

ИСТОРИКО-ЭТИМОЛОГИЧЕСКИЙ СЛОВАРЬ ОСЕТИНСКОГО ЯЗЫКА В.И. АБАЕВА: ПРОБЛЕМЫ СОЗДАНИЯ ЦИФРОВОЙ ДВУЯЗЫЧНОЙ ВЕРСИИ

**О.И. Беляев, И.А. Хомченкова, Ю.В. Сеницына,
В.В. Дьячков, А.А. Бызова, А.О. Бадеев, Д.А. Алексеев,
Ю.Ю. Макаров**

*Московский государственный университет имени М.В. Ломоносова, Институт языкознания РАН, Институт русского языка
имени В.В. Виноградова РАН, НИУ ВШЭ, Москва, Россия;
Кембриджский университет, Кембридж, Великобритания*

Аннотация: В статье подводятся итоги проекта по оцифровке и переводу Историко-этимологического словаря осетинского языка В.И. Абаева, осуществляемого авторами статьи с 2020 г. По своим глубине и охвату этот словарь является уникальным не только среди словарей иранских языков, но и среди этимологических словарей в целом. По замыслу автора, он объединил в себе черты нескольких типов словарей: не только собственно этимологического, но и, в частности, двуязычного (осетинско-русского), документированного и исторического. В результате словарные статьи содержат информацию нескольких типов: перевод осетинского слова на русский, в ряде случаев — с достаточно подробным толкованием; иллюстрирующие примеры; развернутый исторический и культурологический комментарий (для слов, обозначающих значимые реалии); наконец, собственно этимологию, представляющую собой свободный текст с большим количеством цитируемых иноязычных форм (индоевропейских когнатов, источников заимствований и т. д.). В статье обосновывается необходимость максимального сохранения и кодирования этой информации при создании электронной базы данных. По нашему мнению, оптимальным форматом для таких задач является разметка XML, точнее — ее вариант, специально разработанный для разметки текстовых документов: TEI (Text Encoding Initiative, <https://tei-c.org/>). Для редактирования словаря была создана специальная модификация этой системы разметки, а также специализированный модуль для редактора Oxygen XML Editor и набор CSS-файлов, обеспечивающий максимальное приближение визуального облика словаря к исходному. Это же сочетание XML и CSS позволяет без дополнительной обработки сгенерировать печатный вариант словаря в русской и английской версиях при помощи системы Prince (<https://www.princexml.com/>).

Однако, хотя такой формат предпочтителен для максимально полного представления исходных данных, он не очень хорошо подходит для работы



с онлайн-интерфейсом и автоматической обработки данных, а поиск по XML слишком сложен и неоднозначен для обычного пользователя. Поэтому мы предлагаем дополнить XML-базу данных реляционной базой, построенной из словарных статей в автоматическом режиме. В нашем проекте такая база служит основой для электронной «оболочки» этимологического словаря (<https://ossetic.iranica.space>), построенной с использованием платформы OnLex [Makarov et al. 2022]. Таким образом, проект совмещает в себе черты сразу нескольких типов баз данных, что полностью соответствует многостороннему характеру исходного словаря. Мы надеемся, что опыт этого проекта может быть использован при работе над другими аналогичными словарями.

Ключевые слова: этимология; лексикография; осетинский язык; иранские языки; индоевропейские языки; базы данных

Финансирование: Исследование выполнено при поддержке РФФ, проект № 22-28-01639 «Создание двуязычной цифровой версии Историко-этимологического словаря осетинского языка В.И. Абаева».

doi: 10.55959/MSU0130-0075-9-2024-47-02-4

Для цитирования: Беляев О.И., Хомченкова И.А., Синицына Ю.В., Дьячков В.В., Бызова А.А., Бадеев А.О., Алексеев Д.А., Макаров Ю.Ю. Историко-этимологический словарь осетинского языка В.И. Абаева: проблемы создания цифровой двуязычной версии // Вестн. Моск. ун-та. Серия 9. Филология. 2024. № 2. С. 75–86.

V. I. ABAEV'S HISTORICAL-ETYMOLOGICAL DICTIONARY: ISSUES IN THE DEVELOPMENT OF A DIGITAL BILINGUAL EDITION

**O.I. Belyaev, I.A. Khomchenkova, J.V. Sinitsyna,
V.V. Dyachkov, A.A. Byzova, A.O. Badeev, D.A. Alekseev,
Y. Makarov**

*Lomonosov Moscow State University, Institute of Linguistics RAS, Vinogradov
Russian Language Institute RAS, HSE University, Moscow, Russia;
University of Cambridge, Cambridge, UK*

Abstract: In this article, we discuss the conclusions of the project on digitizing and translating the Historical-Etymological dictionary of Ossetic by V.I. Abaev, which has been carried out by the authors since 2020. The entries of the original include information of several kinds: Russian translations of an Ossetic word, related examples, usage and etymological notes. The etymological notes are usually free-form texts with a large number of cited forms from other languages (cognates, loanword sources, etc.). We argue for maximal preservation and thorough encoding of this information when creating a digital database. The optimal format for such purposes is XML markup, more specifically its variant intended for annotating text documents, TEI (Text Encoding Initiative, <https://tei-c.org/>). For digitizing the dictionary, we created a modification of this markup language, a module for Oxygen

XML Editor, and a set of CSS files that ensure maximum similarity between our layout and the original. The same combination of XML and CSS allows generating the print version of the dictionary using the Prince typesetting system (<https://www.princexml.com/>) without any additional processing. For ease of search and access, we suggest supplementing the XML sources with a relational database automatically created from the lexical entries. In our project, this database serves as the source for the online interface of the dictionary (<https://ossetic.iranian.space>), powered by the OnLex platform [Makarov et al. 2022]. Hence, the project combines the features of several database types, which fully corresponds to the multi-faceted character of the original dictionary. We hope that our experience can be used for other digital lexicographic projects.

