н. Б. Кошкарева, «изучение языка и культуры народа ханты заметно интенсифицировалось: в последние годы появились новые работы по хантыйской этнографии, фольклору и языкознанию, опубликовано несколько сборников фольклорных произведений, адресованных разным категориям читателей – и научным сотрудникам, и детям младшего возраста» [Кошкарева 2013, 3]. Сегодня – благодаря цифровизации – в сети Интернет появились российские и зарубежные лингвистические платформы, где представлены материалы по тем или иным диалектам хантыйского языка, которые являются объектом исследования. Прогресс цифровых технологий способствует решению множества задач в сфере сохранения языков, находящихся на грани исчезновения, это и документация языка (резервирование в цифровых хранилищах), и формирование медиасреды по продвижению и популяризации языка (создание видеороликов, мультфильмов, аудио- и видеоуроков, приложений для гаджетов, направленных на изучение языка и др.).

¹ Работа выполнена при поддержке гранта РНФ № 20-18-00403 «Цифровое описание диалектов уральских языков на основании анализа больших данных».

В данной работе рассмотрим цифровые онлайн площадки, в которых осуществляется документация и анализ хантыйского языка.

Материалы и методы

Объектом исследования являются специальные лингвистические платформы по изучению и документации языков, основанные в рамках различных проектов, где представлены массивы данных по хантыйскому языку. Данные виртуальные лаборатории используют различные способы хранения, обработки и анализа данных. Материалы, представленные в них, находятся в открытом доступе.

Для решения поставленных задач использовались: классификация, аналитический и описательный методы.

Результаты

Первые письменные памятники и научное изучение хантыйского языка

В истории хантыйского языкознания имеются сведения о первых памятниках письменности, в которых зафиксированы хантыйские слова или небольшие словники, включающие названия некоторых явлений, предметов. Например, «Первые хантыйские слова и фрагменты текстов (молитв) приводятся в труде Н. Витзена («Noord en Oost Tartarye». Amsterdam, 1692), который в 1664–1665 гг. записал в Москве эти сведения от русских, побывавших в Югорской земле и знакомых с языком ‘остяков’» [Основы финно-угорского языкознания... 1976, 249]. Известны словари, которые составлялись по академической программе с 30-х гг. XVIII века. Один из них был составлен Федором Кушкиным «Название остяцкое написано российским слогом, с переводами на остяцкий язык...», в нем впервые «дается характеристика языка по диалектам с довольно подробным указанием на их территориальное распространение» [Основы финно-угорского языкознания... 1976, 244]. Довольно большой по объёму русско-хантыйский словарь, составленный священником Вологодским (1842 г.), хранится в Архиве РАН в Санкт-Петербурге. Подробный анализ графической системы указанного памятника письменности осуществлен в статье [Кашкин 2020, 30–40].

У истоков научного изучения хантыйского языка стояли такие ученые, как Матиас Кастрен, Антал Регули, Йожеф Папай, Август Альквист, Куста Карьялайнен, Хейкки Паасонен, Серафим Патканов и др. Труды этих ученых до сих пор являются востребованными в научной среде, так как на современном этапе некоторые диалекты (или говоры), грамматическое описание которых имеется в этих работах, уже исчезли с лингвистической карты мира. Например, в своей работе «Опыт перевода остяцкой грамматики с кратким словарем, составленной Др. М. Александром Кастреном» (1849 г.) описывается «наречие, при Иртыше встречаемое» [Кастрен 1849, 5], а также приводятся некоторые материалы по сургутскому диалекту. Лексикографический материал по диалектам хантыйского языка (иртышскому, кондинскому, березовскому, обдорскому, сургутскому, ваховскому и верхнеобскому) собран К. Карьялайненом [Николаева 1995, 3–4]. Словарь Карьялайнена включает диалекты, которые на сегодняшний день утрачены – иртышский и кондинский.

В XX в. значительный вклад в изучение хантыйского языка внес видный немецкий финно-угровед В. Штейниц, издавший ряд исследований по хантыйскому языку и фольклору. Большой популярностью пользуется словарь Вольфганга Штейница [Steinitz 1966–1991], в котором представлен наибольший охват лексики хантыйских диалектов, он включает лексику всех диалектов из всех доступных во время его составления источников. Вклад в изучение хантыйского языка и его диалектов в разные годы XX в. и начала XXI в. внесли зарубежные и отечественные исследователи: К. Редди, Я. Гуя, Э. Вертеш, М. Чепреги, Г. Ганшов, Н. И. Терешкин, Ю. Н. Русская, Е. А. Немысова, Н. Б. Кошарева, В. Н. Соловар, А. Д. Каксин и другие.

