**Правительство Российской Федерации**

**Федеральное государственное автономное образовательное учреждение высшего образования**

**Национальный исследовательский университет «Высшая школа экономики»**

**Факультет гуманитарных наук**

**Образовательная программа**

**«Компьютерные методы в гуманитарных науках»**

**КУРСОВАЯ РАБОТА**

На тему \_\_\_Электронное представление славянских словарей 17 века\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*Название темы на английском*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_Digital representation of the Slavic dictionaries of the XVII century\_\_\_\_\_

Студентка\_I\_\_курса

группы №\_МЦМГН201\_\_\_\_

Левченко Мария Владимировна

(Ф.И.О.)

Научный руководитель

Ляшевская Ольга Николаевна

(Ф.И.О)

\_\_\_канд. филол. наук \_\_\_

(должность, звание)

Москва, 2021 г.

# Содержание:

1. [Введение](#введение)
2. [Епифаний Славинецкий и его «Лексикон греко-славяно-латинский»](#Епифаний)
3. [Электронное представление словаря Епифания](#электро)
   * [Кодирование](#кодирование)
   * [Публикация на веб-сервисе](#публикац)
4. [Заключение](#заклю)

**Введение**

Лексикография, раздел лингвистики, направленный на изучение и составление словарей, в последние годы получила новый толчок к развитию с изобретением автоматизированных, компьютерных методов обработки данных. Помимо стандартных методов сохранения наследия (начиная с простого сканирования) исследователей интересуют процессы конвертации физического объекта в цифровой, машиночитаемый формат (retro-digitization). Строгая структура таких форматов позволяет не только ускорить чтение и навигацию (в нашем случае – по словарю), но и открывает новые возможности для анализа объекта, не доступные при близком чтении.

Наша работа будет попыткой привести в машиночитаемый формат раздел Славяно-греко-латинского словаря, составленного монахом Чудовского монастыря и переводчиком Епифанием Славинецким во второй половине XVII века в Москве.

Работа будет проводиться на материалах, собранных германо-российским проектом «Греко-славяно-латинский словарь Епифания Славинецкого в восточнославянском и европейском языковом пространстве XVI – начала XVIII в.» (руководители проекта - И. А. Подтергера[[1]](#footnote-1), М. О. Новак[[2]](#footnote-2)). Этот проект направлен на изучение культурных и языковых связей в восточной Европе раннего Нового времени на материале ранее не изученных и не изданных текстах так называемого Чудовского круга, оказавшего значительное влияние на развитие русского языка во второй половине XVII века, и особенное внимание уделяется трудам Епифания Славинецкого, одного из ведущих московских книжников. Проект также ставит целью изучить взаимное влияние греческой, латинской, славянской и польской языковых традиций на уровне лексики, ведь восточнославянская лексикография раннего Нового времени и сегодня остается малоизученной областью языковой культуры допетровского времени. Вводя эти тексты в научный оборот, исследователи надеются окончательно опровергнуть гипотезу о «грекофильстве» чудовского круга и раскрыть роль латинских и польских источников.

В рамках проекта планируется создать базу данных, которая включает в себя электронные издания изучаемых текстов, в том числе и трехъязычного словаря Епифания Славинецкого, с XML разметкой по стандарту TEI[[3]](#footnote-3). Для оцифровки рукописей используется программа автоматического распознавания текста Transkribus[[4]](#footnote-4). В дальнейшем этот словарь планируется соединить с другими словарями на уровне гипер/металемм в рамках модели RDF, возможна также имплементация в исторический подкорпус НКРЯ[[5]](#footnote-5). Полученная в результате проекта база данных позволит исследовать тенденции развития письменного церковнославянского языка в контексте взаимодействия с другими европейскими языками, в том числе проследить трансформацию языка от рутенской (украинско-белорусской) и церковно-славянской лексики к русской.

Одним из первых этапов проекта является введение в научный оборот словаря Епифания Славинецкого, и наша курсовая работа пытается выполнить часть задач этого этапа – представление данных в XML формате по стандарту TEI.

Итак, глобальная цель нашей работы совпадает с целью проекта: ввести в научный обиход Славяно-греко-латинский словарь Епифания Славинецкого. Конкретная же цель состоит в том, чтобы попробовать это на небольшом объеме данных – около 600 словарных статей на букву β в формате Excel-таблицы.

Для этого потребуется выполнить следующие задачи:

1. Разработать TEI XML разметку
2. Разработать код для автоматизации процесса разметки
3. Найти способ опубликовать полученный размеченный XML файл на веб-сервисе для тестирования модели

Главное руководство по TEI разметке предоставлено на сайте проекта Text Encoding Initiative. Нам особенно интересен раздел Dictionaries[[6]](#footnote-6), посвященный тэгам и атрибутам, специфичным для словаря. Большую помощь оказало обращение к русскоязычному учебному пособию[[7]](#footnote-7) В. П. Захарова, в котором приведен подробный анализ рекомендаций проекта TEI и перевод на русский язык вышеупомянутого раздела Dictionaries.