Keywords: etymology; lexicography; Ossetic; Iranian languages; Indo-European languages; databases

Funding: The research is supported by Russian Science Foundation, project No. 22-28-01639 “Developing a Bilingual Digital Version of the Historical and Etymological Dictionary of the Ossetian Language by V.I. Abaev.”

For citation: Belyaev O.I., Khomchenkova I.A., Sinitsyna Yu.V., Dyachkov V.V., Byzova A.A., Badeev A.O., Alekseev D.A., Makarov Y. (2024) V.I. Abaev’s Historical-Etymological Dictionary: Issues in the Development of a Digital Bilingual Edition. *Lomonosov Philology Journal. Series 9. Philology*, no 2, pp. 75–86.

1. Введение

В настоящей статье описывается методология и результаты проекта по созданию электронной двуязычной цифровой версии Историко-этимологического словаря осетинского языка В.И. Абаева (далее ИЭСОЯ) [Абаев 1958–1989], выполняемого участниками проекта с 2020 г. Наш опыт может быть использован в широком спектре аналогичных проектов по созданию или цифровизации словарей различных типов, так как ИЭСОЯ является редким примером сочетания в одном источнике сразу нескольких видов словарей: по словам самого автора, «двуязычного (осетинско-русского), документированного, этимологического, исторического и реального» [Абаев 1973: 3]; таким образом, и выбранный нами формат его электронного представления должен был быть таким, чтобы органично сочетать в себе кодирование этих типов данных.

2. Историко-этимологический словарь осетинского языка: общий обзор

Статья ИЭСОЯ состоит из двух основных частей: в первой содержится перевод осетинского слова на русский язык, иллюстративные примеры, идиоматические и деривированные формы, а также — зачастую — социокультурные комментарии о функционировании слова или связанных с ним артефактов. Вторая часть представляет собой собственно этимологию — текст в свободной

форме, включающий большое число перекрестных ссылок на другие словарные статьи, цитируемых иноязычных слов (прежде всего, коннатов из иранских и других индоевропейских языков, но также потенциальных источников заимствования, объектов типологического сопоставления и т. п.), библиографических ссылок. Пример словарной статьи, содержащей все эти элементы, представлен на рис. 1. Первая часть начинается с основных форм лексемы (на иронском и дигорском диалектах, разделенные вертикальной чертой) и продолжается до символа «~», за которым следует этимологическая часть.

Формат электронного словаря должен быть выбран так, чтобы вся эта информация была представлена не в виде простого текста, но в структурированной форме в машиночитаемом формате, доступном для последующей обработки.

zærin | zærinæ 'золото', 'золотой', 'золотистый'. В самостоятельном употреблении находим обычно в сочетании с украшающим эпитетом *syɣ* 'красный': *syɣzærin* (из *syɣzærin*). Но в сложных словах господствует *zærin*: *zæringyrd* 'златокузнец', *zæringdag* 'золотая нить', *zærinbyd* 'золотая тесьма', *zærinbos* 'золотой шнур', *Xürzærin* 'Солнце золотое' (культовое наименование солнца). Ср. также *Zærin nonyx* 'Золотой ноных' — название праздника (Munk. Blüten 218), *Zrūgy zærin Majræm* 'Зругская золотая Богоматерь' (ЮОПам. I 95), *Zærinæ* женское имя. — *ærttytoj zykkūta zærinaw* «(ее) косы блестяли, как золото» (MD 1958 III 17); *zærin xūr arvastæw ærttyvy* «золотое солнце сияет посреди неба» (X а р. Уац. 45); *cæj bærc sæ ts æwzist, yzdy, zærin!* «сколько в них (в горах) серебра, свинца, золота!» (X а р. Кадæж 5); *dæwæj zæringær nlkæj yssardton* «более золотого, чем ты, я никого не нашла» (ЮОПам. III 170); д. *zærin bædtæntæ amonuj* «он (младенец в колыбели) требует золотых повязок» (СОПам. II 156); д. *raburdæj æ c'asæj saw zærinæxær xelagæ* «выползла из своей норы черная с золотистой головой змея» (DZ 204); д. *cæruj sæ xæssæ kalm zærin* «с ними живет золотистая змея» (Iraef 78).