Исследование и сохранение хантыйского языка на современном этапе

В XXI в. изучение хантыйского языка и его диалектов продолжают в Институте языкознания Российской академии наук (г. Москва), Институте филологии СО РАН (г. Новосибирск), в лаборатории «Лингвистические платформы» института системного программирования им. В. П. Иванникова (г. Москва), в научно-учебной лаборатории по формальным моделям в лингвистике научно-исследовательского университета Высшей школы экономики (г. Москва), в Обско-угорском институте прикладных исследований и разработок (г. Ханты-Мансийск), а также в Венгрии, Эстонии, Финляндии и Германии.



В настоящее время исследование хантыйского языка выходит на новый уровень: создаются виртуальные исследовательские лаборатории, специальные программы и платформы, способные обрабатывать лингвистический материал более совершенным способом. Первая виртуальная исследовательская среда под названием «Обско-угорские языки: концептуальные структуры, лексика, конструкции, категории» ("Ob-Ugric languages: conceptual structures, lexicon, constructions, categories") была создана в августе 2009 года как совместный проект европейских университетов: Мюнхенского университета Людвига-Максимилиана (Германия), Венского университета (Австрия), Хельсинкского университета (Финляндия), а также университетов Этвёша Лоранда и Сегеда (Венгрия). Программа исследования заключается в «предоставлении описательных ресурсов в онлайн-формате и современного лингвистического анализа двух родственных и находящихся под угрозой исчезновения обско-угорских языков – хантыйского (остяцкого) и мансийского (вогульского)». Структура проекта включает: Библиографическую базу данных, электронную библиотеку, Текстовый корпус, электронную грамматику, электронный словарь [https://www.babel.gwi.uni-muenchen.de/index.php?abfrage=welcome_eurobabel&navi=about].

Своё развитие данный проект получил в следующей программе «Обско-угорская база данных», запущенной в июле 2014 года. Исследователи решили продолжить и расширить работу «посредством систематизации, оцифровки, анализа ещё двух обско-угорских диалектов (западномансийского и юганских ханты)» [12]. Анализ и описание диалектов охватывал следующие области: а) фонологический анализ в форме IPA (вместо традиционных идиосинкратических систем транскрипции), б) определение морфологических категорий и алломорфов в данных диалектах, а также парадигмы и модели позиций для частей речи, установленных в результате анализа алломорфов. Презентуемые материалы в дальнейшем могут служить целям языковой ревитализации. Кроме того, доступ к базе данных был предоставлен Языковому архиву (TLA) Института психолингвистики Общества Макса Планка (Неймеген, Нидерланды).

С 2012 г. разрабатывается виртуальная лаборатория ЛингвоДок (lingvodoc.ispras.ru). Lingvodoc – совместный проект ИСП РАН, Института языкознания РАН и Томского государственного университета. Система предназначена для совместной многопользовательской документации исчезающих языков, создания многослойных словарей и научной работы с полученными звуковыми и текстовыми данными.

Лингвистическая платформа предлагает анализ звуковых данных с помощью экспериментально-фонетической программы и специальной компьютерной программы Phonology, которая позволяет оценить качество выполненной транскрипции на основе анализа спектрограмм [Норманская 2020, 631]. Представлена опция «Глоттохронологический анализ языков/диалектов», которая создана в 2023 г., также имеется ряд других алгоритмов для анализа лингвистического материала. На сегодняшний день база этого ресурса составляет более 1.000 аудиословарей и 300 корпусов текстов по диалектам языков народов мира, включая хантыйский язык. Представлены уникальные данные по исчезающим языкам России [<http://lingvodoc.ru/?ysclid=lq3fr5zj105932715>]. База по хантыйскому языку включает 48 словарей, как памятников письменности, так и современных полевых материалов.

Реализация пилотного проекта «Типологическая база данных угорских языков» (UTDB) осуществлена на базе Кафедры финно-угроведения университета им. Лоранда Этвёша под руководством Ф. Хаваша [<http://en.utdb.nullpoint.info/>]. В результате работы был создан электронный ресурс языков угорской подгруппы финно-угорской ветви: венгерского (стандартный вариант), мансийского (северный диалект) и хантыйского (сынский и сургутский диалект) с целью типологического исследования морфонологических, морфологических, морфосинтаксических и синтаксических черт рассматриваемых языков. Количество проанализированных параметров указанных языков, в плане как структуры, так и порядка слов, превышает 200 единиц.