Также мы использовали спецификации TEI Lex-0, расширяющие и конкретизирующие некоторые рекомендации основного руководства. Цель проекта – «establishing a baseline encoding and a target format to facilitate the interoperability of heterogeneously encoded lexical resources»[[8]](#footnote-8). Авторы предлагают использовать TEI Lex-0 в новых проектах для облегчения последующей работы с несколькими ресурсами.

Разработка разметки для трехъязычного словаря может немного отличаться от двуязычного. Обратившись к опыту хорватских коллег[[9]](#footnote-9), мы поняли, что наш подход требует отличной схемы. Нам хотелось бы представить менее плоскую и более разветвленную структуру, чтобы в дальнейшем размеченный словарь можно было связать с другими словарями через модель RDF: чтобы перевод на латинский входил в тэг <sense> для славянского переводы, который в свою очередь входит в тэг <sense> греческого слова.

В мировом научном сообществе уже есть опыт оцифровки словарей. Основное отличие других проектов от нашей работы состоит в использовании методов, которые через методы text recognition и machine learning распознают текст на отсканированной странице и напрямую кодируют текст в XML.

Например, так устроена система GROBID-Dictionaries[[10]](#footnote-10) (Mohamed Khemakhem, Laurent Romary), которая на вход принимает PDF файл и строит TEI XML файл с использованием cascading CRF models[[11]](#footnote-11). Так, например, с помощью этой системы был закодирован[[12]](#footnote-12) словарь Миштекского языка с переводами на кастильский конца XVI века. Источник PDF был издан в 2009 году, что облегчило распознавание текста; в нашем случае этот этап был бы несколько осложнен, но, по нашему мнению, было бы интересно попробовать применить этот метод и на рукописи Епифания.

Список реализованных проектов по публикации словарей постоянно растет. Так, проект Lingenio[[13]](#footnote-13) является как структурой базы данных, так и интерфейсной системой, позволяющей создавать иерархическую структуризацию лексических записей для различных исследовательских целей. Именно в Lingenio проектная группа планирует ввести размеченные словари.

Изображение выглядит как текст

Автоматически созданное описаниеРесурс Early Irish Glossaries Database[[14]](#footnote-14) интересен тем, что он предоставляет возможность открыть исходные TEI XML файлы. Они отличаются крайне простой структурой, и навигация по вхождению практически невозможна. Ресурс предоставляет возможность поиска нужного слова по клику в eDIL – электронном словаре ирландского языка.

Особый интерес представляет проект NTeALan[[15]](#footnote-15), направленный на изучение, сохранение и развитие африканских языков посредством оцифровки текстов на этих языках, в частности – словарей. Участниками проекта был разработан электронный ресурс[[16]](#footnote-16), на котором есть возможность поиска и перевода слов с различных африканских языков на французский/английский и обратно. Проект основан на коллаборации: можно записать свое произношение слова или оставить комментарий.

Одним из наиболее близких нам проектов можно считать ресурс Расковник[[17]](#footnote-17), сербский лексикографический портал и платформу сравнительных исследований сербской лексики. На сайте можно вести поиск не только по леммам, но и по грамматическим формам и примерам. Поиск ведется по пяти словарям, при этом можно указать один словарь, можно выбрать сразу все, можно открыть конкретный словарь и искать по указателю. В словарной статье есть грамматическая информация и ссылки на другие вхождения. Основа – база данных eXist.

Из российских проектов хочется выделить портал Церковнославянский язык сегодня[[18]](#footnote-18). Это не только словарь, но и образовательный ресурс с видеолекциями и полезными ссылками. Словарная статья содержит в себе перевод на современный русский язык, примеры из богослужебных текстов, иногда также включены и греческие эквиваленты, однако искать можно только церковнославянские лексемы. Из каждой статьи идет прямая ссылка в корпус НКРЯ, в котором можно посмотреть примеры употребления. Опыт этого ресурса по привязке к НКРЯ может пригодиться в проекте по публикации текстов Чудовского круга.

**Епифаний Славинецкий и его** **«Лексикон греко-славяно-латинский»**

Как было сказано выше, Епифаний являлся монахом Чудовского монастыря, филологом, переводчиком, проповедником и одной из крупнейших фигур, повлиявших на развитие русского языка Нового времени. Постриженик Киево-Печерской лавры, он в 1649 году был приглашен царем Алексеем Михайловичем в Москву для участия в книжных реформах, подготовки нового издания Библии и организации школы для обучения детей греческому языку[[19]](#footnote-19).

Переводческая и филологическая деятельность Епифания чрезвычайно обширна, и многие его работы до сих пор не изданы и не изучены. Он занимался переводом латинских лемм известного в Западной Европе толкового словаря латинского языка авторства Амвросия Калепина, а Арсений Корецкий-Сатановский обратил перевод в славяно-латинскую версию, тем самым введя славянский язык в контекст многочисленных переводов Калепина на европейские языки и положив начало обширному восточнославянскому корпусу Калепина в рутенской и московской традиции. Авторству Епифания также принадлежит так называемый «Филологический лексикон» - объяснения терминов Священного Писания на греческом языке с латинскими соответствиями и вписанными позднее кириллическими[[20]](#footnote-20).