~ Вместе с железом (*fsæn*) и сталью (*ændon*) относится к древнейшим названиям металлов в осетинском. Восходит к иран. **zaranya-*. Ср. перс. *zar(r)* (из **zar-na-*) 'золото', *zarrin* 'золотой', перс. **zarn* в арм. *zarnavaxi* 'златотканый' (Hübсhmann. PSt. 259), пехл. *zarr* 'золото', *zarrēn* 'золотой', парф. *zarnēn* id. (Al IXa 105), пам. ор. *zarnin*, *zarninin* 'золотой' (Зарубин. Орош. 70), согд. **zērn* 'золото', **zērnēnāk* 'золотой' (*zyrn, zyrnyk*), хорезм. *zirnī*, сак. *yisra* (из **zaranya-*) 'золото', *ysarrinaa-* 'золотой', ав. *zaranya-*, др.перс. *daraniya-* 'золото'. Сюда же *Zarina* имя сакской царицы (Ктесий). В топонимии Вост. Ирана: *Zarin, Zarang* (Daffina. L'im-migrazione dei Saka nella Drangiana. Roma, 1967, стр. 90—93). Приводимое Лукьяном (Токсарис) ζιρνυ 'выкуп' есть, вероятно, искаженное *zærin* и означало 'золото' как выкуп [Schmeja. Iranisches bei Lukian. Serta philologica aenipontana, II, Innsbruck, 1972, стр. 25—27 («Sarmatisch Ζιρνυ»)]. Содержащаяся в этих словах индоевропейская база *ghel-* 'желтый', 'красноватый' распознается также в ст.слав. *zlato*, русск. *золото*, нем. *Go'd* и пр. — Из иранского вошло в некоторые угро-финские языки: венг. *arany*, коми, удм. *zarni*, морд. *sirne*, ханты *sarna*, манси *sarñ*, *soññ*. — См. еще *syɣzærin, zærinbos, zæringdag, zæringyrd, Xürzærin*.

Вс. Миллер. ОЭ II 47, 82, III 138; Gr. 16. — Hübсhmann. Oss. 234. — ОЯФ I 33, 39, 52, 83, 190, 200. — Joki. Uul 29, 63, 82, 114, 250. — Bailey. Saka Dictionary 347, 352 sq.

Рис. 1. Словарная статья *zærin* 'золото' [Абаев 1989: 303]

3. Подходы к созданию цифровых словарей

Самый распространенный формат представления словарных данных — это таблица или система таблиц, например набор текстовых файлов в формате CSV (Comma-separated values) или реляционная база данных [Codd 1970]. Его преимущество — простота редактирования и обработки, поэтому он популярен в типологических и сравнительно-лексикографических проектах, в частности, на базе Cross-Linguistic Linked Data (CLLD) и Cross-Linguistic Data Formats (CLDF) [Forkel et al. 2018]. Для наших задач табличный формат является недостаточно гибким: например, в его рамках никак нельзя передать внутреннюю структуру этимологической части статьи.

Вместо этого мы приняли решение использовать язык разметки TEI XML [The TEI Consortium 2023]. Язык Text Encoding Initiative (TEI) представляет собой набор XML-схем для представления различных типов текстовых данных: книг, словарей, рукописей, лингвистических корпусов; в частности, он используется Британским национальным корпусом (BNC, <http://www.natcorp.ox.ac.uk/>) [Burnard, Aston 1998]. Для словарей, помимо базовой схемы TEI, существует также набор стандартных практик TEI Lex-0 [Romary, Tasovac 2018]. Архитектура словаря и процесс работы описаны в [Belyaev et al. 2021]; исходные файлы самого словаря и используемой в нем XML-схемы доступны по адресу <https://github.com/abaevdict>. Начало этимологии приведенной в разделе 2 статьи *zærin* может служить неплохим примером:

```
<etym xml:lang="ru">>Вместе с железом (<ref type="xr" target="»#entry_7fsæn"><w>fsæn</w></ref>) и сталью (<ref type="xr" target="»#entry_ændon">)> относится к древнейшим названиям металлов в осетинском. Восходит к <mentioned xml:lang="ira"><lang/> <w type="rec">zaranya-</w></mentioned>. Ср. <mentioned xml:lang="fa"><lang/> <w>zar(r)</w> <note type="comment">(из <w type="rec">zarna-</w></note> <gloss><q>золото</q></gloss></mentioned>, <mentioned xml:lang="fa"><w>zarrin</w> <gloss><q>золотой</q></gloss></mentioned>, <mentioned xml:lang="fa"><lang/> <w type="rec">zarn</w></mentioned> в <mentioned xml:lang="xcl"><lang/> <w>zar navuxt'</w> <gloss><q>златотканый</q></gloss> <note type="bibl"><bibl><ref type="bibl" target="»ref_Hübschmann._PSt."/>
```

(...)

Для непосредственного ввода словарных данных мы разработали набор CSS-стилей для отображения словарного материала, максимально приближенного к исходному форматированию, а также дополнение к популярному редактору Oxygen XML Editor (<https://www.oxygenxml.com/>), позволяющее редактировать текст в формате,

близком к печатному варианту. Форматирование, таким образом, динамически подбирается в соответствии с ролью и позицией элемента. То, как указанная статья выглядит в интерфейсе редактора с применением CSS-стилей, представлено на рис. 2.

