**LINGUAE ORIENTALES**

|  |
| --- |
|  |

DOI: 10.31696/2618-7302-2020-4-295-301

**КОРПУС ЯЗЫКОВ ГРУППЫ КОРАПУТСКИХ МУНДА:**

**ЗАДАЧИ И СПЕЦИФИКА**

© 2020 **А. С. Крылова, Е. А. Ренковская**[[1]](#footnote-2)

Статья посвящена первому электронному корпусу текстов на языках группы корапутских мунда (сора, гутоб, бонда), размещенному в сети Интернет весной 2020 г. Корапутские мунда распространены в Индии на границе штатов Орисса и Андхра-Прадеш, все в той или иной мере находятся под угрозой исчезновения. Тексты на этих языках были собраны в ходе четырех экспедиций в штат Орисса (Индия) в 2016–2018 гг. Носители корапутских мунда проживают отдельными общинами, различающимися вероисповеданием, историческим родом занятий, диалектами, а также штатом проживания и соответственно влиянием разных официальных языков штатов на родной язык. Так, например, представители народности сора исповедуют более шести разных религий и пользуются четырьмя видами письменности. Одна из основных задач корпуса — представить в своем составе тексты различных жанров и разных сфер бытования языков. На данный момент в корпус вошли устные и письменные тексты, поэтические и прозаические, религиозного, фольклорного и бытового содержания. Особый интерес уделяется песням-импровизациям — автохтонному жанру мунда. Устные тексты представлены как в фонологической записи, так и в аудио- и видеофиксации. Подкорпус письменных текстов, представленных в различных графических записях, содержит как тексты, относящиеся к определенному рукописному жанру, так и образцы печатной продукции. Тексты снабжены морфологической разметкой и переводом на русский и английский языки. Каждый текст сопровождается подробной социолингвистической и жанрово-специфической информацией. Одна из важнейших особенностей корпуса — система меток, включающая формат текста, пол информанта, письменность, жанр, тематику, религию и др. Данный проект призван не только сделать языковые материалы корапутских мунда доступными для мировой науки, но и быть полезным для преподавания и сохранения культурного наследия в регионе распространения этих языков, в частности в рамках государственной программы Индии “Multi-Language Education”.

*Ключевые слова:* языки мунда, корпус, этнолингвистика, социолингвистика.

*Для цитирования:* Крылова А. С., Ренковская Е. А. Корпус языков группы корапутских мунда: задачи и специфика. *Вестник Института востоковедения РАН.* 2020. № 4. С. 295–301. DOI: 10.31696/2618-7302-2020-4-295-301

**CORPUS OF KORAPUT MUNDA LANGUAGES:**

**GOALS AND SPECIAL ASPECTS**

**Anastasiya S. Krylova, Evgeniya A. Renkovskaya**

The paper deals with the first digital corpus of texts in the Koraput Munda languages (Sora, Gutob, Bonda), which became available online in Spring 2020. Koraput Munda are spoken in India on the border between states of Odisha and Andhra Pradesh and they all are more or less endangered. Texts in these languages were collected during four expeditions to the state of Odisha in 2016–2018. Koraput Munda speakers live in communities, which differ in religions, traditional occupations, dialects and are influenced by various official languages depending on the state. For example, Sora speakers belong to more than six religious communities and use four types of writing. Therefore, one of the main tasks of the corpus is to present texts of various genres and different social conditions of language usage. At the moment, the corpus includes oral and written texts, poetry and prose, religious, folklore and traditional everyday content. Oral texts are presented both in phonological transcription and in audio and video recordings. The sub-corpus of written texts presented in various scripts contains both texts related to a particular handwritten genre, as well as samples of printed materials. The texts are provided with morphological markup and translation into Russian and English. Each text is accompanied by detailed sociolinguistic and genre-specific information. One of the most special features of the corpus is the system of tags including text format, speaker’s gender, script, genre, topic, religion etc. This project is intended not only to make linguistic materials of the Koraput Munda languages accessible for the global linguistic and anthropological studies, but also to be useful for teaching and preserving cultural heritage, in particular within the framework of the Multi-Language Education government program.

*Keywords:* Munda languages, corpus, ethnolinguistics, sociolinguistics.