«Лексикон греко-славяно-латинский» был написан, по предположениям исследователей, в 1660е-1680е годы (между 1664 и 1680 годами, как считает С. Н. Брайловский[[21]](#footnote-21)) и включает от 75 до 85 тысяч греческих лемм с их переводами на церковнославянский и латинский языки[[22]](#footnote-22). Автограф рукописи находится в Государственном историческом музее[[23]](#footnote-23). Словарь служил посредником между наследием западной, латинской Европы и грекоориентированной церковнославянской письменной традицией Московского государства и своей целью ставил обеспечение доступа восточнославянских читателей к пониманию греческих и латинских текстов[[24]](#footnote-24).

Словарь Епифания входит в предшествующую ему лексикографическую традицию, учитывая славянскую лексику других словарей. Как утверждает Т.В. Пентковская, «В переводах Максима Грека и особенно Епифания Славинецкого и Евфимия Чудовского действует принцип, характерный для построения словарной статьи в лексиконах: привести как можно больше лексических эквивалентов для одного понятия,»[[25]](#footnote-25) - большой объем словаря может означать обращение Епифания к множеству как славянских, так и латинско-польских источников.

При составлении перевода Калепина Епифаний, знавший польский, пользовался и польско-латинскими источниками, например, польско-латинско-греческим тезаурусом Григория Кнапского – гипотеза О. Б. Страховой о том, что Епифаний пользовался им и при составлении словаря требует своего исследования[[26]](#footnote-26), ведь известно, что польско-латинские источники активно использовались монахами Чудовского круга[[27]](#footnote-27).

На словарь Епифания опирались его последователи, в частности Федор Поликарпов чей словник в свою очередь послужил источником словарей начала XVIII века, в том числе русско-французского словаря Антиоха Кантемира. На примере этих трех словарей было бы интересно проследить эволюцию русского языка на контрасте с устоявшимися греческим и латинским. Это и другие контекстуальные исследования можно будет провести на основе машиночитаемых размеченных текстов.

**Электронное представление словаря Епифания**

**Кодирование**

Изображение выглядит как текст

Автоматически созданное описаниеИтак, перед нами стояла задача разработать структуру разметки словарной статьи и написать программу, которая производила бы разметку автоматически из Excel-таблицы.

Структура самой статьи не так сложна и достаточно последовательна на протяжении словаря.

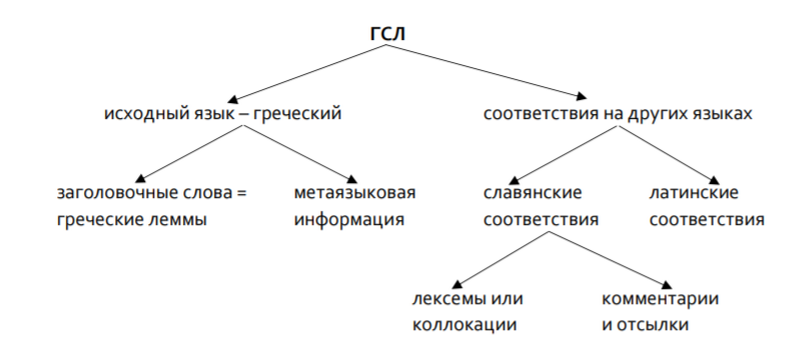
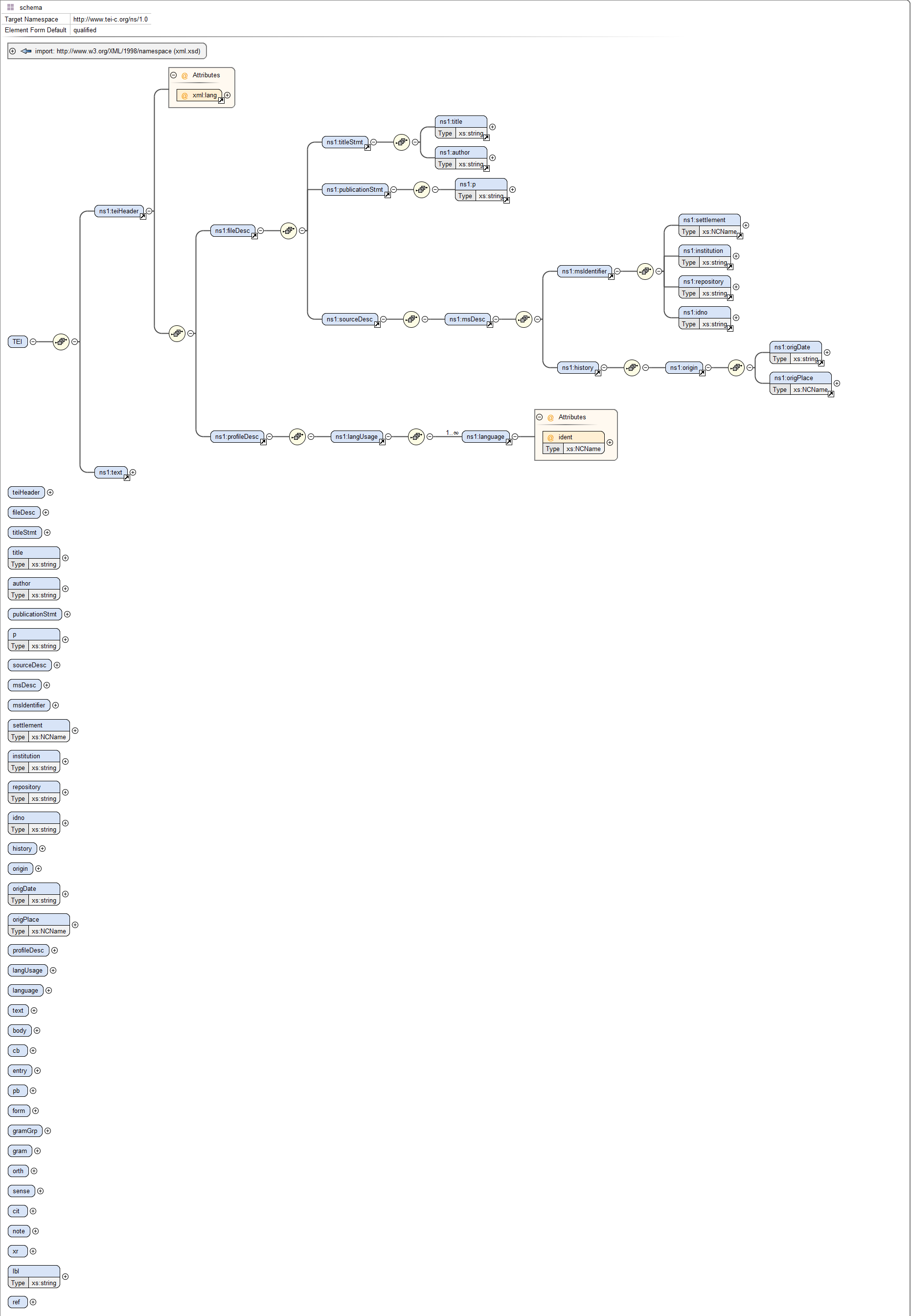
Сначала идет греческая леммма, грамматическая информация (родительный падеж для существительных и прилагательный и указание рода, форма аориста и перфекта для глаголов), затем идут славянские и латинские эквиваленты поочередно. Иногда встречается семантический комментарий: указание на синонимичность с другим словом («тожде єже Βαβάκτης») или смысловая трактовка («Таже имѧ конѧ Ахіллева»).