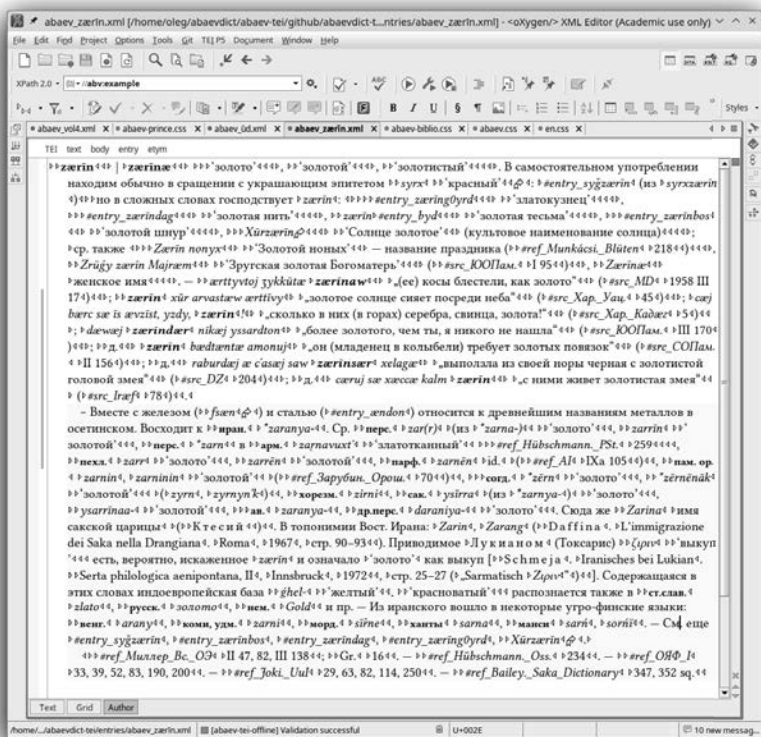


Рис. 2. Вид статьи *zarin* ‘золото’ в редакторе Охуген XML

4. XML-разметка и публикация

Сочетание XML и CSS, используемое для редактирования словаря, может практически без изменений применяться для создания его печатной версии. В нашем проекте мы использовали систему CSS-верстки Prince (<https://www.princexml.com/>). В качестве примера приведем рассматривавшуюся выше словарную статью в сверстанном виде (рис. 3) и ту же статью в английском переводе (с автоматически подставленными названиями языков и ссылками на источники) на рис. 4. Как видно, автоматическое форматирование и подстановка повторяющихся названий значительно облегчает издание словаря и снижает процент ошибок.

zærin | **zærinæ** 'золото', 'золотой', 'золотистый'. В самостоятельном употреблении находим обычно в сращении с украшающим эпитетом *syrx* 'красный': *sygzærin* (из *syrxzærin*)но в сложных словах господствует *zærin*: *zæring yrd* 'златокузнец', *zæringdag* 'золотая нить', *zærinbyd* 'золотая тесьма', *zærinbos* 'золотой шнур', *Xürzærin* 'Солнце золотое' (культовое наименование солнца); ср. также *Zærin nonyx* 'Золотой ноных' — название праздника (MUNKÁCSI. Blüten 218), *Zrúgy zærin Majræm* 'Зругская золотая Богоматерь' (ЮОПам. I 95), *Zærinæ* женское имя. — *ærttyvtjy zykktútæ zærinaw* '(ее) косы блестящи, как золото' (MD 1958 III 17); *zærin xür arvastæw ærttyvy* 'золотое солнце сияет посреди неба' (Хар. Уац. 45); *cæj bærc sæ is ævzist, yzdy, zærin!* 'сколько в них (в горах) серебра, свинца, золота!' (Хар. Кадæг 5); *dæwæj zærinðær nikæj yssardton* 'более золотого, чем ты, я никого не нашла' (ЮОПам. III 170); д. *zærin bædtæntæ amonuj* 'он (младенец в колыбели) требует золотых повязок' (СОПам. II 156); д. *raburdæj æ c'asæj saw zærinæsær xelagæ* 'выползла из своей норы черная с золотистой головой змея' (DZ 204); д. *cæruj sæ hæccæ kalm zærin* 'с ними живет золотистая змея' (Iraef 78).

— Вместе с железом (*fšæn*) и сталью (*ændon*) относится к древнейшим названиям металлов в осетинском. Восходит к иран. **zaranya-*. Ср. перс. *zar(r)* (из **zar-na-*) 'золото', *zarrin* 'золотой', перс. **zarn* в арм. *zarnavuxt* 'златотканый' HÜBSCHMANN. PSt. 259, пехл. *zarr* 'золото', *zarrēn* 'золотой', парф. *zarnēn* id. (AI IXa 105), нам. оп. *zarnin*, *zarninin* 'золотой' (Зарубин. Орoш. 70), согд. **zēr̄n* 'золото', **zēr̄nēnāk* 'золотой' (*zyrn*, *zyrnyrk*), хорезм. *zirnī*, сак. *ysirra* (из **zarnya-*) 'золото', *ysarrinaa-* 'золотой', ав. *zaranya-*, др.перс. *daraniya-* 'золото'. Сюда же *Zarina* имя сакской царицы (Ктесий). В топонимии Вост. Ирана: *Zarin*, *Zarang* (DAFFINA. L'immigrazione dei Saka nella Drangiana. Roma, 1967, стр. 90–93). Приводимое Лукианом (Токсарис) *ζῖπις* 'выкуп' есть, вероятно, искаженное *zærin* и означало 'золото' как выкуп [SCHMEJA. *Iranisches bei Lukian. Serta philologica aenipontana*, II, Innsbruck, 1972, стр. 25–27 ('Sarmatisch Zῖπις')]. Содержащаяся в этих словах индоевропейская база *ghel-* 'желтый', 'красноватый' распознается также в ст.слав. *zlato*, русск. *золото*, нем. *Gold* и пр. — Из иранского вошло в некоторые угро-финские языки: венг. *arany*, коми, удм. *zarni*, морд. *siřne*, ханты *sarna*, манси *sarni*, *sorñi*. — См. еще *sygzærin*, *zærinbos*, *zæringdag*, *zæring yrd*, *Xürzærin*.