На сайте Института филологии СО РАН представлен Электронный диалектологический атлас уральских языков, распространенных на территории Ямало-Ненецкого автономного округа (реализован в 2014–2016 гг.). Указанный электронный ресурс включает мультимедийные диалектологические словари коми-зырянского, ненецкого, селькупского и хантыйского языков (каждый не менее чем на 3500 входов). Отмечается, что в «базе реализован пользовательский поиск по нескольким полям, в частности по лексеме, алломорфу, части речи и переводу» [<https://www.philology.nsc.ru/resources/atlas.php>]. Также есть возможность получать мультимедийное картографическое представление материала для заданного множества языков или лексемы с помощью встроенного модуля «Карта».

На сайте Обско-угорского института прикладных исследований и разработок представлены электронные ресурсы, в том числе, содержащие данные по хантыйскому языку [<https://ouipiir.ru>]. Диалектологический атлас обско-угорских и самодийских языков; объём записей на хантыйском языке – более 6000 словоупотреблений (<https://atlas.ouipiir.ru>). Для пользовательского запроса выдает следующую информацию: язык, лексему в практической орфографии и транскрипции, часть речи, перевод и пример предложения, а также звук. В 2024 году планируется запуск новой опции «Формообразование». Депозитарий, имеющий аудио- и видеозаписи на хантыйском языке, а также текстовые материалы (<https://folk.ouipiir.ru>). Кроме этого, следует отметить «Электронную библиотеку», в которой находятся в оцифрованном виде бумажные издания на родных языках, выходящие в Обско-угорском институте.

Коллектив научных сотрудников из Института народов Севера РГПУ им. А. И. Герцена участвует в разработке электронного ресурса «Идеографическое описание лексики языков коренных малочисленных народов Севера, Сибири и Дальнего Востока РФ (эвенский, нанайский, долганский, нивхский, корякский, юкагирский, мансийский, хантыйский, селькупский, вепский, саамский)». Данный ресурс (<http://dict.su/>) является стартовым для лексикографической базы данных и основан на лексикографических, фольклорных источниках, текстовых и полевых записях.

Заключение

Итак, на современном этапе с развитием диджитал-технологий сохранение и анализ хантыйского языка находится на новом витке развития. Зарубежные и отечественные исследователи внесли большой вклад в документацию и изучение хантыйского языка. Исследование хантыйского языка, наряду с другими языками, ведётся в ведущих российских и зарубежных научных центрах. Организованы специальные лингвистические платформы с различными инструментами анализа, в которых помимо других языков, есть словарный, фото, аудио- и видеоматериал по хантыйским диалектам.

К наиболее прогрессивным относятся такие виртуальные лаборатории, как веб-сайт с двумя лингвистическими проектами OUL (Обско-угорские языки) и OUDB (Обско-угорская база данных); лингвистическая платформа ЛингвоДок 3.0; сайт UTDB (Типологическая база данных угорских языков). Представлены электронные ресурсы с диалектологическими атласами уральских языков, распространённых на территории Ямала и обско-угорских и самодийских языков в ХМАО-Югре, а также формируется онлайн-ресурс по идеографическому описанию лексики языков коренных малочисленных народов Севера, Сибири и Дальнего Востока Российской Федерации.

ЛИТЕРАТУРА

Идеографическое описание лексики языков (эвенский, нанайский, долганский, нивхский, корякский, юкагирский, мансийский, хантыйский, селькупский, вепский, саамский). URL: <http://dict.su/> (дата обращения: 13.12.2023).

Кастрен М. А. Опыт перевода остяцкой грамматики с кратким словарем, составленной Др. М. Александром Кастреном. СПб: Императорская академия наук, 1849. 125 с.

Кашкин Е. В. Хантыйский словарь священника Вологодского: особенности графической системы // Томский журнал лингвистических и антропологических исследований. 2020. 4 (30). С. 30–40.

Кошкарёва Н. Б. Хантыйская письменность. Ханты-Мансийск, 2013. 47 с.

Лингвистическая платформа Lingvodoc 3.0. URL: <http://lingvodoc.ru/?ysclid=lq3frr5zzj105932715> (дата обращения: 13.12.2023).

Николаева И. А. Обдорский диалект хантыйского языка. М., Гамбург: Институт языкознания РАН, *Mitteilungen der Societas Uralo-Altaica*, 1995. 256 с.