Схема и скан словаря © И. А. Подтергера

В нашей разметке мы старались придерживаться данной схемы, однако латинские соответствия было решено опустить ниже в иерархии, так как они всегда шли за славянским переводом и явно по смыслу являлись эквивалентами данного перевода. Получилось, что в тэге <enrty> сначала шла греческая форма, затем шел тэг <sense>, в который входила славянская форма, и затем снова шел тэг <sense> внутри славянской формы, в который входила латинская форма. все три языковых уровня мы старались приблизить по структуре и оформлению, чтобы в дальнейшем можно было инверсировать словарь и принимать любой язык за исходный.

Пробную разметку словарной статьи мы сделали вручную в Oxygen XML Editor 23.1 как образец и шаблон для последующего написания кода в Python. Ниже представлена схема тэга <teiHeader>, в котором представлена информация о самой рукописи словаря.

<teiHeader xml:lang="ru">

<fileDesc>

<titleStmt>

<title>Лексикон греко-славяно-латинский</title>

<author>Епифаний Славинецкий</author>

</titleStmt>

<publicationStmt>

<p>Не опубликован</p>

</publicationStmt>

<sourceDesc>

<msDesc>

<msIdentifier>

<settlement>Москва</settlement>

<institution>Государственный исторический музей</institution>

<repository>Отдел рукописей и старопечатных книг, Синодальное собрание</repository>

<idno>№ 488</idno>

</msIdentifier>

<history>

<origin>

<origDate>XVII в.</origDate>

<origPlace>Москва</origPlace>

</origin>

</history>

</msDesc>

</sourceDesc>

</fileDesc>

<profileDesc>

<langUsage>

<language ident="grc">Древнегреческий язык</language>

<language ident="chu">Церковнославянский язык</language>

<language ident="lat">Латинский язык</language>

</langUsage>

</profileDesc>

</teiHeader>

Так как исходные данные находились в .csv таблице, наш [код](https://github.com/wildmary/TEI-encoded-Dictionary/blob/main/словарь.ipynb) обходит каждую строчку и в соответствии с заданными условиями задает тэги. Крайне удобным оказалось использование пакета pandas[[28]](#footnote-28), который облегчает навигацию по таблице с помощью индексов строк и названий столбцов, так как большой проблемой стала структура таблицы, легко понимаемая человеком, но сложная для простого хождения по строкам.