Миллер Вc. ОЭ II 47, 82, III 138; GR. 16. — ОЯФ I 33, 39, 52, 83, 190, 200. — JOKI. Uul 29, 63, 82, 114, 250. — BAILEY. Saka Dictionary 347, 352 sq.

Рис. 3. Печатная версия статьи *zærin* (русский язык)

zærin | **zærinæ** 'gold', 'golden', 'aureate'. In independent use, it is usually found with the epithet *syrx* 'red': *sygzærin* (from *syrxzærin*)but *zærin* dominates in compound words: *zæring yrd* 'goldsmith', *zæringdag* 'golden thread', *zærinbyd* 'gold lace', *zærinbos* 'golden cord', *Xürzærin* 'golden sun' (cult name of the sun); cf. also *Zærin nonyx* 'gold nonyx' – holiday name (MUNKÁCSI. Blüten 218), *Zrúgy zærin Majræm* 'Zrug golden mother of God' (SOSSrc. I 95), *Zærinæ* female name. — *ærttyvtjy zykktútæ zærinaw* '(her) braids were shining like gold' (MD 1958 III 17); *zærin xür arvastæw ærttyvy* 'the golden sun is shining in the middle of the sky' (Хар. Уац. 45); *cæj bærc sæ is ævzist, yzdy, zærin!* 'how much silver, lead, gold are in them (in the mountains)!' (Хар. Кадæг 5); *dæwæj zærinðær nikæj yssardton* 'I have not found anyone more golden than you' (SOSSrc. III 170); *Digor zærin bædtæntæ amonuj* 'he (the baby in the cradle) demands gold bandages' (NOSSrc. II 156); *Digor raburdæj æ c'asæj saw zærinæsær xelagæ* 'a black snake with a golden head crawled out of its hole' (DZ 204); *Digor cæruj sæ hæccæ kalm zærin* 'a golden snake lives with them' (Iraef 78).

— Together with iron (*fšæn*) and steel (*ændon*), belongs to the most ancient names of metals in Ossetic. It goes back to Iranian **zaranya-*. Cf. Persian *zar(r)* (from **zar-na-*) 'gold', *zarrin* 'golden', Persian **zarn* in Armenian *zarnavuxt* 'golden-woven' HÜBSCHMANN. PSt. 259, Middle Persian *zarr* 'gold', *zarrēn* 'golden', Parthian *zarnēn* id. (AI IXa 105), Oroshi *zarnin*, *zarninin* 'golden' (ZARUBIN. Orosh. 70), Sogdian **zēr̄n* 'gold', **zēr̄nēnāk* 'golden' (*zyrn*, *zyrnyrk*), Khwarezmian *zirnī*, Saka *ysirra* (from **zarnya-*) 'gold', *ysarrinaa-* 'golden', Avestan *zaranya-*, Old Persian *daraniya-* 'gold'. See here also *Zarina* name of the Saka queen (CTESIAs). In the toponymy of Eastern Iran: *Zarin*, *Zarang* (DAFFINA. L'immigrazione dei Saka nella Drangiana. Roma, 1967, p. 90–93). Given by LUCIAN (Toxaris) *ζῖπις* 'ransom' is probably a distorted *zærin* and it meant 'gold' as ransom [SCHMEJA. *Iranisches bei Lukian. Serta philologica aenipontana*, II, Innsbruck, 1972, p. 25–27 ('Sarmatisch Zῖπις')]. The Indo-European base *ghel-* 'yellow', 'reddish' contained in these words is also recognized in Old Church Slavonic *zlato*, Russian *zoloto*, German *Gold* etc. — From Iranian it entered some Finno-Ugric languages: Hungarian *arany*, Komi, Udmurt *zarni*, Mordvinic *siřne*, Khanty *sarna*, Mansi *sarni*, *sorñi*. — See also *sygzærin*, *zærinbos*, *zæringdag*, *zæring yrd*, *Xürzærin*.

MILLER Ws. ОЭ II 47, 82, III 138; GR. 16. — ОЯФ I 33, 39, 52, 83, 190, 200. — JOKI. Uul 29, 63, 82, 114, 250. — BAILEY. Saka Dictionary 347, 352 sq.

Рис. 4. Печатная версия статьи *zærin* (английский язык)

5. Онлайн-интерфейс к базе данных

Одной из задач настоящего проекта было создание электронного веб-интерфейса к словарю. Для этого в нашем проекте на базе XML-кода автоматически генерируется реляционная база данных формата SQLite, совместимая с системой OnLex [Makarov 2023], которая уже успешно применялась для создания электронного словаря памирских языков [Makarov et al. 2022]. Так, в таблицу Units входят все леммы словаря (как осетинские — включая дигорские варианты и дериваты, — так и все иноязычные слова, встретившиеся в этимологиях); она содержит поля для идентификатора леммы, ее основной формы, исходного XML-кода статьи; в таблицы Examples и Meanings входят примеры и значения соответственно, и все три таблицы связаны ссылками через идентификаторы. Такой формат в дальнейшем может быть достаточно легко конвертирован в CSV-формат (например, CLDF) для последующей обработки.