Норманская Ю. В. Коми-язвинский – диалект коми-пермяцкого или отдельный язык? // Ежегодник финно-угорских исследований. Том 14. № 4. 2020. С. 630–643.

Обско-угорский институт прикладных исследований и разработок. URL: <https://ouipiir.ru> (дата обращения: 13.12.2023).

Основы финно-угорского языкознания. Марийские, пермские и угорские языки. М.: Наука, 1976.

Электронный диалектологический атлас уральских языков, распространённых на территории Ямало-ненецкого автономного округа. URL: <https://www.philology.nsc.ru/resources/atlas.php> (дата обращения: 13.12.2023).

About Ob-Ugric languages: conceptual structures, lexicon, constructions, categories. URL: https://www.babel.gwi.uni-muenchen.de/index.php?abfrage=welcome_eurobabel&navi=about (дата обращения: 13.12.2023).



About Ob-Ugric Database. URL: https://www.babel.gwi.uni-muenchen.de/index.php?abfrage=-welcome_oudb&navi=about_oudb&subnavi (дата обращения: 13.12.2023).

Castrén M. A. Versuch einer ostjakischen Sprachlehre nebst kurzem Wörterverzeichnis. St.-Petersburg: Buchdruckerei der Kaiserlichen akademie der wissenschaften, 1858.

Steinitz W. Dialektologisches und etymologisches Wörterbuch der ostjakischen Sprache. Berlin: Akademie Verlag, 1966–1991. 2018 p.

Typological Database of the Ugric Languages. URL: <http://en.utdb.nullpoint.info/> (дата обращения: 13.12.2023).

Поступила в редакцию 15.01.2024

Молданова Ирина Максимовна

кандидат филологических наук, ведущий научный сотрудник
Обско-угорский институт прикладных исследований и разработок
628011, Россия, ХМАО-Югра, г. Ханты-Мансийск, ул. Мира, д. 14А
младший научный сотрудник Лаборатории «Лингвистические платформы»
Институт системного программирования им. В. П. Иванникова РАН
109004, Россия, Москва, ул. Солженицына, 25
E-mail: Uil.gera@yandex.ru
ORCID ID: 0000-0003-2992-2247

I. M. Moldanova

KHANTY LANGUAGE: FROM PAPER MANUSCRIPTS TO DIGITAL PLATFORMS

DOI: 10.35634/2224-9443-2024-18-1-109-114

The article examines a modern virtual environment in which, through linguistic platforms, the Khanty language is preserved and studied. The scientific study of the Khanty language, which began in the XIX century, was preceded by fragmentary records of the Khanty vocabulary known since the XVII century. At the present stage the study and backup of language data is carried out using the latest digital technologies, while the research materials are publicly available, which increases the value of language databases. The purpose is to consider modern online platforms where material on the research and documentation of the Khanty language is accumulated. As a result of the research, domestic and foreign digital databases have been identified, which also contain material on the Khanty language. These online platforms have special programs with high instrumental capabilities in the field of computational linguistics, which greatly facilitates the study of languages at the present stage, among them: the virtual research environment "Ob-Ugric languages: conceptual structures, lexicon, constructions, categories" and "About Ob-Ugric Database" (2009, 2014), virtual laboratory "Lingvodoc" (2012), "Typological Database of the Ugric Languages" (2012-2015), etc. These resources contain digitized materials on Khanty dialects for the purpose of systematization, analysis, and in the future they can serve to revitalize the language. The projects offer description and analysis at all levels of the language system – phonological, morphological and syntactic. The value also lies in the fact that these virtual platforms offer a flexible user query on various aspects of linguistic research.

Keywords: Khanty language, computational linguistics, digital platform, information and communication technologies.

Citation: Yearbook of Finno-Ugric Studies, 2024, vol. 18, issue 1, pp. 109–114. In Russian.

REFERENCES

Ideograficheskoe opisanie leksiki yazykov (evenskiy, nanayskiy, dolganskiy, nivkhskiy, koryakskiy, yukagirskiy, mansiyskiy, khantyyskiy, sel'kupskiy, vepsskiy, saamskiy) [Ideographic description of the vocabulary of languages (Even, Nanai, Dolgan, Nivkh, Koryak, Yukaghir, Mansi, Khanty, Selkup, Veps, Sami)]. Available at: <http://dict.su/> (Accessed December 13, 2023). In Khanty, In Russian.