Разберем на примере:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Βακτηρία | ας, ἡ | палица |  | baculus | baculus· |
|  |  |  | сᲄопоблюстило |  | quod | quod gress[us] servuat |
|  |  |  |  |  | gressus |  |
|  |  |  |  |  | servare |  |
|  | Βακτήριον | ου, τὸ |  | тожде |  |  |
|  | Βάκτρευμα | ατος, τὸ | паличество |  |  |  |
|  |  |  | поⷣпираніе |  | nitor | nixus in baculũ· |
|  |  |  |  |  | in |  |
|  |  |  |  |  | baculus |  |
|  |  |  | палица |  |  |  |
|  | Βακτρεύω | μ ευσω |  | тожде єже βακτήρευ· |  |  |
|  |  | π ευκα |  |  |  |  |

У греческой леммы Βακτηρία форма родительного падежа и указание на женский род и два славянских перевода – палица и сᲄопоблюстило. Первая строчка кодируется легко, однако латинский эквивалент второго славянского перевода нужно преобразовать: гиперлемма, первый латинский столбец, находится сразу на трех строчках, хотя должна быть на одной, как лемма (второй латинский столбец). Для этого нужно использовать условия и обращаться по индексу к ячейкам ниже текущей строчки и вписывать gressus и servare в ячейку к quod, а затем пропускать эти две строки, так как в них нет ни греческой, ни славянской леммы.

Далее идет семантический комментарий для леммы Βάκτρευμα «тожде», который обращается к предыдущей словарной статье, но не к предыдущей строчке – нужно проходить циклом и смотреть, когда в столбце греческой леммы появится непустая ячейка и к ней добавлять тэги <xr> и <ref> и указывать атрибутом @target греческую лемму.

у Βάκτρευμα проблема такая же, как у Βακτηρία.

У Βακτρεύω проблема в грамматическом описании: μ указывает на форму аориста, а π – на форму перфекта. Они относятся к одному слову, но находятся на разных строчках.

Итак, на выходе у нас получилась следующая разметка статьи для слова Βακτηρία (стандарт схемы можно посмотреть в [Приложении 1](#Приложение), код – на [GitHub](https://github.com/wildmary/TEI-encoded-Dictionary/blob/main/словарь.ipynb)):

<entry xml:id="Βακτηρία" xml:lang="grc">

<form type="lemma">

<orth>Βακτηρία</orth>

</form>

<form type="inflected">

<gramGrp>

<gram type="pos" value="noun"/>

<gram type="number" value="sg"/>

<gram type="case" value="genitiv"/>

<gram type="gen" value="f">ἡ</gram>

</gramGrp>

<orth extent="part">ας</orth>

</form>

<sense xml:id="Βακτηρία.1">

<cit type="translationEquivalent" xml:lang="chu">

<form xml:id="палица">

<orth>палица</orth>

</form>

<sense xml:id="палица.1">

<cit type="translationEquivalent" xml:lang="lat">

<form type="hyperlemma" xml:id="baculus">

<orth>baculus</orth>

</form>

<form type="lemma" xml:id="baculus·">

<orth>baculus·</orth>

</form>

</cit>

</sense>

</cit>

<cit type="translationEquivalent" xml:lang="chu">

<form xml:id="сᲄопоблюстило">

<orth>сᲄопоблюстило</orth>

</form>

<sense xml:id="сᲄопоблюстило.1">

<cit type="translationEquivalent" xml:lang="lat">

<form type="hyperlemma" xml:id="quod">

<orth>quod gressus servare</orth>

</form>

<form type="lemma" xml:id="quod gress[us] servuat">

<orth>quod gress[us] servuat</orth>

</form>

</cit>

</sense>

</cit>

</sense>

</entry>

Тэги словарного описания (<entry> <sense> или <cit>) обладают атрибутами xml:id или xml:lang для облегчения навигации по документу.

Данные из столбца с метаинформацией попадают в тэг <form type="inflected">. Разбиение на нужные нам группы происходит с помощью регулярных выражений. В соответствии с рекомендациями TEI Lex-0 мы обобщили грамматические тэги <pos>, <number>, <case> и <gen> в атрибуты одного тэга <gram>. Такой подход все еще поддерживается основной классификацией TEI. Также наш выбор обусловлен тем, что в тексте словарной статьи нет прямого указания на число, падеж или часть речи: все эти характеристики вложены в артикль (ἡ). Мы бы хотели кодировать как текст лишь то, что напрямую есть в статье.

Также были закодированы указания на начало страницы или столбца:

<pb n="149v">Kustode (f. 149v, linke Spalte, rechtsbündig): βακχάω</pb>

<cb pageside="149v" side="rechte">f. 149v, rechte Spalte</cb>

**Публикация на веб-сервисе**

Изображение выглядит как стол

Автоматически созданное описаниеЧтобы протестировать созданный файл с TEI XML разметкой, мы решили найти веб-сервис, который мог бы отобразить словарь в желаемом виде: с возможностью поиска и навигации по вхождениям, как в упомянутых проектах NTeALan, Расковник и церковнославянский.онлайн.

Сперва мы обратились к приложению TEI Publisher[[29]](#footnote-29) на основе базы данных eXist-db[[30]](#footnote-30), которое позволяет публиковать собственные XML файлы и настраивать их отображение. К сожалению, отсутствие возможности настроить поиск по документу (без знаний веб-разработчика) и очень сложный процесс настройки отображения показали, что этот ресурс нам не подходит на данном этапе. Мы попытались использовать опыт исследователей проекта TEI Lex-0, которые работают над созданием приложения в TEI Lex-0 Publisher[[31]](#footnote-31), но начальная стадия проекта и технические ограничения не позволили нам и здесь добиться результата.