Пилотный вариант основанного на такой базе данных веб-интерфейса доступен онлайн по адресу <https://ossetic.iranica.space/dict>. В нем возможен поиск по осетинским формам, примерам и переводам. Каждая лемма представлена в виде трех вкладок: «Статья» (структурированное представление тех данных, которые непосредственно представлены в SQLite-базе, рис. 5), «Оригинальный текст» (исходный текст статьи, с возможностью перехода по гиперссылкам на перекрестные ссылки, другие языки и их слова, рис. 6) и «Карта» (карта языков, слова из которых цитируются в этимологии,

Статья Оригинальный текст Карта

СЛОВО mad Матер и./Иом

См. также
◆ [madæ \(дигорское\)](#)

Значение	мать
	<ul style="list-style-type: none">○ <i>ysææ læppijæn... sæ mad æmæ sæ fyd zægtoj; Ysææ kyzy iw mad æmæ fydej ærkeniut, kærnod næ bæfidawlysty</i> '(своим) двадцати сыновьям сказали их мать и отец: «Приведите (в жены) двадцать девушек от одной матери и отца, иначе они не поладят»'○ <i>dy tyn nujææg madoj adyndææ</i> 'ты мне милее родной матери'○ <i>xir... oly cæææg idgojmaŷ madaw rævdylla</i> 'солнце ласкало, как мать, всякое живое существо'○ <i>næ zæonp tæe mady rævdyd</i> 'не знаю материнской ласки'○ <i>(wærykkytæn) sæ madæly dææ ærken, kærnod wæzysu</i> 'приведи агнцам их маток, не то они будут блеять'○ <i>mæxi mady zææægŷ k uzæm dæ wæzyp</i> 'я люблю тебя, как родных братьев и сестер'○ <i>noŷdææ mæ dyp dæ sæææn kad ysæægysætæ mady bæææg</i> 'кроме того, мы почтим тебя (подарком) «конем матери»'
Устойчивые сочетания и идиомы	<p>mady zææg родные братья и сестры</p> <p>mady bæx конь, которого жених дарит матери невесты не в счет калыма</p>

Рис. 5. Структурированный вариант представления статьи

mad | **madæ** 'мать', мн. ч. *madætiæ* | *maddætiæ*, *mady zænæg* | *madi zænæg*, *madi gubun* 'родные братья и сестры', *mady bæx* 'конь, которого жених дарит матери невесты не в счет калыма', — *ysæz læppijæn...* sæ *mad æmæ* sæ *fyd ægitoj*; 'Ysæz Kyzgy iw *mad æmæ fydej ærkænūt, kænōd næ bafidawæzlyt*...' (своим) двадцати сыновьям сказали их мать и отец: «Приведите (в жены) двадцать девушек от одной матери и отца, иначе они не поладят» (ЮОПам. III 56); *mady mardej stjær* x,yz, *fyd mardej margyr* x,yz 'от смерти матери — сиротский вид, от смерти отца — бедный вид' (Емб. 19); *dy myn nyjjaræg madæj adgyndær* 'ты мне милее родной матери' (С е к а 51); *xiæ...* *aly cæræg idgojmaŷ madaw rævdýda* 'солнце ласкало, как мать, всякое живое существо' (С е к а 121); *næ zonyl næ mady rævdýd* 'не знаю материнской ласки' (К о с т а 31); д. *avd læqwaeni æma ewnæg kizgi madæ fæwwo* 'стань матерью семи мальчиков и одной девочки' (MD 1949 II 37); (wærýkkytaæn) sæ *madelty dær ærkaen, kænōd woszysty* 'приведи ягнтятам их маток, не то они будут блеять' (С е к а 60); д. *wæd ku nimboxuj, ku nyzinazuj i madi zæronđ* 'тогда как зарывает, запричитает старая мать' (Iraef 89); д. *Asægo...* *ranæxtæræg æ madæ ma æ fidi agorunmæ* 'Асаго отправился искать свою мать и отца' (MSt. 15); *mæxi mady zænægý* x,yzæn *dæ warzyl* 'я люблю тебя, как родных братьев и сестер' (F. 1928 IV 7); д. *warzon madi gubun basabur kodta æ kæwungæles* 'любимая сестра приглушила свой плач' (MD 1937 I 35); *nogydær ma dyn dæ særæn kad ysænzysytæm mady bæxæj* 'кроме того, мы почитим тебя (подарком) «конем матери»' (С е к а 126)

- Общеиндоевропейский термин родства. Восходит к иран. **matar-* (имен. падеж **mātā*), и.е. *māter-*. Форма *madæ* отражает старый имен. падеж **mātā*: основа мн. ч. *madel-* восходит к **matar-*. Ср. перс. *mādar* 'мать', *mādā* 'самка', исл. *matar*, *māt* 'мать', *mātak* 'самка', ирл. (Healing. BBV 112), бел. *māt*, *mās*, *māl*, вурд. *māk* (из **mādā*), тал. *mo*, афг. *mōr* (= **mātr*), нам. ш. *mād*, нам. руш. *mōd*, мн. ч. *mōdēj*, согд. **māt-* (*mth-*, *mʔ-*), парф. *mād*, хорезм. *mād*, сак. *māta-*, ав. *mātar-* 'мать', др.перс. *ha-mātā* 'происходящий от одной матери', др.инд. *mātār-*, арм. *matr*, ст.слав. *мати*, мн.ч. *матере*, русск. *мотё* 'жена', латыш. *māte*, англос. *mōdor*, нем. *Mutter*, гр. *μήτηρ*, *μήτηρ*, лат. *mater*, др.ирл. *māthir*, тохар. А *mācar*, тохар. В *mācer* 'мать'. — См. *madel*, *madard* (*hærgnæg*). Ср. *fyd* 'отец'.