*Для цитирования:* Krylova A. S., Renkovskaya E. A. Corpus of Koraput Munda Languages: Goals and Special Aspects. *Vestnik Instituta vostokovedenija RAN*. 2020. 4. Pp. 295–301. DOI: 10.31696/2618-7302-2020-4-295-301

Статья посвящена разрабатываемому авторами электронному корпусу текстов на языках группы корапутских мунда. Сайт корпуса был размещен в сети Интернет по адресу mundastudies.info в апреле 2020 г., на данный момент имеет русскую и английскую версии и содержит тексты на трех языках — сора (саора, савара), гутоб (гадаба) и бонда (ремо). Тексты и другие языковые материалы по корапутским мунда были собраны авторами непосредственно в ходе четырех экспедиций в город Бхубанешвар и округа Гаджапати, Раягада и Корапут штата Орисса (Индия) в 2016–2018 гг.

Корапутские мунда распространены в Индии на территории двух граничащих между собой штатов Орисса и Андхра-Прадеш. К ним относятся такие языки, как сора, джурай (статус этого языка до сих пор не определен, часто его относят к диалектам сора), гутоб, бонда, дидаи и горум. Генетически мунда в целом представляют собой самую западную ветвь аустроазиатских языков, однако являются ли корапутские мунда таксоном генетической классификации или лишь ареальным объединением, до сих пор в точности не установлено.

Социолингвистическая ситуация в ареале языков мунда представляет собой на данный момент сложную многоплановую картину. Языки корапутских мунда находятся в непосредственном соседстве с индоарийскими, а также дравидийскими племенными языками. На территории проживания корапутских мунда, таким образом, можно наблюдать употребление в разных сферах следующих идиомов: собственно родные языки племен и их диалекты, языки соседних племен и народностей, индоарийский десия как лингва франка в округе Корапут, ория и телугу как официальные языки штатов, хинди и английский как официальные языки страны [Крылова, Ренковская, 2019, c. 65–66]. Все языки корапутских мунда в той или иной степени находятся под угрозой исчезновения. Современные процессы модернизации, господство индоарийских языков и английского в СМИ, интернете, школе и государственных учреждениях Индии, «непрестижность» языков мунда вследствие традиционного для индуизма представления о «неприкасаемости» и «неразвитости» племен, — все эти явления ведут к языковой и культурной ассимиляции и утрате языка младшим поколением.

Несмотря на то что изучение языков мунда является достаточно популярным направлением ввиду генетической специфики этих языков, корапутские мунда остаются малоизученными. До сих пор не существует полноценных грамматических описаний этих языков. Современные исследования в области языков корапутских мунда в основном сосредоточены на сравнительно-исторических и типологических аспектах, тогда как социолингвистическая ситуация и диалектология остаются за пределами внимания ученых, поскольку их изучение невозможно без полевой работы. При этом любое исследование данных языков сопряжено с большим числом проблем, среди которых не только труднодоступность районов проживания носителей этих языков, но и сложная и неоднозначная система получения допуска на племенные территории для иностранных граждан. Если работа европейских ученых в округах Гаджапати и Раягада осложняется негласным противодействием правительства Индии христианскому миссионерству из западных стран, то округ Малкангири, основное место проживания племен бонда и дидаи, закрыт для посещения иностранными гражданами ввиду нестабильной ситуации с наксалитским повстанческим движением, действующим в регионе.

**Актуальность и новизна задачи по созданию корпуса**

Созданный нами корпус является первым электронным корпусом языков мунда. В основу создания корпуса легли социолингвистическое и этнолингвистическое исследования языков корапутских мунда, включающие среди всего прочего изучение сфер употребления этих языков, вопроса о наличии диалектов, о статусе и распространении различных видов письменности, а также о процессах языкового активизма и языкового строительства, проводимого в регионе. Согласно изначально поставленным задачам, текстовый материал, представленный в корпусе, призван отразить различные сферы и условия бытования корапутских мунда, создавая, таким образом, наиболее полную картину их функционирования в языковом сообществе. На данный момент в корпусе представлены 14 текстов на языке сора, 6 на гутоб и 5 на бонда.