Далее мы обратились к ресурсу ELEXIS – European Lexicographic Infrastructure[[32]](#footnote-32), который известен своим сервисом SketchEngine и множеством других инструментов. Нас заинтересовала облачная платформа для публикации словарей Lexonomy[[33]](#footnote-33). Пользователь имеет возможность создать свой веб-словарь с нуля или загрузить XML файл.

Так как Lexonomy не публикует словарь в виде приближенного к оригиналу текста, загружаемый XML файл не должен содержать других элементов кроме <entry>, которые находятся все на нулевом уровне – мы вручную удалили шапку и пометки о столбцах и страницах. Отображение и организацию загруженного словаря можно настраивать.

Изображение выглядит как текст

Автоматически созданное описаниеЕсть возможность загрузить собственную схему (DTD – Document Type Definition – или Xonomy Document Specification): после неудачной попытки сформировать структуру на основе сгенерированной в Oxygen XML Editor DTD-схемы мы добавили элементы вручную:

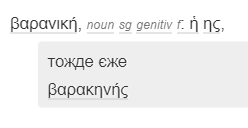
Настройки отображения позволяют показывать не только текст элементов, но и атрибуты: например, атрибут @value в грамматическом тэге. Можно настроить шрифт, положение, внутреннюю и внешнюю пунктуацию, кликабельную перекрёстную ссылку, которая производит поиск слова во всем словаре. К сожалению, мы не нашли способа сделать ссылки на другие вхождения в случае указания на них в словаре («тожде»).

Поиск в словаре можно осуществлять по любому элементу вхождения кроме метаинформации (@value): по греческим, славянским или латинским словам и по комментариям.

Некоторые кириллические буквы (ᲀ, ᲆ, ᲄᲄ) не отображаются в браузере: надеемся, что в будущем это можно исправить. Часть выносных букв отображаются, часть также выглядит как ᲄ.

Изображение выглядит как текст

Автоматически созданное описаниеLexonomy оказался очень удобным инструментом в публикации словаря на данном этапе. Ниже представлены некоторые примеры статей; [ссылка](https://www.lexonomy.eu/gsldict) на страницу словаря.

Изображение выглядит как текст

Автоматически созданное описание

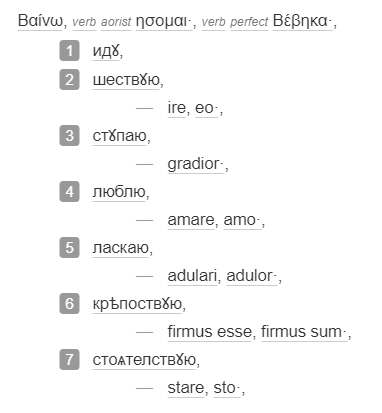


Схема словаря на сайте Lexonomy

**Заключение**

Итак, мы разработали TEI XML разметку для «Лексикона греко-славяно-латинского» Епифания Славинецкого и разместили его на веб-ресурсе Lexonomy. В ходе исследования были испробованы разные подходы, техники, инструменты и сервисы – как нам кажется, мы остановились на лучших, однако в дальнейшем при развитии главного проекта будут необходимы коррективы (например, в отображении кириллицы).

Удобство словаря для использования в базе данных и сравнительного анализа вхождений с другими словарями пока остается неизвестным ввиду начальной стадии проекта, однако уже сейчас данный словарь можно использовать как полноценный веб-ресурс – нужно только добавлять остальные вхождения (напомним, мы работали всего лишь с ~600 вхождениями).

Поученный опыт в организации и структурировании TEI XML разметки можно будет использовать и в будущих исследованиях, а также для кодирования церковнославянско-латинского словника к Библии Пискатора и церковнославянско-греческого реестра лексем, появившихся в отредактированном Евфимием Чудовским переводе „Андриант“, оцифровкой которых также занимается проект «Греко-славяно-латинский словарь Епифания Славинецкого в восточнославянском и евро­пейском языковом пространстве XVI – начала XVIII в.».

[Ссылка на репозиторий курсовой работы на GitHub](https://github.com/wildmary/TEI-encoded-Dictionary)

**Литература и ресурсы**

Брайловский С. Н. Филологические труды Епифания (Славинецкого) // Русский филологический вестник. 1890. Т. 24. №2. С. 241-250. Цит. по: Пентковская Т.В. Адаптирующие глоссы в поздних церковнославянских переводах с греческого // Вестник Московского университета. Серия 9. Филология. 2016. №1. С. 42.

Литература и культура Древней Руси: Словарь-справочник / О.М. Анисимова, В.В. Кусков, М.П. Одесский, П.В. Пятнов. Под ред. В.В. Кускова. М.: Высш. шк., 1994. С. 38.

Пентковская Т.В. Адаптирующие глоссы в поздних церковнославянских переводах с греческого // Вестник Московского университета. Серия 9. Филология. 2016. №1. С. 41.

1. Пентковская, Т. В.: Новый Завет в переводе книжного круга Епифания Славинецкого и польская переводческая традиция XVI в.: Перевод аргументов к Апостолу // Русский язык в научном освещении № 1 (31), 2016. С.  184–229.

Фонкич, Б. Л.: Греческие рукописи и документы в России в XIV – начале XVIII в. (= Россия и Христианский Восток. Библиотека. Вып. 4). М., 2003. С. 281.