Рис. 6. Исходный текст статьи в онлайн-базе



Рис. 7. Карта цитируемых слов в онлайн-версии словаря

рис. 7). Таким образом, не теряя ничего из оригинальных данных словаря (включая авторский порядок элементов и форматирование), онлайн-интерфейс включает все преимущества современных словарных проектов, а интерактивная карта на примере визуализации иллюстрирует богатые возможности работы с электронными лексическими данными.

6. Заключение

В настоящей статье были представлены основные результаты четырехлетнего проекта по созданию электронной двуязычной версии ИЭСОЯ В.И. Абаева. Описан используемый в нашей работе формат на базе TEI XML и система его редактирования в Охуген

XML; обоснован выбор этого инструментария для такого типа словарей со сложной структурой. На примере печатной версии словаря и онлайн-базы продемонстрированы возможности дальнейшей обработки XML-данных. Представляется, что наш опыт может быть полезен как для других подобных проектов цифровизации печатных словарей, так и для проектов по созданию новых лексикографических ресурсов.

СПИСОК ЛИТЕРАТУРЫ

1. *Абаев В.И.* Историко-этимологический словарь осетинского языка. Т. I: А–К'. М.–Л., 1958. Т. II: L–R. Л., 1973. Т. III: S–T'. Л., 1979. Т. IV: U–Z. Л., 1989. Указатель (сост. Е.Н. Сченснович, А.В. Лушникова, Л.Р. Додыхудоева). М., 1995.
2. *Кузнецова А.И., Сердобольская Н.В., Усачёва М.Н., Бирюк О.Л., Идрисов Р.И.* (редакторы-составители). Словарь бесермянского диалекта удмуртского языка. М., 2013.
3. *Belyaev O., Khomchenkova I., Sinityna J., Dyachkov V.* Digitizing print dictionaries using TEI: The Abaev Dictionary Project // IWCLUL 2021: The Seventh International Workshop on Computational Linguistics of Uralic Languages. Proceedings of the Workshop. Syktyvkar, 2021.
4. *Burnard L., Aston G.* The BNC handbook: exploring the British National Corpus. Edinburgh, 1998.
5. *Codd E.F.* A relational model of data for large shared data banks // Communications of the ACM. 1970. № 13 (6). P. 377–387.
6. *Forkel R., List J.-M., Greenhill S.J., Rzymiski C., Bank S., Cysouw M., Hammarström H., Haspelmath M., Kaiping G.A., Gray R.D.* Cross-Linguistic Data Formats, advancing data sharing and re-use in comparative linguistics // Scientific Data. 2018. № 5, 180205.
7. *Makarov Y., Melenchenko M., Novokshanov D.* Digital Resources for the Shughni Language // Proceedings of The Workshop on Resources and Technologies for Indigenous, Endangered and Lesser-Resourced Languages in Eurasia within the 13th Language Resources and Evaluation Conference. Marseille, 2022. P. 61–64.
8. *Makarov Y.* OnLex: a platform for online lexicography. Рукопись. М., 2023.
9. *Romary L., Tasovac T.* TEI Lex-0: A Target Format for TEI-Encoded Dictionaries and Lexical Resources // Miyazaki T. (ed.). The 18th Annual TEI Conference and Members' Meeting: Book of Abstracts. Tokyo, 2018. P. 274–275.
10. *The TEI Consortium.* TEI P5: Guidelines for Electronic Text Encoding and Interchange. 2023. URL: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html> (дата обращения: 5 декабря 2023 г.)

REFERENCES

1. *Abaev V.I.* *Istoriko-jetimologicheskij slovar' osetinskogo jazyka* [A historical-etymological dictionary of Ossetic]. Vol. I: A–K'. Moscow and Leningrad, Nauka Publ., 1958. 655 p. Vol. II: L–R. Leningrad, Nauka Publ., 1973. 448 p. Vol. III: S–T'. Leningrad, Nauka Publ., 358 p. 1979. Vol. IV: U–Z. Leningrad, Nauka, 1989. 325 p. *Ukazatel'* [Index] (compiled by E. N. Schensnovich, A. V. Lushnikova, L. R. Dodyhudoeva). Moscow, Nauka, 1995. 447 p.

2. Kuznecova A. I., Serdobol'skaja N. V., Usachjova M. N., Birjuk O. L., Idrisov R. I. (eds.). *Slovar' besermjanskogo dialekta udmurtskogo jazyka* [A dictionary of the Besermen dialect of Udmurt]. Moscow, Tezaurus Publ., 2013. 257 p.
3. Belyaev O., Khomchenkova I., Sinitsyna J., Dyachkov V. Digitizing print dictionaries using TEI: The Abaev Dictionary Project. In: *IWCLUL 2021: The Seventh International Workshop on Computational Linguistics of Uralic Languages. Proceedings of the Workshop*. Syktyvkar, Association for Computational Linguistics, 2021.
4. Burnard L., Aston G. *The BNC handbook: exploring the British National Corpus*. Edinburgh, Edinburgh University Press, 1998. 256 p.
5. Codd E.F. A relational model of data for large shared data banks. *Communications of the ACM*, 1970, no. 13 (6), pp. 377–387.
6. Forkel R., List J.-M., Greenhill S. J., Rzymiski C., Bank S., Cysouw M., Hammarström H., Haspelmath M., Kaiping G. A., Gray R. D. Cross-Linguistic Data Formats, advancing data sharing and re-use in comparative linguistics. *Scientific Data*, 2018, № 5, 180205. <https://doi.org/10.1038/sdata.2018.205>
7. Makarov Y., Melenchenko M., Novokshanov D. Digital Resources for the Shughni Language. In: *Proceedings of The Workshop on Resources and Technologies for Indigenous, Endangered and Lesser-Resourced Languages in Eurasia within the 13th Language Resources and Evaluation Conference*. Marseille, European Language Resources Association, 2022. P. 61–64.
8. Makarov Y. *OnLex: a platform for online lexicography*. Manuscript. Moscow, 2023.
9. Romary L., Tasovac T. TEI Lex-0: A Target Format for TEI-Encoded Dictionaries and Lexical Resources. In: Miyazaki T. (ed.). *The 18th Annual TEI Conference and Members' Meeting: Book of Abstracts*. Tokyo, 2018. P. 274–275.
10. The TEI Consortium. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. 2023. URL: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html> (accessed 5 December 2023)