Kastren M. A. Opyt perevoda ostyatskoy grammatiki s kratkim slovarem, sostavlennoy Dr. M. Aleksandrom Kastrenom [The experience of translating the Ostyak grammar with a short dictionary compiled by Dr. M. Alexander Kastren]. S.-Peterburg: Imperatorskaya akademiya nauk Publ., 1849. 125 p. In Russian.

Kashkin E. V. Khantyyskiy slovar' svyashchennika Vologodskogo: osobennosti graficheskoy sistemy [The Khanty dictionary of the Vologda priest: features of the graphic system]. Tomskiy zhurnal lingvisticheskikh i antropologicheskikh issledovaniy [Tomsk Journal of Linguistic and Anthropological Research], 2020, no. 4 (30), pp. 30–40. In Russian.



- Koshkareva N. B.** Khantyyskaya pis'mennost' [Khanty writing]. Khanty-Mansiysk, 2013. 47 p. In Russian. Lingvisticheskaya platforma Lingvodoc 3.0 [Lingvodoc 3.0 linguistic platform]. Available at: <http://lingvodoc.ru/?ysclid=lq3fr5zzj105932715> (Accessed December 13, 2023). In English, In Russian, In Khanty.
- Nikolaeva I. A.** Obdorskiy dialekt khantyyskogo yazyka [Obdorsky dialect of the Khanty language]. M., Gamburg: Institut yazykoznaniya RAN, Mitteilungen der Societas Uralo-Altaica Publ., 1995. 256 p. In Russian.
- Normanskaya Yu. V.** Komi-yaz'vinskiy – dialekt komi-permyatskogo ili otdel'nyy yazyk? [Is Komi-Yazvinsky a dialect of Komi-Permian or a separate language?]. Ezhegodnik finno-ugorskikh issledovaniy [Yearbook of Finno-Ugric Studies], Vol. 14, no. 4. 2020. pp. 630–643. In Russian.
- Obsko-ugorskiy institut prikladnykh issledovaniy i razrabotok [Ob-Ugric Institute of Applied Research and Development]. Available at: <https://ouipiir.ru> (Accessed December 13, 2023). In Russian, In Khanty.
- Osnovy finno-ugorskogo yazykoznaniya. Mariyskie, permskie i ugorskie yazyki [Fundamentals of Finno-Ugric linguistics. Mari, Permian and Ugric languages]. M., Nauka Publ., 1976. pp. 249. In Russian.
- Elektronnyy dialektologicheskiy atlas ural'skikh yazykov, rasprostranennykh na territorii Yamalonenetskogo avtonomnogo okruga [Electronic dialectological atlas of the Uralic languages spoken in the Yamalo-Nenets Autonomous Okrug]. Available at: <https://www.philology.nsc.ru/resources/atlas.php> (Accessed December 13, 2023). In Russian, In Khanty.
- About Ob-Ugric languages: conceptual structures, lexicon, constructions, categories. Available at: https://www.babel.gwi.uni-muenchen.de/index.php?abfrage=welcome_eurobabel&navi=about (Accessed December 13, 2023). In English, In Russian, In Deutsch, In Khanty.
- About Ob-Ugric Database. Available at: https://www.babel.gwi.uni-muenchen.de/index.php?abfrage=welcome_oudb&navi=about_oudb&subnavi (Accessed December 13, 2023). In English, In Russian, In Deutsch, In Khanty.
- Castrén M. A.** Versuch einer ostjakischen Sprachlehre nebst kurzem Wörterverzeichnis. St.-Petersburg: Buchdruckerei der Kaiserlichen akademie der wissenschaften Publ., 1858. pp. 125. In Deutsch.
- Steinitz W.** Dialektologisches und etymologisches Wörterbuch der ostjakischen Sprache. Berlin: Akademie Verlag Publ., 1966–1991. 2018 p. In German.
- Typological Database of the Ugric Languages. Available at: <http://en.utdb.nullpoint.info/> (Accessed December 13, 2023). In English, In Russian, In Khanty.

Received 15.01.2024

Moldanova Irina Maksimovna

Candidate of Philological Sciences, Leading Researcher
Ob-Ugric Institute of Applied Research and Development
14A, Mira st., Khanty-Mansiysk, KHMAO-Yugra, 628011, Russia
Junior Researcher of the Laboratory «Linguistic Platforms»
Institute for System Programming by V. P. Ivannikov of the Russian Academy of Sciences
25, Solzhenitsyn st., Moscow, 109004, Russia
E-mail: Uil.gera@yandex.ru
ORCID ID: 0000-0003-2992-2247