Электронный обменный формат для словарей проекта TEI (Text Encoding Initiative): Учебное пособие. СПб.: СПбГУ. РИО. Филологический факультет, 2013. 80 с.

Bago, Petra & Boras, Damir. Interoperability of an 18th century Italian-Latin-Croatian dictionary. 211-222. 10.17234/INFUTURE.2015.25. <https://www.researchgate.net/publication/314636203_Interoperability_of_an_18th_century_Italian-Latin-Croatian_dictionary>

Elvis Mboning Tchiaze, Daniel Baleba, Jean Marc Bassahak, Ornella Wandji, Jules Assoumou. NTeALan Dictionaries Platforms: An Example Of Collaboration-Based Model. Proceedings of the 1st International Workshop on Language Technology Platforms 2020, pp.11 - 16. <https://hal.archives-ouvertes.fr/hal-02701912/document>

Jack Bowers, Mohamed Khemakhem, Laurent Romary. TEI Encoding of a Classical Mixtec Dictionary Using GROBID- Dictionaries. ELEX 2019: Smart Lexicography, Oct 2019, Sintra, Portugal. hal02264033f <https://hal.inria.fr/hal-02264033/document>

1. Keipert, H.: Rezension zu [Leksykon latynsʹkyj Je. Slavynecʹkoho ta A. Kore­cʹko­ho-Satanovsʹkoho. (Pamʹjatky ukrajinsʹkoji movy XVII st. Serija naukovoji literatury.) by V. V. Nim­čuk] // *Zeitschrift für Slavische Philologie* 38.2 (1975). S. 390.

Lavergne, T., Cappé, O., and Yvon, F. (2010). Practical very large scale CRFs. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pages 504–513. Association for Computational Linguistics <https://www.aclweb.org/anthology/W11-2168.pdf>

Strakhov, O. B.: Jepyfanij Slavynecʹkyj’s Greek-Slavic-Latin Lexicon: The History, Contents, and Principles Underlying the Composition of Its Greek Portion (Preliminary Re-marks) // Harvard Ukrainian Studies 28.¼ (2006). P 271.

Strakhov, O. B.: Jepyfanij Slavynecʹkyj’s Greek-Slavic-Latin Lexicon: The History, Contents, and Principles Underlying the Composition of Its Greek Portion (Preliminary Re-marks) // Harvard Ukrainian Studies 28.¼ (2006). P 275.

Toma Tasovac, Laurent Romary et al. 2018. TEI Lex-0: A baseline encoding for lexicographic data.. DARIAH Working Group on Lexical Resources: <https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html>

9 Dictionaries // P5: Guidelines for Electronic Text Encoding and Interchange: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html>

Early Irish Glossaries Database, словарь Irsan: <https://www.asnc.cam.ac.uk/irishglossaries/texts.php?versionID=17>

ELEXIS : <https://elex.is>

eXist-db: <http://exist-db.org>

GROBID-Dictionaries: <https://github.com/MedKhem/grobid-dictionaries>

Lex-0 Publisher: <http://tacheles.humanistika.org/s/SyaZ6HR2B>

1. Lexonomy: <https://www.lexonomy.eu>
2. Lingenio: <https://lingenio.de>

NTeALan: <https://ntealan.net/>

1. pandas: <https://pandas.pydata.org>

TEI Publisher: <https://teipublisher.com/index.html>

Text Encoding Initiative: <https://tei-c.org>

Transkribus: <https://readcoop.eu/transkribus>

НКРЯ, исторический церковнославянский подкорпус: <https://ruscorpora.ru/new/search-orthlib.html>