Поступила в редакцию 12.08.2023

Принята к публикации 13.02.2024

Отредактирована 09.03.2024

Received 12.08.2023

Accepted 13.02.2024

Revised 09.03.2024

ОБ АВТОРАХ

Беляев Олег Игоревич — кандидат филологических наук, доцент кафедры теоретической и прикладной лингвистики филологического факультета МГУ имени М.В. Ломоносова, старший научный сотрудник сектора иранских языков Института языкознания РАН; belyaev@ossetic-studies.org

Хомченкова Ирина Андреевна — преподаватель кафедры теоретической и прикладной лингвистики филологического факультета МГУ имени М.В. Ломоносова, младший научный сотрудник сектора иранских языков Института языкознания РАН, научный сотрудник отдела корпусной лингвистики и лингвистической поэтики Института русского языка имени В.В. Виноградова; irina.khomchenkova@yandex.ru

Синицына Юлия Вячеславовна — преподаватель кафедры теоретической и прикладной лингвистики филологического факультета МГУ имени М.В. Ломоносова, младший научный сотрудник сектора иранских языков Института языкознания РАН; jv.sinitsyna@gmail.com

Дьячков Вадим Викторович — кандидат филологических наук, старший научный сотрудник отдела африканских языков Института языкознания РАН; vadim.dyachkov@iling-ran.ru

Бызова Анна Александровна — студент бакалавриата отделения теоретической и прикладной лингвистики филологического факультета МГУ имени М.В. Ломоносова, лаборант-исследователь лаборатории исследования и сохранения малых языков Института языкознания РАН; byzova@iling-ran.ru

Бадеев Артём Олегович — лаборант-исследователь сектора иранских языков Института языкознания РАН, студент магистратуры Школы лингвистики НИУ ВШЭ; badeev@iling-ran.ru

Алексеев Данил Андреевич — лаборант-исследователь сектора иранских языков Института языкознания РАН, студент бакалавриата отделения теоретической и прикладной лингвистики филологического факультета МГУ имени М.В. Ломоносова; dnl.alksv@gmail.com

Макаров Юрий Юрьевич — младший научный сотрудник сектора типологии Института языкознания РАН, научный сотрудник отдела корпусной лингвистики и лингвистической поэтики Института русского языка имени В.В. Виноградова, студент магистратуры Кембриджского университета; yurmak@iling-ran.ru

ABOUT THE AUTHORS

Oleg Belyaev — PhD, Associate Professor, Department of Theoretical and Applied Linguistics, Faculty of Philology, Lomonosov Moscow State University; Senior Researcher, Section of Iranian Languages, Institute of Linguistics RAS; belyaev@ossetic-studies.org

Irina Khomchenkova — Teaching Fellow, Department of Theoretical and Applied Linguistics, Faculty of Philology, Lomonosov Moscow State University; Junior Researcher, Department of Iranian Languages, Institute of Linguistics RAS; Research Fellow, Department of Corpus Linguistics and Linguistic Poetics, Vinogradov Russian Language Institute RAS; irina.khomchenkova@yandex.ru

Julia SinitSYna — Teaching Fellow, Department of Theoretical and Applied Linguistics, Faculty of Philology, Lomonosov Moscow State University; Junior Researcher, Department of Iranian Languages, Institute of Linguistics RAS; jv.sinitSYna@gmail.com

Vadim Dyachkov — PhD, Senior Researcher, Department of African Languages, Institute of Linguistics RAS; vadim.dyachkov@iling-ran.ru

Anna Byzova — BA Student, Department of Theoretical and Applied Linguistics, Faculty of Philology, Lomonosov Moscow State University; Assistant, Department of Iranian Languages, Institute of Linguistics RAS; byzova@iling-ran.ru

Artyom Badeev — Assistant, Section of Iranian Languages, Institute of Linguistics RAS; MA Student, School of Linguistics, HSE University; badeev@iling-ran.ru

Danil Alekseev — Assistant, Section of Iranian Languages, Institute of Linguistics RAS; BA Student, Department of Theoretical and Applied Linguistics, Faculty of Philology, Lomonosov Moscow State University; dnl.alksv@gmail.com

Yury Makarov — Junior Researcher, Section of Typology, Institute of Linguistics RAS; Research Fellow, Department of Corpus Linguistics and Linguistic Poetics, Vinogradov Russian Language Institute RAS; MPhil Student, University of Cambridge; yurmak@iling-ran.ru