Отражение социолингвистической ситуации и сфер функционирования языка в корпусе представляется нам основной из задач по нескольким причинам. Во-первых, традиционный образ жизни племен корапутских мунда остается по сей день практически неизученным, антропологические исследования в этой области единичны. Во-вторых, в жизни племен в настоящее время происходит огромное количество преобразований, во многом изменяющих их традиционный уклад жизни. Такие преобразования обусловлены не только такими общими факторами, как распространение в регионе образования и средств массовой информации, но и уникальными, как, например, успешный прозелитизм различных религий и конфессий среди сора, массовое переселение гутоб из мест компактного проживания в связи со строительством гидроэлектростанций и затоплением деревень, а также внедрение программы MultiLingual Education (MLE) — обучения на родных языках в младших классах школы среди сора и бонда.

На данный момент языковые сообщества корапутских мунда являются помимо всего прочего внутренне социально неоднородными. Наиболее неоднородным является языковое сообщество сора. По вероисповеданию сора относятся более чем к шести различным религиям и конфессиям [Krylova, Renkovskaya, 2020]. Помимо сорских общин, придерживающихся традиционных верований, а именно исповедующих существовавший у сора на протяжении веков анимизм, есть баптистская, католическая и индуистская общины (сформировавшиеся в результате христианского и индуистского прозелитизма), индуистское религиозное течение Махима Дхарма (более известное как Алекх Дхарма, распространенное в штате Орисса среди низких каст и племен как протест против брахманизма), а также религиозное течение Матар Баном — неоанимисты, почитающие в качестве божества новоизобретенную для языка сора письменность [Zide, 1999, p. 200; Крылова, 2018, c. 123]. Внутри каждой религиозной общины складывается своя особая социолингвистическая ситуация, основанная на различной роли родного языка в религии, а также роли самой религии в жизни общины, стремлении к религиозной ассимиляции или, наоборот, сегрегации и, как следствие, разной степени утраты языка.

Язык в различных религиозных общинах сора также претерпевает некоторые изменения. Так, антрополог П. Витебски, много лет занимающийся сора, в [Vitebsky, 2017, p. 142, 217 и др.] замечает, что язык баптистских текстов отличается заметной искусственностью. При этом, поскольку в баптистской общине язык сора активно используется в церкви и именно там его практикует молодежь, с детства обучающаяся за пределами ареала сора, то язык Библии становится некоторым эталоном сора, к тексту Библии обращаются для уточнения тех или иных оборотов. Таким образом, можно говорить, что в баптистской общине язык сора используется уже в несколько модифицированном виде.

Пришедший к сора индуизм (в адаптированном для племен виде) ориентируется на уже сложившуюся баптистскую традицию, и индуистские религиозные тексты составляются на основе баптистских (вплоть до случаев элементарной замены имени Иисус на Рама и Кришна). Так, как отмечено в [Vitebsky, 2017, p. 222], для перевода английского *god* главой баптистской миссии, канадкой Анной Мунро, было выбрано слово *kituŋ*, при этом в анимистских верованиях оно обозначало класс существ, создавших мир, но не играющих большой роли в повседневной жизни людей в отличие от духов-«сонумов». Именно в баптистской общине сора *kituŋ* становится важным и часто употребляемым словом, и то, что в текстах неоиндуистов используется именно оно, а не, например, заимcтвование из санскрита или ория, свидетельствует об ориентации на христианские тексты. О сходстве баптистских и индуистских гимнов можно, в частности, судить по текстам, уже имеющимся в корпусе: так, например, вступительная формула *kituŋ asıntǝm*, не характерная для анимистских текстов,встречается в гимнах обеих этих религий, ср. (1)–(2):

(1) *ameŋ kituŋ <a>sın<tǝm>*

живой бог <NPST.1PL.EXCL/2SG> благодарить

Живой бог, мы благодарим тебя (начало баптистского гимна)

(2) *e gǝnʊgʊ kituŋ <a>sın<tǝm>*

 VOC великий бог <NPST.1PL.EXCL/2SG>благодарить

О великий бог, мы благодарим тебя (начало индуистского гимна).

При этом ориентации на текст Библии в индуизме нет, и большой роли язык сора в общине не играет.