Расковник: <http://raskovnik.org>

1. Церковнославянский язык сегодня: <https://церковнославянский.онлайн>

**Приложения**

Приложение 1. Схема разметки словарной стать**Изображение выглядит как карта

Автоматически созданное описание**и

1. Ирина Александровна Подтергера , канд. филол. наук: <https://www.slav.uni-heidelberg.de/personal/ipodtergera.html> [↑](#footnote-ref-1)
2. Мария Олеговна Новак, доктор филол. наук: <http://www.ruslang.ru/node/1736> [↑](#footnote-ref-2)
3. Text Encoding Initiative: <https://tei-c.org> [↑](#footnote-ref-3)
4. Transkribus: <https://readcoop.eu/transkribus> [↑](#footnote-ref-4)
5. НКРЯ, исторический церковнославянский подкорпус: <https://ruscorpora.ru/new/search-orthlib.html> [↑](#footnote-ref-5)
6. 9 Dictionaries // P5: Guidelines for Electronic Text Encoding and Interchange: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html> [↑](#footnote-ref-6)
7. Электронный обменный формат для словарей проекта TEI (Text Encoding Initiative): Учебное пособие. СПб.: СПбГУ. РИО. Филологический факультет, 2013. 80 с. [↑](#footnote-ref-7)
8. Toma Tasovac, Laurent Romary et al. 2018. TEI Lex-0: A baseline encoding for lexicographic data.. DARIAH Working Group on Lexical Resources: <https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html> [↑](#footnote-ref-8)
9. Bago, Petra & Boras, Damir. Interoperability of an 18th century Italian-Latin-Croatian dictionary. 211-222. 10.17234/INFUTURE.2015.25. <https://www.researchgate.net/publication/314636203_Interoperability_of_an_18th_century_Italian-Latin-Croatian_dictionary> [↑](#footnote-ref-9)
10. GROBID-Dictionaries: <https://github.com/MedKhem/grobid-dictionaries> [↑](#footnote-ref-10)
11. Lavergne, T., Cappé, O., and Yvon, F. (2010). Practical very large scale CRFs. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pages 504–513. Association for Computational Linguistics <https://www.aclweb.org/anthology/W11-2168.pdf> [↑](#footnote-ref-11)
12. Jack Bowers, Mohamed Khemakhem, Laurent Romary. TEI Encoding of a Classical Mixtec Dictionary Using GROBID- Dictionaries. ELEX 2019: Smart Lexicography, Oct 2019, Sintra, Portugal. hal02264033f <https://hal.inria.fr/hal-02264033/document> [↑](#footnote-ref-12)
13. Lingenio: <https://lingenio.de> [↑](#footnote-ref-13)
14. Early Irish Glossaries Database, словарь Irsan: <https://www.asnc.cam.ac.uk/irishglossaries/texts.php?versionID=17> [↑](#footnote-ref-14)
15. Elvis Mboning Tchiaze, Daniel Baleba, Jean Marc Bassahak, Ornella Wandji, Jules Assoumou. NTeALan Dictionaries Platforms: An Example Of Collaboration-Based Model. Proceedings of the 1st International Workshop on Language Technology Platforms 2020, pp.11 - 16. <https://hal.archives-ouvertes.fr/hal-02701912/document> [↑](#footnote-ref-15)
16. NTeALan: <https://ntealan.net/> [↑](#footnote-ref-16)
17. Расковник: <http://raskovnik.org> [↑](#footnote-ref-17)
18. Церковнославянский язык сегодня: <https://церковнославянский.онлайн> [↑](#footnote-ref-18)
19. Литература и культура Древней Руси: Словарь-справочник / О.М. Анисимова, В.В. Кусков, М.П. Одесский, П.В. Пятнов. Под ред. В.В. Кускова. М.: Высш. шк., 1994. С. 38. [↑](#footnote-ref-19)
20. Фонкич, Б. Л.: Греческие рукописи и документы в России в XIV – начале XVIII в. (= Россия и Христианский Восток. Библиотека. Вып. 4). М., 2003. С. 281. [↑](#footnote-ref-20)
21. Брайловский С. Н. Филологические труды Епифания (Славинецкого) //

    Русский филологический вестник. 1890. Т. 24. №2. С. 241-250. Цит. по: Пентковская Т.В. Адаптирующие глоссы в поздних церковнославянских переводах с греческого // Вестник Московского университета. Серия 9. Филология. 2016. №1. С. 42. [↑](#footnote-ref-21)
22. Strakhov, O. B.: Jepyfanij Slavynecʹkyj’s Greek-Slavic-Latin Lexicon: The History, Contents, and Principles Underlying the Composition of Its Greek Portion (Preliminary Re-marks) // Harvard Ukrainian Studies 28.¼ (2006). P 271. [↑](#footnote-ref-22)
23. ГИМ, Москва, Син. гр. №383 (по каталогу архим. Владимира №488). [↑](#footnote-ref-23)
24. Keipert, H.: Rezension zu [Leksykon latynsʹkyj Je. Slavynecʹkoho ta A. Kore­cʹko­ho-Satanovsʹkoho. (Pamʹjatky ukrajinsʹkoji movy XVII st. Serija naukovoji literatury.) by V. V. Nim­čuk] // *Zeitschrift für Slavische Philologie* 38.2 (1975). S. 390. [↑](#footnote-ref-24)
25. Пентковская Т.В. Адаптирующие глоссы в поздних церковнославянских переводах с греческого // Вестник Московского университета. Серия 9. Филология. 2016. №1. С. 41. [↑](#footnote-ref-25)
26. Strakhov, O. B.: Jepyfanij Slavynecʹkyj’s Greek-Slavic-Latin Lexicon: The History, Contents, and Principles Underlying the Composition of Its Greek Portion (Preliminary Re-marks) // Harvard Ukrainian Studies 28.¼ (2006). P 275. [↑](#footnote-ref-26)
27. Пентковская, Т. В.: Новый Завет в переводе книжного круга Епифания Славинецкого и польская переводческая традиция XVI в.: Перевод аргументов к Апостолу // Русский язык в научном освещении № 1 (31), 2016. С.  184–229. [↑](#footnote-ref-27)
28. pandas: <https://pandas.pydata.org> [↑](#footnote-ref-28)
29. TEI Publisher: <https://teipublisher.com/index.html> [↑](#footnote-ref-29)
30. eXist-db: <http://exist-db.org> [↑](#footnote-ref-30)
31. Lex-0 Publisher: <http://tacheles.humanistika.org/s/SyaZ6HR2B> [↑](#footnote-ref-31)
32. ELEXIS : <https://elex.is> [↑](#footnote-ref-32)
33. Lexonomy: <https://www.lexonomy.eu> [↑](#footnote-ref-33)