Язык первых гимнов Матар Баном, входящих в религиозный канон, представляет собой искусственно созданный язык, явно являющийся подражанием санскриту и содержащий минимальное количество слов сора. Создание отдельного языка для священных текстов, как и создание автохтонной письменности, имеет типологическую параллель у племени хо группы северных мунда [Pinnow, 1972, p. 837] и требует дальнейшего изучения. Хотя в сложившейся у сора ситуации, вероятно, еще рано говорить об этнолектах, тенденция к их формированию несомненно присутствует.

На данный момент насчитывается пять разных видов письменности для языка сора [Крылова, Ренковская 2019б, с. 140]. Два типа латиницы были разработаны в баптистской общине. Первая сорская латиница содержала символы алфавита МФА, но впоследствии в целях упрощения стали использоваться только символы латинского алфавита. В католической общине были переняты оба типа орфографии. В религии Матар Баном используется письменность соранг сомпенг на оригинальной графической основе. Кроме того, в рамках программы «Многоязычного образования» (MultiLingual Education, MLE) для сора были разработаны еще два вида письменности — на графических основах ория (штат Орисса) и телугу (штат Андхра-Прадеш), поскольку, согласно программе, письменность для малых языков должна быть разработана на графической основе официальных языков штатов, где эти языки распространены. Были осуществлены попытки стандартизации этих двух письменностей, однако на данный момент ни одна из письменностей сора не является полностью нормализованной.

Все вышеперечисленные сферы употребления сора значимы для представления функционирования языка в целом. Таким образом, в корпусе для сора представлены религиозные и другие тексты, специфичные для разных религиозных общин, а также тексты на имеющихся видах письменности. Языковые общины гутоб и бонда не столь религиозно разнородны: в основном эти племена придерживаются анимистических верований, хотя у гутоб существует также достаточно многочисленная община религиозного течения Алекх Дхарма. Письменность существует из этих двоих языков только на бонда на основе графики ория, поскольку на бонда ведется образование в рамках программы MLE.

**Форма представления и организации материала в корпусе**

В первую очередь тексты сгруппированы по языковым разделам, а в случае сора и бонда разделены на письменные и устные. Страница текста в корпусе содержит метаданные, социолингвистическую информацию о тексте, включающую сведения о сфере функционирования и жанре, видео- и аудиоматериалы, транскрипцию в случае устных текстов или нормализованную транслитерацию в случае письменных текстов, грамматическую аннотацию, а также, в некоторых случаях, комментарии, в частности страноведческие, к отдельным словам текста. Аудиоматериалы представляют собой аудионарезку, соотносимую с небольшими фрагментами текста (предложениями или их частями). Видео снабжены субтитрами.

На данный момент в корпусе пока нет возможности поиска по лексемам, морфемам и грамматическим значениям, однако нами разработан поиск по различным социолингвистическим параметрам. Так, «Поиск по тэгам (меткам)» позволяет группировать тексты в зависимости от пола информанта, наличия того или иного формата представления текста (видео, аудио, глоссы), вида графики для письменных текстов, области функционирования, жанра и тематики текста. Тэги группируются в рубрики. Так, в рубрику «Области функционирования» входят тэги «Религия», «Фольклор» и «Многоязычное образование (MLE)», которыми маркируются соответственно тексты религиозного содержания и узуса, фольклорные тексты, а также тексты из учебной литературы программы MLE, отражающие методику преподавания родного языка в школах.

Рубрика «Тематика текста» применяется к тем бытовым нарративам, которые являются рассказами либо про традиционный уклад жизни и занятия племен (тэг «Традиции»), либо про жизнь племен, меняющуюся под влиянием современных процессов модернизации (тэг «Модернизация»): и то, и другое представляется важной с этнографической точки зрения информацией. В дальнейшем возможно изменение названия тэгов, их количества, а также названия и содержания рубрик по мере появления в корпусе новых текстов.

Внимание в корпусе уделяется не только сферам функционирования, но и жанровому своеобразию. У народов мунда есть специфические жанры, нехарактерные для индоарийских народов. К таким жанрам, например, относятся песни-импровизации ― песни, сочиняемые непосредственно в момент исполнения на тему того, что происходит вокруг. Среди письменных текстов представлены не только опубликованные тексты, но и тексты особых письменных жанров. К таким жанрам относится, например, письмо с предложением брака в баптистской общине сора: чисто эпистолярный жанр, содержащий специальные формулы. На данный момент в корпусе представлены тексты следующих жанров: песня-импровизация (3), сказка (4), индуистский гимн (2), баптистский гимн (1), легенда (1), бытовой нарратив (5), брачное письмо в баптистской традиции (1), христианская молитва (2) и др.

**Заключение**

Разрабатываемый нами корпус сделает данные языков корапутских мунда доступными для мировой науки и, как мы надеемся, будет полезен как для специалистов различных гуманитарных областей, как то лингвисты, филологи, фольклористы, этнографы и антропологи, так и для самих носителей этих языков, а также сможет быть использован в рамках преподавания, языкового строительства и сохранения культурного наследия в регионе распространения корапутских мунда.

Сокращения / Abbreviations

|  |  |
| --- | --- |
| 1 — первое лицо2 — второе лицоEXCL — эксклюзивNPST — непрошедшее время | PL — множественное числоSG — единственное числоVOC — вокативная частица |

Литература / References

Крылова А. С. История, структура и происхождение автохтонных письменностей языков мунда. *Восток (Oriens)*. 2018, № 3. C. 119–132 [Krylova A. S. History, Structure and Origin of the Munda Autochthonous Scripts. *Vostok (Oriens).* 2018. 3. P. 119–132 (in Russian)].

Крылова А. С., Ренковская Е. А. Некоторые аспекты социолингвистической ситуации племени гутоб. *Труды Института востоковедения РАН*. Вып. 27. *Проблемы общей и востоковедной лингвистики. Языки Азии и Африки*. М., 2020. С. 62–67 [Krylova A. S., Renkovskaya E. A. Some Aspects of the Sociolinguistic Situation of the Gutob Tribe. *Proceedings of the Institute of Oriental Studies RAS*. Vol. 27. *Problems of general and oriental linguistics. Languages of Asia and Africa*. Moscow, 2020. Pp. 62―67 (in Russian)].

Крылова А. С., Ренковская Е. А. Пять видов письменности языка сора: причины появления и процессы становления. *Международная конференция «Лингвистический форум — 2019: Коренные языки России и мира»* (Москва, Институт языкознания РАН, 4–6 апреля 2019 г.): Тезисы докладов. М., 2019. С. 140–141 [Krylova A. S., Renkovskaya E. A. Five types of Sora writing systems: causes, genesis and formation. *International Conference “Linguistic Forum — 2019: Indigenous Languages of Russia and the World” (Moscow, Institute of Linguistics, RAS, April 4–6, 2019): Abstracts*. Moscow, 2019. Pp. 140–141 (in Russian)].

Krylova A., Renkovskaya E. Religion and Language Preservation: Case of Sora. *Journal of South Asian Languages and Linguistics*. 2020. Vol. 7.1 (in print).

Pinnow H.-J. Schrift und Sprache in den Werken Lako Bodras im Gebiet der Ho von Singhbhum (Bihar). *Anthropos*. 1972. 67. S. 822–856.

Vitebsky P. *Living without the Dead: Loss and Redemption in a Jungle Cosmos.* Chicago, 2017.

Zide N. Three Munda Scripts. *Linguistics of Tibeto-Burman Area*. 1999. 22.2. Pp. 199–232.

1. Анастасия Сергеевна КРЫЛОВА, научный сотрудник Института востоковедения РАН, Москва; krylova\_anastasi@bk.ru

 Anastasiya S. KRYLOVA, Research Fellow, Institute of the Oriental Studies RAS, Moscow; krylova\_anastasi@bk.ru

 ORCID ID: 0000-0002-2286-3304

 Евгения Алексеевна РЕНКОВСКАЯ, младший научный сотрудник Института языкознания РАН, Москва; jennyrenk@gmail.com

 Evgeniya A. RENKOVSKAYA, Junior Research Fellow, Institute of Linguistics RAS, Moscow; jennyrenk@gmail.com

 ORCID ID: 0000-0003-1944-0746 [↑](#footnote-ref-2)